



HAL
open science

Evaluation Framework for the Design of an Engineering Model

Walid Ben Ahmed, Mounib Mekhilef, Bernard Yannou, Michel Bigand

► **To cite this version:**

Walid Ben Ahmed, Mounib Mekhilef, Bernard Yannou, Michel Bigand. Evaluation Framework for the Design of an Engineering Model. *AI EDAM*, 2010, 24 (1), pp.107-125. 10.1017/S0890060409000171 . hal-00748720

HAL Id: hal-00748720

<https://hal.science/hal-00748720>

Submitted on 21 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evaluation Framework for the Design of Engineering Model

Short title: Evaluation Framework in Engineering

Walid Ben Ahmed^{*}, Mounib Mekhilef^{**1}, Bernard Yannou^{*}, Michel Bigand^{***}

^{*} *Laboratoire Genie Industriel, Ecole Centrale Paris*

^{**} *University of Orleans, IUFM, Department of Mathematics, rue Emile Hilaire Amagat 18000 Bourges, France. Mounib.mekhilef@univ-orleans.iufm.fr.*

Tel: +33 2 48 64 21 55 – Fax +33 2 48 64 21 55

^{***} *Ecole Centrale de Lille, Laboratoire de Génie Industriel*

Abstract

According to both *cybernetics* and *general system theory*, a subject develops and uses an adequate model of a system to widen his/her knowledge about the system. Models are then the interface between a subject and a real-world system to solve problem and to construct knowledge. Hence, evaluating these models is crucial to ensure the quality of the constructed knowledge. We propose here an evaluation framework to assess complex models based on the intrinsic properties of these models as well as the properties of the derived knowledge. A series of 40 evaluation criteria are proposed under the four systemic axes: *ontology*, *functioning*, *evolution* and *teleology*. Through a case study, we show how our evaluation model allows both presenting a given model and assessing it.

Keywords: model evaluation, knowledge evaluation, cybernetics, evaluation criteria.

1 Introduction

Modelling is a human process intrinsic to any human task [26.]. System's behaviour is administrated explicitly or implicitly by at least one model, which is directly related to a perception of the world. Models are then the basis of problem solving and knowledge construction. In an industrial engineering context as well as in a social area, models are used to construct systems. Indeed, any designed system is based on a given representation of the context and the environment in which it is supposed to evolve. For instance, to launch a transport company, an investor has to implement a representation of the market. Models are also used to analyze an existing system and therefore to understand and predict its behavior in order to steer it. For instance, a decision maker (DM) in a transport company implements a representation of the transportation system rationale as well as of its environment stating constraints to be satisfied thus determining the system behavior and consequently its performances. Thereafter, the DM's actions and decisions are guided explicitly or implicitly by this representation. Hence, as the *constructivism theory* suggests, models found any knowledge construction.

Since models are, in a sense, the interface between a subject and a real-world system, the evaluation of these models is crucial to ensure the quality of the constructed knowledge. Evaluation has been well studied in the fields of education, health, business, industry and management to mention a few. Many

¹ Corresponding author

journals and conferences deal with evaluation issues in various areas². However, the main issue considered in this paper is conceptual and addresses the epistemology of evaluation. In other words, we do not address the issue of evaluating a given *real system*, but the issue of evaluating the *quality of the model construction stage* and thus the *model* itself.

In the next section of this paper, we present a state-of-the-art of the evaluation issue and the epistemological foundations of our research. In the third section, we present an evaluation framework intended to allow a subject to assess existing or under construction models. The following section is dedicated to a case study explaining how our evaluation framework has been applied in the Kansei Engineering³ field to be used as a guideline in a modelling process intended to build road accident models. In the fifth section, we propose to characterize the interrelationships between *model* evaluation criteria and *knowledge* evaluation criteria on the one hand, and within *model* evaluation criteria themselves on the other hand.

2 State-of-the-Art

Most theories and epistemologies agree the fact that models are the interface between a *subject* and the *real world*. However, these epistemologies give different definitions to the notions of *system*, *model* and *knowledge*. Therefore, we stress that it would be misleading to deal with *model evaluation* without defining these notions as well as the notion of *evaluation* itself.

2.1 Definitions

The definitions of the following notions are required to understand the epistemological foundation of our work.

Subjectivism: the doctrine that states that knowledge and value are dependent on and limited by our subjective experience.

Relativism: the philosophical doctrine that all criteria of judgment are relative to the individuals and situations involved.

Positivism: a doctrine taught by *Auguste Comte* (1798-1857) that states that positivism is a form of empiricism that bases all knowledge on perceptual experience (not on intuition or revelation).

Constructivism: a philosophical perspective derived from the work of *Immanuel Kant* which views reality as existing mainly in the mind, constructed or interpreted in terms of one's own perceptions. Note: In this perspective, an individual's prior experiences, mental structures, and beliefs bear upon how experiences are interpreted. Constructivism focuses on the process of how knowledge is built rather than on its product or object.

2.2 Systems, models, knowledge and evaluation

Epistemology is known as the branch of philosophy that deals with questions related to the nature, the scope, and the sources of knowledge. According to Heylighen in [21.], the most fundamental question that any epistemology must answer is “*how an infinitely complex environment can be represented by a model that is necessarily much simpler than this environment and that allows a subject to derive knowledge leading to valuable predictions*”. We may distinguish two main epistemologies: *positivism* and *constructivism*.

² Performance Evaluation, American Journal of Evaluation, International Journal of Value Based Management, Business Ethics, European Journal of Engineering Education, AI-EDAM, for instance.

³ Kansei Engineering or Emotional Engineering [30.], [36.] aims at providing designers with models to help them understanding customers' needs and thereby predict their appreciation level of a new product.

Positivism, as Platonic idealism and empiricism, stresses the absolute, passive and permanent character of *knowledge*. It assumes that science should not pretend to more than what is observable and measurable. A *real system* in a positivist perspective (also called ‘hard’ perspective) is seen as a set of existing and real entities. In other words, it has features, which are universally valid, embedded in its nature and can be identified and studied as such. Thus, a *model* in such a perspective is *universal, objective and independent* from the subject who builds it. The value of a model (or of an object in general) is then independent from the evaluation context and from the subject who performs the evaluation.

However, the constructivist epistemology points out the relativity and context-dependence of *knowledge* as well as its continuous evolution. *Cybernetics* [1.],[42.] and *general system theory* [7.] are two approaches derived from this epistemology. They claim that *real systems* are open to, and interact with, their environments, and that they can qualitatively acquire new properties through emergence, resulting in continual evolution. Rather than reducing an entity only to the properties of its parts or constituting elements, cybernetics and the general system theory focus on the relationship between the parts, which gathers them as a whole (the *holism* principle). Hence, a *model* is considered as a *perception* of the real-world in a given context. It is constructed by a subject for a given purpose. Then, in contrast to positivism, a model in constructivism is not dissociated from the subject who builds it.

2.3 Evaluation’s influence on attitudes, perceptions and actions

In this paper, we assimilate an evaluation task to an interaction process between the *subject* who performs the evaluation and the *evaluated object* (which is the *model* in our case). As it is emphasized by relativism, subjectivism and constructivism, the perception of the subject and his/her personal background influence the values he/she assigns to the object. But, on the other hand, during the interaction process, the evaluation tasks may in turn influence the evaluators’ perception. In fact, in [19.],[18.][24.], the authors have noticed that assessments can influence perceptions of social problems, selection and implementation of social policies. Furthermore, they encourage evaluators to rethink the outcomes influenced by assessments.

Based on these researches, one may notice that assessment can influence perceptions and thereby actions. Since models are the result of perceptions, we can assume that model evaluation can influence the modelling task itself. This can be explained as feedback behaviour when a subject evaluates a model (or an object in general), his/her perception may be influenced: he/she may notice some incompleteness, misrepresented aspects, etc. Then, he/she may carry out new actions in order to tackle the identified limits and thereby the model may be changed.

Since evaluation affects the subject’s perception and since modelling is based on perceptions, it would be misleading to carry out a modelling task without focusing on the issue of model evaluation. This is why it’s worth determining some criteria to assess, for each stage of the modelling process, the adequacy of the model with the initial modelling objectives. The aim of this paper is to develop a framework for model evaluation and to use this framework as a guideline in the modelling process or as a guide to select a given model for a given objective. The context dependency and the subjectivity of models as well as of their evaluation from a constructivist point of view may lead to pure relativism. Nevertheless, the present paper advocates that, in spite of the variability and subjectivity of models, a number of criteria can be formulated in order to help a user selecting an adequate model among a list of existing alternatives or to validate the *a priori* quality of a model being implemented.

2.4 Knowledge evaluation

As we mentioned in the introduction, a model is not an objective in itself, but a tool to develop a goal-dependent knowledge. An ‘adequate model’ is then the one that permits to derive an ‘adequate knowledge’. Thus, the question about model assessment may be transformed into a question about knowledge assessment as shown in Figure 1.

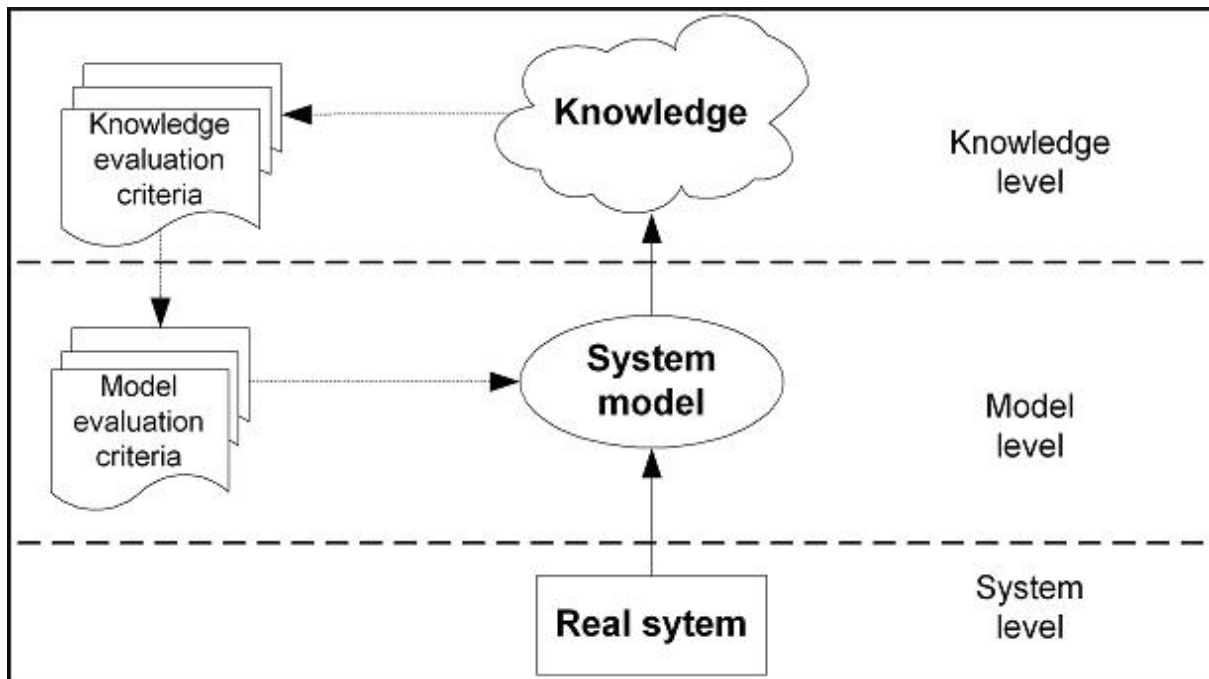


Fig. 1. The relationship between model evaluation and knowledge evaluation.

The basic question the epistemology attempts to answer is: what distinguishes *true (or adequate) knowledge* from *false (or inadequate) knowledge* [9.],[21.][20.]. In other words, how can knowledge quality, soundness, etc... be evaluated?

The dualistic debate between absolutism and relativism in philosophy arises in epistemology. Indeed, on the one hand positivist theories stress the absolute, passive and permanent character of knowledge and thereby try to formulate unambiguous and fixed criteria for distinguishing "true" or "real" knowledge from "false". On the other hand, constructivist theories stress the relativity and evolution of knowledge and therefore try to formulate subjective criteria that are more context-dependent (see for instance [9.],[21.][33.].

In spite of the variability and subjectivity of knowledge, a number of researches have been carried out in order to formulate criteria that allow distinguishing between adequate knowledge and inadequate knowledge (see [9.],[21.][33.]).

As a matter of fact, Turchin [41.] claims that the essential function of knowledge is prediction and since there is no universal and absolute criterion of truth, the unique criterion of truth is the prediction power that the concerned knowledge is able to provide. In other words, 'true' knowledge is the one that allows a system to handle different types of perturbations by anticipating them and testing (and further selecting among) possibly adequate actions that could contribute to its survival [21.].

Another point of view provides a natural definition of what 'true' or 'real' knowledge means: it is the *selectionist* point of view which states that 'true' or 'real' knowledge is knowledge that can survive. This *selectionist* point of view stems from Campbell's "*evolutionary epistemology*" [9.] and Heylighen's "*evolutionary-cybernetic epistemology*" [21.]. Hence, *knowledge assessment criteria* may result in *knowledge selection criteria*.

Reich [34.] using a constructivist approach addressed the issue of the measure of knowledge. In particular he demonstrated the need to use several different measures simultaneously rather than a single assessment.

Heylighen [21.] distinguishes three super-classes of criteria that are used by a subject to select a given knowledge: *objective criteria*, *subjective criteria* and *inter-subjective criteria*:

- **Objective criteria** are those used for judging "*objectivity*" or "*reality*" of knowledge or a given perception in general. The first objective criterion is related to knowledge *invariance*. Indeed, there is a part of "solid"/objective knowledge related to a given phenomenon that must persist even when

its perception (i.e. how perception is carried out, perception context, perception means, time of perception etc.) is no more active or changed. Heylighen distinguishes three types of invariance: (a) *invariance over modalities*: perception should be the same even though it is performed through different senses, points of view, or means of observation; (b) *invariance over time*: perception should be the same even though it is performed at different moments in time; (c) *invariance over persons*: perception should be the same even though it is performed by different observers. The second objective criterion is related to knowledge *distinctiveness*: a “real” perception is the one which can be characterized in details, structured in a coherent manner and represented by a distinct pattern. Dreaming, for example, is not “real” because it is a coarse-grained and fuzzy set of perceptions. The third objective criterion is *controllability*: a knowledge that reacts differentially to the different actions performed on it is more likely to be real than one that changes randomly or not at all.

- **Subjective criteria** are those related to how efficiently knowledge can be assimilated by the individual subject. For instance, despite its objectivity, the relativistic quantum field model of the beryllium atom is assimilated by very few people. Since the capacity of a cognitive system is limited and learning is based on strengthening associations, useless knowledge, complex knowledge and knowledge into conflict with existing knowledge burdens the subject and reduces the chances for survival. Therefore, the first subjective criterion is related to the *individual utility* of knowledge: it is postulated that a subject will only make the effort to learn and retain an idea that can help him/her reach his/her goals. The second subjective criterion is related to the *simplicity* of knowledge (easy to learn): the more complex an idea, the higher the burden on the cognitive system, the lower the chance for the knowledge in question to be selected. This idea is the same as the information axiom within the axiomatic design theory by Suh [39.]: the lighter the information required for the design process of a product to put on the market, the more likely the product is to be inexpensive, robust in terms of adaptability to a usage context, easy to reengineer and finally the more competitive it is likely to be and the more certain to survive. This is also related to the information entropy theory. The third subjective criterion is related to knowledge *consistency*: the ease with which a cognitive system assimilates new ideas depends on the support it gets from ideas assimilated earlier. In other words, ideas that do not connect to existing knowledge simply cannot be assimilated. The last subjective criterion is *novelty*: new, unusual or unexpected ideas or perceptions tend to attract the attention, and thus arouse the cognitive energy that will facilitate their assimilation.
- **Inter-subjective criteria** are related to the capacity of knowledge to be transmitted and assimilated easily. Heylighen [20.] proposes the following criteria: (a) *publicity*: it may be related to the subject’s motivation (the effort the subject carrying the idea invests in making it known to others) or to knowledge itself (simplicity, consistency, novelty, etc.); (b) *expressivity*: it depends on the whether the knowledge can be expressed in a clear and easy language; (c) *formality*: the possibility for an idea to be formulated in a less context-dependent way, so it can be assimilated equally by different subjects; (d) *collective utility*: some forms of knowledge benefit to the community, while being useless for an isolated individual; (e) *conformity*: Campbell stresses that a community achieves a selective pressure that removes individual selfish deviations from these collective beliefs; (f) *authority*: the backing of a recognized expert contributes to the acceptance and the legitimacy of a given idea.

3 A generic framework for model evaluation

One of the main issues when considering model evaluation is how complete the evaluation framework is. Moreover, many viewpoints may be used to evaluate a model; what are the relevant viewpoints? Which criteria must be satisfied to produce an “adequate knowledge”?

To address these issues, we use the cybernetic and systemic approaches: we consider a model as ontology (ideas, expressions, rules, patterns) open to and that interacts with its environments through a

given *functioning*. It can qualitatively acquire new properties resulting in *continuous evolution* in order to fulfil a given *teleology* (goal/motivations of the subject prior to the model implementation).

Hence, our evaluation framework consists in four generic viewpoints: *ontology*, *functioning*, *evolution* and *teleology*. We are hereunder proposing a collection of evaluation criteria according to each of these systemic viewpoints.

3.1 Evaluating the model ontology

The *model ontology* consists of concepts used to represent the real system and/or the phenomena we are modelling. A concept is an abstract idea or a mental symbol, typically associated with a corresponding representation in a language or symbology. Hence, two important aspects must be considered in the evaluation of model ontology: the *model concepts* and the *model representation formalism*. To assess model ontology, we propose the following criteria:

- **Self-descriptiveness** of the model ontology: it is the ability of the model concepts to embed enough information to explain the model objectives and properties. This criterion is related to the choice of the model concepts as well as the representation formalism in which these concepts are expressed. There exist several representation techniques such as graphs [37], text, mathematical grammars, frames, rules etc. to represent a model ontology. The model representation formalism is crucial to help, for instance, a subject to present and transmit his/her models or a group to share a common model. The more *self-descriptive* the model, the more *expressive* the knowledge expressed through the model (i.e. easy to be expressed in a clear and easy language) and the easier the *publicity* of this knowledge.
- **Consistency** of the model ontology is a second criterion to ensure the model coherence and self-descriptiveness. It is related to the degree of uniformity, standardization, and freedom from contradiction among the model concepts. Consistency is crucial to satisfy the two following knowledge subjective criteria: *simplicity* and *consistency* and thereby the *publicity* inter-subjective criterion. Indeed, the more *consistent* the model ontology, the easier the knowledge expressed through this ontology (i.e. *simplicity*), the higher the support this knowledge gets from ideas assimilated earlier (i.e. *consistency*) and thereby the better the concerned knowledge is transmitted (i.e. *publicity*).
- **Incompleteness** of the model ontology: it is related to the lack of a concept or a misspecification of one of the concepts. An incomplete model might make the concerned knowledge more difficult to formulate and therefore more difficult to transmit and assimilate.
- **Independence** of the model ontology: it is related to the independency of the model from the subject who has elaborated it. Model ontology satisfying this criterion would improve the *formality* of the concerned knowledge (i.e. possibility to formulate knowledge in a less context-dependent way), its *collective utility* (i.e. its benefit to the community, while being useless for an isolated individual) and its *invariance over persons*.

3.2 Evaluating the model functioning

The *model functioning* is characterized by the model interaction with its environment (constraints of use, objective of use, inputs, etc.) to satisfy the model teleology (i.e. goal). Three important aspects must be considered to correctly assess the model functioning: the *model interaction with users*, the *model behavior under normal conditions* and the *model behavior under stressful conditions* (e.g. erroneous input, varied constraints, etc.). In other words, criteria that should be satisfied by model functioning are related to these three aspects. Furthermore, these criteria should be defined such that the knowledge expressed through the concerned model satisfies the knowledge criteria we have defined. Based on these assumptions, we define the following criteria grouped into the three super-classes already mentioned:

3.2.1 Evaluating the model interaction with users

The evaluation of a model interaction with its users consists in characterizing the facility of use and the reusability of the model. This leads to the following criteria:

- The **attractiveness** of the model: it is related to how attractive the model may be to the user. This refers to attributes of the model ontology intended to make the model more attractive for the user, especially attributes related to the representation formalism such as the use of color, the nature of the graphical design, etc. This criterion is also related to the previous criteria (i.e. *consistency*, *self-descriptiveness* and *independence*). This criterion may improve the *publicity* criterion of the expressed knowledge.
- The **reusability** of the model: it is related to the efficiency of the model in facilitating a selective use of its components or sub-models.
- The **usability** of the model: it is related to how the model allows the user to learn in order to operate, prepare inputs for, and interpret outputs.
- The **abstractness** of the model: how a model allows a user to perform only the necessary functions relevant to a particular purpose.
- The **understandability** of the model: it is related to how the model permits the user to understand whether the model is suitable for a given modelling purpose, and how it can be used for particular tasks and conditions of use.
- The **learnability** of the model: it is related to how the model itself helps the user learn more on the modelled phenomena and application.
- The **adaptability** of the model: it is related to the ease with which the model meets contradictory and variable users' constraints and users' needs.
- The **operability** of the model: is related to how the model allows the user to operate and control it. Aspects of *suitability*, *changeability* and *adaptability* may affect the model *operability*. Operability corresponds to *controllability*, *error tolerance* and *conformity* with users' expectations that we will present in the following paragraphs.

Criteria related to the model-user interaction such as *reusability*, *understandability*, *adaptability*, *learnability*, etc. play a relevant role to ensure certain subjective criteria of the knowledge expressed through the model. Indeed, the more *usable*, *reusable*, *understandable*, *adaptable*, *learnable* and *operable* the model is, the higher the *individual utility*, the *simplicity*, and *consistency* of the inherent knowledge.

3.2.2 Evaluating the model behavior under normal conditions

- The **controllability** of the model: it is related to how efficiently the model reacts differentially to the different actions it is submitted to.
- The **repeatability** of the model: it is related to how the model generates the same results under the same functioning conditions.
- The **generality** of the model: it is related to how the model performs a broad range of functions.
- The **interoperability** of the model: it is related to the ability of two or more models or model components to exchange information and to use the information exchanged.
- The **replaceability** of the model: it is related to how the model can be used instead of another specified model for the same purpose in the same environment.

- The **usability compliance** of the model: it is related to how the model complies with standards, conventions, style guides or regulations relating to usability.

3.2.3 Evaluating the model behavior under stressful conditions:

Stressful conditions may be related to input quality (e.g. errors, incompleteness, noise, inconsistency etc.), model component faults, and constraints of use (e.g. use duration, use period, validity domain, different types of stimulation allowed, etc.). The general criteria referring to the assessment of a model functioning under stressful conditions is **robustness** and **reliability**. It is defined as the ability of a model or a model component to function correctly in the presence of invalid inputs or stressful environment conditions or unexpected circumstances. Robustness and reliability can be characterized through the following criteria:

- **Error tolerance**: it is related to the ability of the model to continue an operation normally despite the presence of erroneous inputs.
- **Fault tolerance**: it is related to the ability of a model to continue an operation normally despite the presence of model component faults.
- **Error proneness**: it is related to the ability of a model to allow the user to intentionally or unintentionally introduce errors into the model or misuse the model.

The model **robustness** criteria (i.e. *Error tolerance*, *Error proneness*, *Reliability*, *Controllability*, etc.) make the knowledge expressed through the model concerned satisfy especially the objective criteria introduced previously. Indeed, **robustness** criteria improve knowledge *invariance over input modalities*, knowledge *invariance over time*, and knowledge *invariance over persons* and knowledge *controllability*.

3.3 Evaluating the model evolution

The *model evolution* is characterized by its transformation (i.e. structural or functional) due to an internal or external change. An *internal transformation* may affect a given function, component or attribute of the model itself. Example: a function or a component is defective, another component or function is added or improved, etc. An *external transformation* may affect the model environment. Example: a new use environment, a new input, a new application, a new requirement, a new user, new constraints, etc.

The evaluation of a model evolution consists in assessing the **modifiability** of the model: the ease with which a model or model component can be modified to correctly fit evolutions and changes. To handle changes, a model should be able to evolve. Hence, *model evolution* refers the following criteria:

- **Flexibility** depends on how easily modifications can be carried out in order to use the model in applications or environments other than those for which it has been specifically designed.
- **Extendibility** (or *Expandability*) is related how easily modifications can be performed in order to increase the model functional capacity.
- **Maintainability** is related to how easily modifications can be carried out in order to correct model faults.
- **Testability** is related to how easily modifications can be performed within the validation stage of the complete model under construction.

3.4 Evaluating the model teleology

The *model teleology* is the goal of its elaboration. Assessing model teleology consists in measuring the gap between the users' needs and the effective functions the model fulfils. This gap is measured through the following criteria:

- **Accuracy/Precision:** how well the model provides the right or agreed results or effects with the expected degree of accuracy;
- **Efficiency:** how well the model provides an appropriate performance, relatively to the amount of resources used (time, human resources, etc.), under stated conditions;
- **Effectiveness:** the ability of the model to target all aspects of the goal.

4 A case study of model evaluation

In the following sections, we show how our evaluation model allows both presenting a given model and evaluating it. We choose as a case study a model we constructed and presented in [3].

A set of 11 automotive experts (of sales departments) have been gathered for a whole day of evaluation of 10 dashboards of recent cars belonging to the same marketing segment (of small cars), namely: (1) Audi A2, (2) Citroën C2, (3) Fiat Idea, (4) Lancia Ypsilon, (5) Nissan Micra, (6) Peugeot 206, (7) Renault Clio, (8) Renault Modus, (9) Toyota Yaris, (10) Volkswagen Polo. The 11 subjects have been immersed in a decision context. They have been described a target user profile and a purchasing situation. During this workshop, the 11 subjects are asked to assess dashboard pictures without actually seeing or touching these dashboards. We are conscious that there is a bias but it is also a way to isolate the dashboards since the car brands are not displayed and they are even removed from the pictures.

4.1 Presentation of our model

4.1.1 Presentation of the model teleology (objective)

A design process can be seen as an iterative and complex process guided by a final and ultimate objective, which is to make the developed product fitting the customer aspirations. Hence, predicting customers' satisfaction level when one develops a new product is fundamental. That is the aim of the model we use here as a case study. It is stemmed from Kansei Engineering (or Emotional Engineering) [30] [36], which provide designers with models to help them understanding customers' needs and thereby predict their appreciation level of a new product.

In other words, the teleology of our model is to allow designers answering the two following questions:

- 1) *What is the impact of a given decision related to the design parameters (i.e., technical and/or functional parameters) on the final customer perception?* And on the other hand
- 2) *Given a customer expected need, what are the optimal technical choices designer has to perform in order to satisfy the customer need?*

The relevance of the answer to these questions depends on the quality of our kansei model. Thereby, the evaluation of this model is crucial.

4.1.2 Presentation of the model ontology

As we noticed in section 3.1, the model ontology includes the *model concepts* as well as the *representation formalism*.

4.1.2.1 Presentation of the model concepts

The several concepts of our kansei model are described in Figure 2.

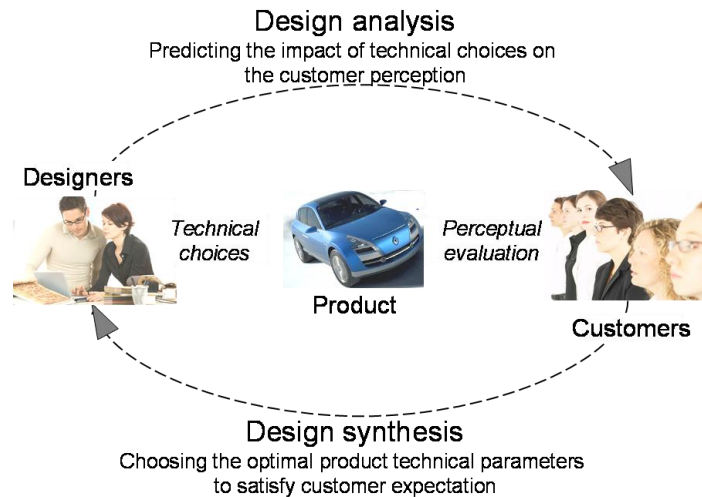


Fig. 2. Several concepts of a kansei model

A kansei model can be seen as an interaction between the following concepts:

- The **product** to be designed: in our case: car dashboards (As those represented in Figure 3)
- The **customer**: car users
- The **designer**: dashboard designers.

The interaction between these three concepts is expressed through two types of attributes:

- **Technical attributes**: they characterize the dashboards. The role of a designer is to choose the adequate technical attributes. In a sense, technical attributes are the result of the interaction between designers and dashboards
- **Perceptual attributes**: they describe the customer assessment of the dashboards. In a sense, perceptual attributes are the result of the interaction between customers and dashboards.

As we noticed in (section 4) the model building is based upon a data collecting protocol that has been described in [43] and already experimented on another case study in [44, 32]. Ten automotive dashboards (AUDI A2, CITROEN C2, FIAT Idea, LANCIA Ypsilon, NISSAN Micra, PEUGEOT 206, RENAULT Clio, RENAULT Modus, TOYOTA Yaris, VW Polo} are evaluated by 11 customers (cf. Figure 3).

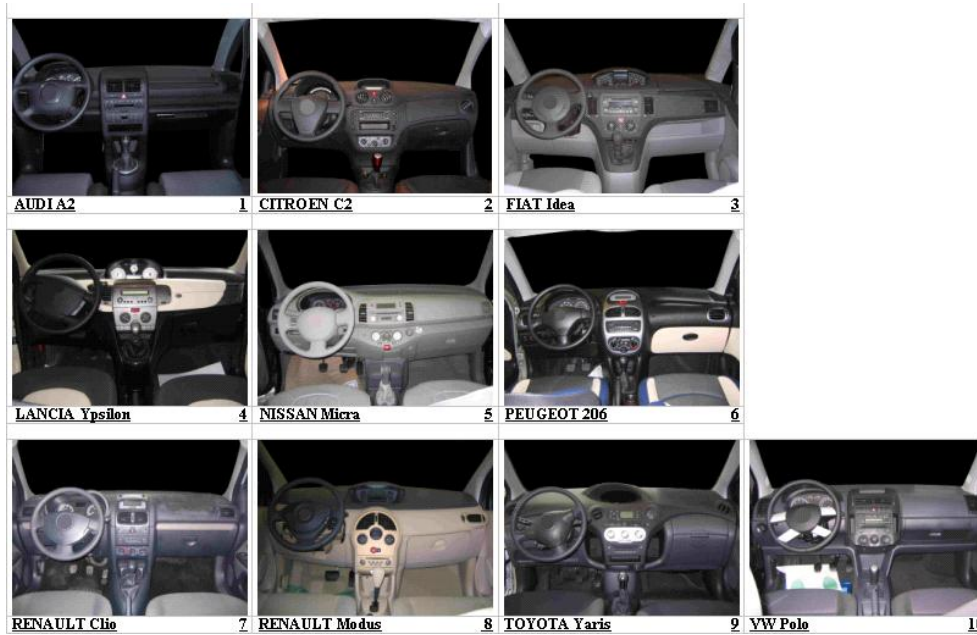


Fig. 3. The 10 dashboards evaluated by customers

We defined a set of 8 technical attributes characterizing the dashboards with corresponding modalities (two at least but the number may increase): the “*Speedometer dial position*”={*behind steering wheel, at the center of the dashboard*}, “*Display lay-out*”={*Analogue, Digital*}, “*Air conditioner control*”={*Button, Other*}, “*Air vent shape*”={*Rounded, Square*}, “*Dashboard color*”={*Single color, Two colors*}, “*Aerator shape*”={*Rounded, Square*}, “*Arrangement space*”={*Many, Few*} and “*Style lay-out*”={*Curved lines, Straight lines*}. The characterization of the 10 dashboards according to the technical attributes is objective and do not depend on the preference of customers. It is presented in table 1.

Table 1. The technical characterization of the 10 dashboards

Dashboards	Speedometer Dial Position	Display layout	Air Conditioner Control	Air Vent Shape	Dashboard colour	Aerator shape	Arrangement Space	Style layout
AUDI A2	behind steering wheel	analogue	button	square	single colour	square	many	straight lines
CITROEN C2	behind steering wheel	digital	other	rounded	single colour	rounded	few	curved lines
FIAT Idea	at the centre	analogue	other	square	two colours	square	many	straight lines
LANCIA Ypsilon	at the centre	analogue	other	square	two colours	square	many	curved lines
NISSAN Micra	behind steering wheel	analogue	button	rounded	single colour	rounded	few	straight lines
PEUGEOT 206	behind steering wheel	analogue	other	rounded	two colours	rounded	few	curved lines
RENAULT Clio	behind steering wheel	analogue	other	square	single colour	square	few	straight lines
RENAULT Modus	at the centre	digital	button	rounded	two colours	rounded	many	curved lines
TOYOTA Yaris	at the centre	digital	other	rounded	single colour	rounded	many	curved lines
VW Polo	behind steering wheel	analogue	other	square	single colour	square	few	straight lines

Ben Ahmed W., Mekhilef M., Yannou B., Bigand M., (2010), *Evaluation Framework for the Design of an Engineering Model*. AI EDAM: Artificial Intelligence for Engineering Design, Analysis and Manufacturing, vol. 24(1): p. 107-125, doi:10.1017/S0890060409000171.

We also defined a set of 11 perceptual attributes, which describe the customer assessing of the “*Space organization*”, “*Control button comprehensibility*”, “*Aerator lay-out*”, “*Arrangement space*”, “*Comfort*”, “*Simplicity*”, “*Sportive lay-out*”, “*Masculinity lay-out*”, “*Quality*”, “*Novelty*” and “*Harmony*” (see [17] for details on attributes). The customer evaluations of the dashboard perceptual attribute levels is made in qualitatively pairwise comparing the 10 dashboards under each of the 11 perceptual attributes (see [27] for mathematical details). It leads to 11 normalized score vectors. The advantage of this method is that the value scale is automatically built thanks to the pairwise comparison mechanism without the need to define a specific metrics (for instance, a score of 0.1 for the “*Masculinity lay-out*” means much more feminine than a score of 0.3). Next, each normalized score vector (the scores sum is 1) is transformed to fit into a standard scale of [0, 20]. Finally, continuous attribute levels are projected into discrete categories: [0, 5]=*Very low*, [6, 10]=*Low*, [10, 14]=*Medium*, [15, 17]=*High*, [18, 20]=*Very high*.

As 11 customers have participated to this study, a 110×19 matrix is then constructed: rows=10 dashboards \times 11 customers, columns=8 technical attributes & 11 perceptual attributes.

4.1.2.2 Presentation of the model representation formalism

As we noticed in section 3.1, we can use several representation techniques such as graphs [37], text, mathematical grammars, frames, rules, etc. to construct our model. In [44], we used the Principle component Analysis (ACP). While in [43], we used Bayesian Networks (BN) [23] as the representation formalism. In this paper we briefly describe the second model.

BN are directed acyclic graphs used to represent uncertain knowledge in Artificial Intelligence [23]. A BN is defined as a couple: $G=(S, P)$, where:

- $S=(N, A)$ represents the structure (i.e. the graph);
 - “ N ” is a set of *nodes*. Each node represents a discrete variable X having a finite number of mutually exclusive states (modalities). In our case study, X may be a perceptual attribute as well as a technical attribute;
 - “ A ” is a set of *edges*; the relation “ N_1 is a parent of N_2 ” is represented by an edge linking N_1 to N_2 . In our case study, an edge may be interpreted as a causal relation.
- P represents a set of *probability distributions* that are associated to each node. When a node is a *root node* (i.e. it does not have a parent), P corresponds to the probability distribution over the node states. When a node is not a root node, i.e. when it has some parent nodes, P corresponds to a *conditional probability distribution* that quantifies the probabilistic dependency between that node and its parents. It is represented by a *Conditional Probability Tables (CPT)*.

Figure 4 represents the BN we obtained through an automatic learning on the data. The presentation of the learning approach is out of the scope of this paper (See [25] and [3] for more details on the learning approach we used).

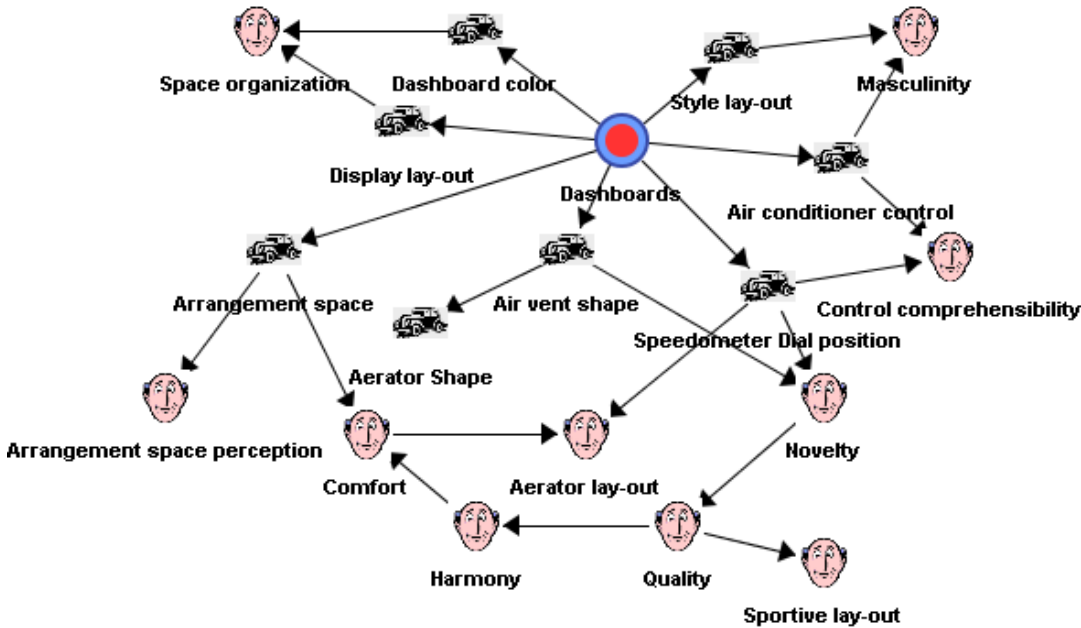


Fig. 4. Unsupervised learning to identify probabilistic relationships within the data (i.e. between dashboard physical - car icon - and perceptual - face icon - attributes)

Edges in this bayesian network can be interpreted as causal relationships. For instance, according to Figure 4, the subjective attribute “Novelty” depends on the two physical attributes “Air vent Shape” and “Speedometer position”. Each relation (i.e. edge) is expressed through a conditional probability table, which is automatically computed. For example, the relation between “Novelty”, “Air vent Shape” and “Speedometer position” is represented through Table 2.

Table 2. Conditional probabilities representing the causal relation between “Air vent Shape”, “Speedometer position” and “Novelty”. According to this table: $P(\text{novelty}=\text{very low}/\text{Speedometer dial position}=\text{at the center} \ \& \ \text{Air vent shape}=\text{rounded})=13.6\%$

Speedometer dial position	Aerator shape	Novelty				
		Very low	Low	Medium	High	Very High
At the centre	Rounded	13.6	36.4	31.8	9.1	9.1
	Square	27.3	36.4	27.3	0.0	9.1
Behind steering wheel	Rounded	24.2	60.6	9.1	6.1	0.0
	Square	75.8	24.2	0.0	0.0	0.0

4.1.3 Presentation of the model functioning

We notice here that the constructed model (cf. Figure 5) allows identifying three types of relationships:

- (a) Relationships within technical attributes. For example, “Air vent shape” has a direct impact on the “Aerator shape”

- (b) **Relationships within perceptual attributes.** For example, “*harmony*” perception has a direct impact on “*comfort*” perception
- (c) **Relationships between technical and perceptual attributes.** For example, the two physical attributes “*Air vent Shape*” and “*Speedometer position*” have an impact on the “*Novelty*” perception.

Because a Bayesian network is a complete model for the attributes and their relationships, it can be used to answer probabilistic queries about them. For example, the network can be used to find out updated knowledge of the state of a subset of attributes when other attributes (the evidence attributes) are observed. This process of computing of the posterior distribution of attributes given evidence is called *probabilistic inference*. Inference in BN [22] allows then taking any state attribute observation (an event) into account so as to update the probabilities of the other attributes. Without any event observation, the computation is based on a priori probabilities. When observations are given, this knowledge is integrated into the network and all the probabilities are updated accordingly.

A kansei bayesian network provides designers with several use – or simulation - scenarios. We present here only the main scenarios: the analysis scenario and the synthesis scenario (see [3] for all the use scenarios presentation).

4.1.3.1 Analysis scenario

The analysis scenario allows answering the question “what is the probable impact of the choice related to physical attributes on the other design attributes and especially on the perceptual attributes”. Let us consider the speedometer dial position as an example of such a design impact. According to the model presented in Figure 4, the speedometer dial position has an impact on the dashboard “novelty perception” as well as on the “control comprehensibility”. This model not only helps the design to identify the relevant relations between this particular technical attribute and the other design attributes, but also allows him knowing in which proportions it impacts them. For instance, the model states that a dashboard whose speedometer dial is located at the center is perceived by customers as more novel than a dashboard whose speedometer dial is located behind the steering wheel. However, that choice deteriorates the control comprehensibility. In a sense, the model allows designer comparing the two possible technical choices related to the speedometer dial position (i.e. at the center or behind the steering wheel) in a multicriteria way (cf. Figure 5) with a certain confidence depending on the learning set of assessed dashboards.



Fig. 5. The influence of the speedometer dial position on the dashboard novelty lay out as well as on the control comprehensibility: a dashboard whose speedometer dial is located at the center is perceived by customers as more novel than a dashboard whose speedometer dial is located behind the steering wheel. However, that choice may deteriorate the control comprehensibility.

4.1.3.2 *Synthesis scenario*

The synthesis scenario allows answering the question “what are the best choices (related to technical attributes) the designer must make so as to configure the level of a perceptual attribute as expected”. The same model presented in Figure 4 allows a designer identifying all possible design choices that let him optimizing the level of a given perceptual attribute (or performance). As an example, we take the “dashboard novelty perception” as target attribute to optimize and show how our BN model allows identifying the best technical choices designers can perform to improve that attribute.

Figure 6 shows that to improve “dashboard novelty perception”, designers should carry out the following choices: a speedometer dial position at the center of the dashboard, two colors instead of single color, digital display instead of analogue, rounded air vent shape, many arrangement spaces and curved lines, etc.

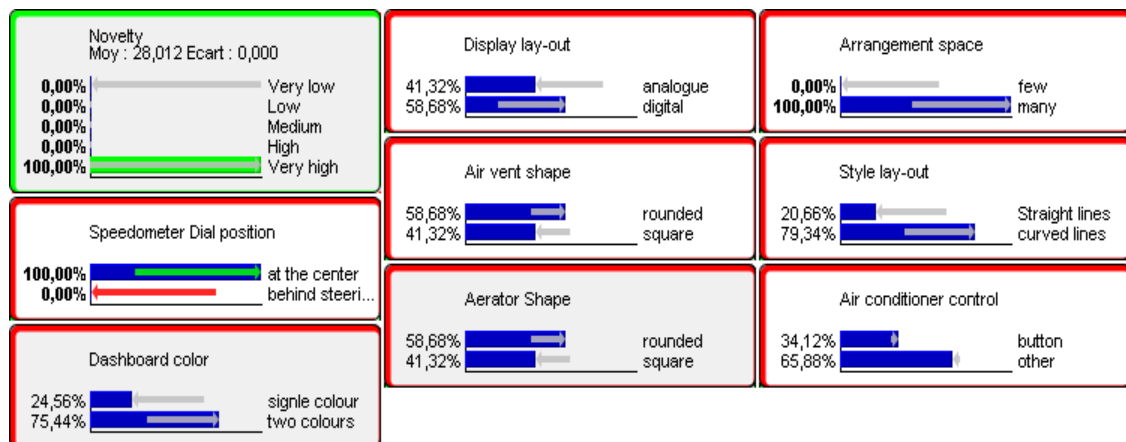


Fig. 6. The optimal technical choices designer should carry out in order to improve the novelty perception of a dashboard.

4.1.4 Presentation of the model evolution

As noticed in section 3.3, a model evolution is characterized by its transformation (i.e. structural or functional) due to an internal or external change. One of the main determining advantages of Bayesian network approach is its ability to evolve in order to integrate changes. Many reasons may be a cause of change:

- A structural inconsistency: Since input data may be not representative of the reality, there may be an inconsistent relationship between two nodes (i.e. attributes). In this case, the user may easily modify the model structure to handle such an inconsistency instead of using the causal network computed after a given learning algorithm. Then, the user may remove an edge if he believes that there is no apparent causal relation between the correspondent nodes and restart a quantitative updating of inner conditional probability tables. Likewise, the user may add an edge between two nodes if he believes there is a causal relationship between them even if the learning algorithm has not detected the relation. He may also modify the orientation of a given edge. Let us take the example of the model presented in Figure 4. This model states a strong probabilistic correlation between “comfort” perception and “aerator style-out” perception. However, the edge orientation states that “comfort” perception has an impact on “aerator style-out” perception. It is easy to detect this “structural” inconsistency since the inverse is more coherent. In such a case, the user has just to change the edge orientation to make this relationship causally more relevant. There is apparently no change in the levels of node modalities but there is a local recomputation of the conditional probability table and a next simulation through the BN will lead to different results.
- An analytical incoherence: it is related to conditional probabilities characterizing attributes relationships. Let us take the example presented in Table 2: based on his experience, the user can change the figures that represent the conditional probabilities linking attributes if he believes that the figures do not represent the reality (when there is a lack of data form example)
- An update of the mode inputs: if there is an evolution of the input data used to learn the BN model, the user has just to perform a new learning of the BN model on the new data. The structure as well as the conditional probabilities is automatically updated.

In a sense, a BN model allows a user to integrate his knowledge as well as the knowledge embedded in new data.

4.2 Evaluation of our kansei model

In the following we present the different assessment of our model along the four systemic axis as developed above.

4.2.1 Evaluation of our model ontology

In the following table, we assess the model ontology (concepts and representation formalism) according to the criteria we presented in section 3.1.

Table 3. Evaluation of the model ontology

Criteria	Evaluation
Evaluation criteria related to model concepts	Assessment of our kansei model concepts
Incompleteness: the risk of missing a concept or a misspecification of one of the concept.	As we noticed in section 4.1.1, the aim of a kansei model is to provide designers with models to help them understanding customers' needs and thereby predict their appreciation level of a new product. The concepts of product (car dashboard), customer (car user) and designer (car dashboard designer) as well as the interaction between them (perceptual attributes and technical attributes) are sufficient, complementary and consistent enough to embed all needed information required to perform a complete model. Hence the completeness and consistency are satisfied.
Consistency: the degree of uniformity, standardization, and freedom from contradiction among the model concepts.	Self-descriptiveness is one of the main powerful characteristics of models built using Bayesian Network approach. In fact, graphical formalism of Bayesian Network (cf. Figure 2). Allows at the same time to represent all the concepts as well as their qualitative relationship (expressed through edges) and quantitative relationship (expressed through Conditional probabilities such as those presented in Table 2).
Self-descriptiveness: the ability of the model to embed enough information to explain the model objectives and properties.	- The attributes (technical and perceptual) choice process as well as the collect data protocol was performed in a way ensuring their independence from the persons who elaborated them (see [19, 50]. Of course, despite a clear elaboration process to obtain a list of major perceptual attributes and technical parameters, it remains a part of non determinism.
Independence: the independency of the model from the subject who has elaborated it.	- Moreover, the model is automatically constructed using BN learning on data. Thereby, the structural relationship (edges) as well as the analytical relationship (conditional probabilities) between nodes (i.e. attributes) are automatically computed. They are independent from the person who performed the model.

4.2.2 Evaluation of our model functioning

Table 4. *Evaluation of the model interaction with the user*

Criteria	Evaluation
<p>Attractiveness: how attractive the model may be to the user (the use of color, the nature of the graphical design etc.) This criterion is also related to the previous criteria (i.e. <i>consistency, self-descriptiveness</i> and <i>independence</i>).</p>	<p>Graphical formalism in general and Bayesian Network particularly are more expressive and thereby more attractive compared to other representation formalisms. Moreover, the graphical tool we used (BayesiaLab®) allows distinguishing graphically in the user interface the technical and perceptual attributes (see icon in Figure 4). Causal relationships between attributes are also easy to represent and to interpret. This makes the model easy to understand and to use even by a non expert user.</p>
<p>Reusability: the efficiency of a model in facilitating a selective use of its components or sub-models</p>	<p>A BN model can be considered as a combination of sub-models. Indeed, each group of nodes can be analyzed separately as a single model interacting with the other nodes. The tool we used allows carrying out selective analysis through a monitor (see Figure 4 and Figure 5). Hence, an user can only focus on a given sub-set of the model nodes.</p>
<p>Usability: how the model allows the user to learn in order to operate, prepare the model inputs, and interpret its outputs.</p>	<p>The construction of our kansei BN model can be seen as a classical task of extraction information from data automatically. In other words, it is a machine learning task. It is a well structured task in artificial intelligence consisting of a sequence of sub-tasks: <i>data preparation, model learning</i> and <i>model interpretation</i>. When a user constructs such model, he learns at the same time how to perform each of these sub-tasks.</p>
<p>Learnability: how the model itself helps the user learn more on its application.</p>	
<p>Abstractness: how a model allows a user to perform only the necessary functions relevant to a particular purpose.</p>	<p>As we noticed in section 4.1.3, thanks to BN approach, the same model allows representing at the same time the relationship within technical attributes, the relationship within perceptual attributes and the relationship between technical and perceptual attributes. Thereby, it allows performing at least two several use scenarios: analysis and synthesis, the two main tasks in kansei engineering.</p>
<p>Understandability: how the model permits the user to understand whether the model is suitable, and how it can be used for particular tasks and conditions of use.</p>	<p>As we noticed in the previous point, a BN model embeds the several inter-attributes relationship as well as the several use scenarios required in kansei engineering. It is sufficiently intuitive to allow a given user, even he is not expert in machine learning or BN approach or kansei engineering to understanding and use such model.</p>
<p>Operability: how the model allows the user to operate and control it.</p>	<p>The use of BN approach allows a user easily performing the several tasks of kansei engineering (see 4.1.3). It also allows an easy update as well as incoherence overtaking (see section 4.1.4).</p>
<p>Adaptability: the ease with which the model meets contradictory users' constraints and users' needs.</p>	<p>A same BN model allows performing several kansei tasks. It also allows a user to integrate his own knowledge.</p>

Table 5. Evaluation of the model behavior under normal conditions

Criteria	Evaluation
The controllability : how efficiently the model reacts differentially to the different actions it is submitted to.	A same BN model allows performing several kansei tasks (analysis scenario and synthesis scenario) (see 4.1.3)
The repeatability : how the model generates the same results under the same functioning conditions.	The structural relationship (edges) as well as the analytical relationship (conditional probability) between attributes is automatically computed. Thereby, the result is repeatable.
The generality : how the model performs a broad range of functions.	A same BN model allows performing several kansei tasks (analysis scenario and synthesis scenario) (see 4.1.3)
The interoperability : the ability of two or more models or model components to exchange information and to use the information exchanged.	Probabilistic inference in BN allows taking any state attribute observation (an event) into account so as to update the probabilities of the other attributes. Without any event observation, the computation is based on a priori probabilities. When observations are given, this knowledge is integrated into the network and all the probabilities are updated accordingly (see section 4.1.3)
The replaceability : how the model can be used instead of another specified model for the same purpose in the same environment.	Neither of the other models we constructed using the same data but other data mining techniques (Logistic Regression, Principal Component Analysis, etc.) allows performing at the same time the use scenario and the synthesis scenario. A BN model can replace all of these models.
The usability compliance : how the model can comply with standards, conventions, style guides or regulations relating to usability.	Not usable for our case

Table 6. Evaluation of the model behavior under stressful conditions

Criteria	Evaluation
<p>Error tolerance: the ability of the model to continue an operation normally despite the presence of erroneous inputs.</p>	<p>As we noticed in section 4.1.4, a BN model allows a user integrating his own knowledge as well as knowledge embedded in new data. Hence, if there is erroneous input data, a BN model allows overtaking structural and analytical incoherencies.</p>
<p>Fault tolerance: the ability of a model to continue an operation normally despite the presence of model component faults.</p>	<p>As we noticed previously, a BN model can be seen as a combination of sub-models. Thereby, even though there are incoherencies in some relationship between a set of attributes, the other relationship can be used normally. In other words, if an inference does not include a fault part of the model, we can use its result without risk.</p>
<p>Error proneness: the ability of a model to allow the user to intentionally or unintentionally introduce errors into the model or misuse the model.</p>	<p>User can easily introduce errors into the model: he can add, remove or change the orientation of an edge. He also can modify conditional probabilities characterizing a relationship between nodes.</p>

4.2.3 Evaluation of our model evolution

Table 7. Evaluation of the model evolution

Criteria	Evaluation
<p>Flexibility: how easily modifications can be carried out in order to use the model in applications or environments other than those for which it has been specifically designed.</p>	<ul style="list-style-type: none"> - The same Bayesian model may be used as a cause analysis tool. Suppose a new dashboard has negative customer appreciation. Our model can be used to attempts to identify the cause of this negative evaluation provided that the present design context is compatible with the BN hypotheses such as the car segment market and the customer segment. The synthesis scenario is then used. - The same model can be used to carry out supervised and unsupervised learning to carry out local or global optimization of a given design (see [3]).
<p>Extendibility (or Expandability): how easily modifications can be performed in order to increase the model functional capacity</p>	<p>In our case, increasing the model functional capacity consists in increasing its precision and accuracy. The fact that we can easily introduce expert knowledge as well as knowledge embedded in new data (see section 3.3) improves the scope of our model.</p>
<p>Maintainability: how easily modifications can be carried out in order to correct model faults.</p>	<p>As we noticed in section 4.1.4, a BN model allows user integrating easily his own knowledge as well as knowledge embedded in new data in order to handle structural and analytical incoherencies.</p>
<p>Testability: how easily modifications can be performed within the validation stage of the model.</p>	<p>See previous point.</p>

4.2.4 Evaluation of our model teleology

Table 8. *Evaluating of the model teleology*

Criteria	Evaluation
Accuracy/Precision: how well the model provides the right or agreed results or effects with the expected degree of accuracy.	In [3] we showed how to assess the accuracy of supervised and unsupervised Bayesian models and to appropriately choose between them depending on the design goal.
Efficiency: how well the model provides an appropriate performance, relatively to the amount of resources used (time, human resources, etc.), under stated conditions.	
Effectiveness: the ability of the model to target all aspects of the goal.	

5 Interrelationships between evaluation criteria

5.1 Introduction to the selection of criteria

The main technical issue that this work faced was related to the criteria identification. Coming from various fields such as education, policy making, information theory, economy, philosophy, the criteria evolved with the progress in understanding the processes. In most cases these criteria appear as single to undertake the assessment of a specific character. Based on the work of Reich (see [33, 34]) and the cybernetic of the second order we have considered all the potential scientific fields that have explicitly addressed the evaluation theory and methodology, and their associated criteria. We have then suggested consensus on their definitions based on the work of Heylighen (see [20, 21]).

5.2 Interrelationships between model evaluation criteria and knowledge evaluation criteria

Since a model does not constitute an objective in itself, but is a means to create new knowledge, a satisfactory model must be the one which allows deriving adequate knowledge in given contexts. In other words, the model evaluation criteria must fulfil the knowledge evaluation criteria (see section 2.4 and figure 1).

The question of links between the two types of criteria sets is worth studying. We propose, in this paper, a first suggestion of such links, based on our experience (example used in this paper and other initiatives). Table 9 is the result of a first generic correlation that can exist between the two sets of evaluation criteria.

Table 9 may be interpreted in both directions. In the vertical direction, let us take the example of the criterion presented in the first column, i.e. ‘*knowledge invariance*’: to improve this criterion (i.e. +), we can improve the ‘*model consistency, self-descriptiveness, independency, etc.*’. We can also weaken the criterion ‘*model completeness*’. In the horizontal direction, let us take the example of the criterion presented in the third row, i.e. ‘*model ontology independency*’: the improvement of this criterion (i.e. +) may lead to the improvement of the ‘*knowledge invariance, simplicity and consistency*’ and/or the degradation (i.e. –) of ‘*the knowledge distinctiveness, controllability and formality*’.

Only the approach related to the relationship between knowledge evaluation criteria and model evaluation criteria must be considered here and the reader must not pay too much attention to the table content as it should be confirmed by more model implementations and post-validations. We intend to provide Table 9 to a panel of researchers to figure out whether it is possible and relevant to refine this

general correlation table. But for the time being, we consider this table as architecture to adapt (a pattern to instantiate) to any domain of application.

Table 9. *Interrelationships within model evaluation criteria. A “+” (resp. “-”) in case (i,j) means that model criterion i is positively (resp. negatively) correlated to the improvement of model criterion j.*

CRITERIA				Knowledge evaluation criteria												
				Objective criteria			Subjective criteria				Inter-subjective criteria					
				Invariance	Distinctiveness	Controllability	Individual utility	Simplicity	Consistency	Novelty	Publicity	Expressivity	Formality	Collective utility	Authority	
Model evaluation Criteria	Evaluating the model ontology	Representation formalism	Consistency	+				+	+		+	+	+	+		
			Self-Descriptiveness	+	+			+			+	+				
			Independency	+	-	-	-	+	+		+	+	-	+		-
			Completeness	-	+	+	+	-	-			+	-			
		Model-user Interaction	Attrativeness				+					+			+	
			Usability	+			+					+		+	+	
			Reusability	+			+					+		+	+	
			Understandability	+			+	+				+		+	+	
			Learnability	+			+	+				+		+	+	
			Operability	+			+	+				+		+	+	
	Adaptability		-	-	-	+				+		+	+			
	Evaluating the model functioning	Stressful conditions	Error Tolerance	+	-	-	+							+	+	
			Fault Tolerance	+	-	-	+							+	+	
			Reliability	+	+	+	+								+	
			Error propeness	+	-	-	+							+	+	
		Normal conditions	Controlability	+	+	+	+									
			Geanarlity		-	-	-	+				+		+	+	
			Replaceability		-	-	-	+							+	
	Evaluating the model evolution	Modifiability	Flexibility		-	-	-						+		+	+
			Maintainability													+
			Testability	+			+	+				+		+	+	
			Extendibility		-	-	-					+		+	+	
			Stability	+	+	+	+									+
	Evaluating the model teleology	Accuracy	-	+	+	+		-			+		-			
		Efficiency				+					+				+	
		Effectiveness				+					+				+	

5.3 Interrelationships within model evaluation criteria

We have so far considered a complete independence between the evaluation criteria of a model. But, in practice, the levels of compliance to the criteria turn out to be correlated. Again, we have found no

existing study on that subject in the literature. We propose in table 10 the generic correlation matrix between the model evaluation criteria filled by the knowledge gathered during this experiment. This result has to be considered as a framework and should not be adopted without an extensive validation. Table 10 may be interpreted in its two directions. Vertically: let us take the example of the criterion presented in the 18th column, i.e. ‘*model flexibility*’: to improve this criterion (i.e. +), we can improve ‘*model consistency, independency, etc.*’. We can also weaken (i.e. -) ‘*model completeness*’. Horizontally: let us take the same example of the criterion ‘*model flexibility*’: the improvement of this criterion (i.e. +) may lead to the improvement of ‘*model attractiveness, reusability, etc.*’. This may also lead to the weakening (i.e. -) of ‘*model controllability, model precision, etc.*’. We notice here (as in the previous section) that only the approach related to the relationship within model evaluation criteria, must here be considered and the reader must not pay too much attention to the table content as this content should be confirmed by more model implementations and post-validations.

Table 10. *Interrelationships within model evaluation criteria. A “+” (resp. “-”) in case (i,j) means that model criterion i is positively (resp. negatively) correlated to the improvement of model criterion j.*

CRITERIA		Model Evaluation Criteria																								
		Evaluating the model ontology										Evaluating the model functioning					Evaluating the model evolution			Evaluating the model teleology						
		Representation formalism				Model-User interaction						Stressful conditions			Normal conditions		Modifiability									
		Consistency	Self-descriptiveness	Independency	Completeness	Attractiveness	Usability	Reusability	Understandability	Learnability	Operability	Adaptability	Error Tolerance	Fault Tolerance	Error propeness	Controllability	Generality	Replaceability	Flexibility	Maintainability	Testability	Extendibility	Stability	Accuracy	Efficiency	Effectiveness
Evaluating the model ontology	Representation formalism	Consistency	+	+		+	+	+	+	+	+															
		Self-descriptiveness		+			+	+	+	+	+															
		Independency		+	-		+	+	+	+	+															
		Completeness			-	+	+	+	+	+	+															
	Model-User interaction	Attractiveness				+	+	+	+	+																
		Usability					+			+																
		Reusability					+	+			+															
		Understandability					+	+	+	+	+															
		Learnability					+	+	+	+																
		Operability					+	+	+	+	+															
Evaluating the model functioning	Stressful conditions	Adaptability				+	+	+		+		+	+	+												
		Error Tolerance					+	+	+	+	+															
		Fault Tolerance					+	+	+	+	+															
	Normal conditions	Error propeness					+	+	+	+	+															
		Controllability																								
		Generality																								
		Replaceability																								
		Flexibility																								
		Maintainability																								
		Stability																								
Evaluating the model evolution	Modifiability	Accuracy																								
		Efficiency																								
		Effectiveness																								
		Consistency																								
		Self-descriptiveness																								

6 Discussion of the approach

There are several potential approaches to the representation of the perceived world. Modelling is a natural human process that started to be studied since the Greek civilization. The understanding of the explicit and implicit behaviour of the “modeller” has been influenced by most of the school of thoughts in philosophy. It is too early to state that a Cartesian ontological description of the world is obsolete. However, there is a consensus in the scientific community for a need to describe the component of the perceived real or artificial world in term of its components and its behaviour or

functionality. There are much more doubt and critics against the need to describe the teleology of a system.

The approach used here in order to set the list of criteria to be considered is based on two stages. A top-down perspective based on the general system theory that forces the consideration of the four levels of description; and the bottom-up approach based on a deep analysis of the criteria used in several disciplines. We provide here this classification.

The main drawback of the approach used belongs to the intrinsic characteristic of the approach dealing with the concept of recursively. In fact, while at the epistemological level it leaves the door open for a refinement of the description, at the same time it closes the door for a perfect control of the system behaviour and thus lead to a risk of incompleteness.

Nevertheless, when applied in the design field (presented here) and in more areas since the beginning of the ninetens (Knowledge management [5], Design process documentation [10, 29], and industrial maintenance [2]) it provided us with a real new approach leading to a more mature description of the systems under studies.

Nevertheless, from the application perspective, one has to consider that all the criteria might not be considered at the same time. It is best suitable to introduce some weighting or hierarchisation according to the modelling objectives. The use of house of quality method or QFD is recommended.

7 Conclusion

Is my model of the *real world* or my model of an *artificial world* a satisfactory *model*? Here is the question that a biologist⁴ or, an industrial engineer⁵ could ask when confronted to a modelling process aiming at generating the necessary *knowledge* that could result in the best set of actions in a given *context*.

This paper has adopted an *evolutionary-cybernetic epistemology* to state that the *model assessment criteria* may also derive from the assessment criteria of the generated *knowledge*. This paper has also adopted a systemic approach in systematically considering four viewpoints in the evaluation process, namely: *ontology, functioning, evolution/transformation, teleology*.

A generic *model of a model evaluation* has been defined through the proposal of 28 model evaluation criteria and 12 knowledge evaluation criteria. We have been using this approach in several case studies and presented a specific case in this paper.

In addition, we have proposed two correlation tables between evaluation criteria that should help the modeller to better characterize his/her application domain in terms of expected modelling difficulties.

We hope that this *model of a model evaluation* will bring a valuable aid to modellers in the future. The matrix presented might be extended to include any missed criteria. The ultimate question could then be "Is our model satisfactory?".

8 References

- [1.] Ashby. W.R. **An introduction to cybernetics**, ed. Hall, C., London, (1965).
- [2.] Baud.N, Mekhilef. M & Bocquet. J-C. **Proposal of a Functional Model of Logistics for Spare Parts Preservation**. Integrated Design Manufacturing in mechanical Engineering, Kluwer Academic Publishers. May, pp 561-568. (1999).
- [3.] Ben Ahmed W. & Yannou B., **A bayesian learning of probabilistic relations between perceptual attributes and technical characteristics of car dashboards to construct a perceptual evaluation model**, International Journal of Product Development, under press. (2008).

⁴ relatively to a model of bacteria

⁵ relatively to a model of a production system or of a product system (*digital mockup*)

Ben Ahmed W., Mekhilef M., Yannou B., Bigand M., (2010), *Evaluation Framework for the Design of an Engineering Model*. AI EDAM: Artificial Intelligence for Engineering Design, Analysis and Manufacturing, vol. 24(1): p. 107-125, doi:10.1017/S0890060409000171.

- [4.] Ben Ahmed. W, Mekhilef. M, Bigand. M & Page. Y. **A Communication Tool between Designers and Accidentologists for the Development of Safety Systems**. Galsgow, UK, (2005).
- [5.] Ben Ahmed. W, Mekhilef. M, Bigand. M & Page. Y. **Intégration des connaissances du domaine pour la fouille de données complexes**. Extraction et Gestion de Connaissances, Université Blaise Pascal, Clermont Ferrand, France, (2003).
- [6.] SAFE-NEXT: a Systemic Approach For Knowledge Discovery In Databases. Application In Accident Scenario Development And Interpretation, Dissertation Thesis, Ecole Centrale Paris, 2005
- [7.] Bertalanffy. L. V. **General system theory: foundations, development, applications**, George Braziller, New York, (1969).
- [8.] Booch G., Rumbaugh J., Jacobson I. **Unified Modelling Language User Guide**, Addison Wesley Professional, (1999).
- [9.] Campbell. D. T, **Evolutionary Epistemology**. in Schilpp P.A. (ed.), The Philosophy of Karl Popper., p. 413-463, Open Court Publish., La Salle, Ill, (1974).
- [10.] Cantzler. O, Mekhilef. M & Bocquet.J-C. **A Systemic Approach to Corporate Knowledge: An Ontology to Modeling in a Design Department**. IEEE Transaction on Systems and Human. pp 153-158, Vancouver, Canada. (1995).
- [11.] Chovan. J.D, Tijerina. L, Everson. J. H, Pierowicz. J. A. & Hendricks. D. L, **Examination of Intersection, Left Turn Across Path Crashes and Potential IVHS Countermeasures**, Rapport N°: DOT HS 808 154, National Highway Traffic Safety Administration, (1994).
- [12.] Condillac.E.B, **Commerce and government considered in their mutual relationship**, Cheltenham, UK; Northampton, Mass., USA: E. Elgar, (1776).
- [13.] Fleury. D., Fline. C & Peytavin.J.F **Diagnostic local de sécurité, outils et méthodes**, Editions SETRA, Collection Etudes de sécurité, Bagneux, (1991).
- [14.] Frey.D.D, and Dym.C.L, **Validation of design methods: lessons from medicine**, Research in Engineering Design, 17:45-57. (2006).
- [15.] Fuller. R, **The Task-Capability Interface Model of The Driving Process**. RTS, Recherche Transports Sécurité, vol. 66, Tome 1, p. 35-45, (2000).
- [16.] Fuller.R & Santos. J. A. **Human Factors For High-way Engineers**, Elsevier, Pergamon, 2002.
- [17.] Harvey. A. **Application of an integrated method to a study of the consumer perceptions of automobile dashboards**, Research Internship report in Ecole Centrale Paris, University of Bath. (2005)
- [18.] Henry. G. T & Mark. M. M. **Beyond use: Understanding assessment's influence on attitudes and actions**. American Journal of Evaluation, vol. 24(3), p. 293-314, (2003).
- [19.] Henry. G.T. **Influential Evaluation**. American Journal of Evaluation, vol. 24(4), p. 515-524, (2003).
- [20.] Heylighen. F. **Objective, subjective and intersubjective selectors of knowledge**. Evolution and Cognition, vol. 3(1), p. 63-67, (1997).
- [21.] Heylighen. F, **Selection Criteria for the Evolution of Knowledge**. in: Proc. 13th Int. Congress on Cybernetics, Association Internationale de Cybernétique, Namur, (1993).
- [22.] Huang. C, A. Dawiche, **Inference in Belief Networks: A Procedural Guide**. International Journal of Approximate Reasoning, 15, p225-263. (1996).
- [23.] Jensen F.V. **An Introduction to Bayesian Networks**. (UCL Press (Ed)), London. (1996).
- [24.] Kirkhart. K. **Reconceptualizing evaluation use: An integrated theory of influence**. in V. Caracelli & H. Preskill (Eds.) The expanding scope of evaluation use. New directions for evaluation, p. 5-24, San Fransisco: Jossey-Bass, (2000).

Ben Ahmed W., Mekhilef M., Yannou B., Bigand M., (2010), *Evaluation Framework for the Design of an Engineering Model*. AI EDAM: Artificial Intelligence for Engineering Design, Analysis and Manufacturing, vol. 24(1): p. 107-125, doi:10.1017/S0890060409000171.

- [25.] Lam, W. & Bacchus. F. **Learning Bayesian belief networks: An approach based on the MDL principle**, Computational Intelligence, Vol. 10, pp.269-293. (1994)
- [26.] Le Moigne. J-L, **La modélisation des systèmes complexes**, Dunod, (1999).
- [27.] Limayem F., Yannou B. **Generalization of the RCGM and LSLR Pairwise Comparison Methods**. Computers and Mathematics with Applications, vol. 48: p. 539-548. (2004).
- [28.] MacQueen. J. B. **Some methods for classification and analysis of multivariate observations**. Proc. Symp. Math. Stat. And probability (5th), Univ. Of California, Berkeley, p 281-297, (1967).
- [29.] Mekhilef. M, Bocquet. J.C, Cantzler. O & Gallardo.J-F. **Towards a Conceptual Architecture for the Capitalization of Design Process**. Integrated Design and Manufacturing for Mechanical Engineering, IDMME. Mai 1998, pp 1037-1044, Compiègne France. (1998).
- [30.] Nagamachi, M. **Kansei Engineering: The Framework and Methods**. Kansei Engineering 1, Nagamachi, M. (ed.), Kaibundo Publishing Co. Ltd., Kure, pp. 1-9. (1997).
- [31.] Najm. W. G, Smith. J. D & Smith. D. L. **Analysis of Crossing Path Crashes**, Rapport N° DOT HS 809 423, National Highway Traffic Safety Administration, (2001).
- [32.] Petiot J.-F, Yannou. B. **Measuring consumer perceptions for a better comprehension, specification and assessment of product semantics**. International Journal of Industrial Ergonomics, vol. 33(6): p. 507-525. (2004).
- [33.] Reich. Y. **Layered models of research methodologies**, Artificial Intelligence for Engineering Design, Analysis, and Manufacturing, 8(4):263-274, (1994).
- [34.] Reich. Y. **Measuring the value of knowledge**, International Journal of Human-Computer Studies, 42(1), pp. 3-30, (1995).
- [35.] Schreiber. G, Wielinga. B, Akkermans. H, Van de Velde. W & Anjewierden. A. **CML: The CommonKADSCONCEPTUAL MODELING LANGUAGE**. 8th European Knowledge Acquisition Workshop EKAW'94, Allemagne, p 283-300, (1994).
- [36.] Schütte, S. **Engineering emotional values in product design: Kansei Engineering in development**, Dissertaion 951, Linköping Universitet Institute of Technology. (2005).
- [37.] Sowa. J.F, **Conceptual Structures, information processing in mind and machine**, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, (1984).
- [38.] Subrahmanian. E, Konda. S. L, Levy. S. N, Reich.Y, Westerberg. A. W & Monarch.I.A. **Equations aren't enough: Informal modeling in design**, Artificial Intelligence in Engineering Design, Analysis, and Manufacturing, vol. 7, no. 4, pp. 257-274, (1993).
- [39.] Suh. N. **The Principles of Design**, Oxford University Press, Oxford, (1993).
- [40.] Summala. H. **Automatization, automation, and modeling of driver's behavior**. RTS, Recherche Trans-ports Sécurité, vol. 66, Tome 1, p. 35-45, (2000).
- [41.] Turchin. V. **Cybernetics and Philosophy, in: The Cy-bernetics of Complex Systems - Self-organization, Evolution and Social Change**, F. Geyer (ed.), Inter-systems, Salinas, CA, (1991).
- [42.] Von Foerster. H. **The Cybernetics of Cybernetics** (2nd edition), Future Systems Inc., Minneapolis, (1995).
- [43.] Yannou B., Coatanea E., **Easy and flexible specifications and product evaluations by expert and customer comparisons with existing products**. in International Conference on Engineering Design: ICED'07, August 28-31, Cité des Sciences et de l'Industrie, Paris, France. (2007).
- [44.] Yannou. B. & Petiot J.-F. **A methodology for integrating the customers' assessments during the conceptual design**. In DETC/DTM: ASME Design Engineering Technical Conferences / Design Theories and Methodologies, Sept. 28 - Oct. 2, Salt Lake City, Utah, USA. (2004).

Vitae



Dr Walid BEN AHMED Walid Ben Ahmed received his PhD (2005) in Knowledge Engineering and Management in design from the Industrial Engineering Department at Ecole Centrale Paris (ECP) in 2004. He received his MSc (2001) in Design and Product System Development from Ecole Centrale Paris (ECP).

He received his Degree in Mechanical Engineering in 2000 from the National Engineering School of Tunis (ENIT). He is now an expert on product reliability, in charge of innovation risk management in the Powertrain Engineering Division in RENAULT. His research interests include product reliability analysis, complex system modelling, product modelling, innovation, evaluation, knowledge engineering and data mining in design.



Dr Mounib MEKHILEF is an Associate Professor at the university of Orléans in France. He obtained his Mechanical engineering degree in 1982. In 1991 he defended his Ph.D thesis at the Ecole Centrale Paris. In 2000 he got his habilitation degree to manage research from the University of Nantes. Teaching Modeling techniques and computer aided design, in the mechanical Engineering Department at the University of Orléans. His main research fields are design optimization.



Dr Bernard YANNOU is a Full Professor of Industrial and Mechanical Engineering at the Laboratoire Génie Industriel of Ecole Centrale of Paris, France. He received a M. Sc. (1988) in Mechanical Engineering from Ecole Normale Supérieure of Cachan, and a second M. Sc. (1989) in Computer Science from Paris-6 University. He received a Ph. D. (1994) in Industrial Engineering from Ecole Normale Supérieure of Cachan. His research interests are centered on the preliminary stages of product design: defining the design requirements, synthesizing product concepts, rapid evaluation of product performances, preference aggregation of the product and the project performances for the supervision of the design process.



Dr Michel BIGAND is an Associate Professor of Computer Science and Project Management at Ecole Centrale de Lille (France). He received an MSc in Mechanical Engineering (1980) from Ecole Normale Supérieure de Cachan, a PhD (1988) from Paris 6 University and enabling to manage research since 2005. His research activities are done at the Industrial Engineering Laboratory of Lille (LGIL) and concern the design systems engineering and their associated information systems. More specifically, he works on models integration for knowledge sharing and interoperability.