Can dissonance engineering improve risk analysis of human-machine systems?

F. Vanderhaegen[1,2] and O. Carsten[3]
[1] UVHC, LAMIH, F-59313 Valenciennes, France
[2] CNRS, UMR 8201, F-59313 Valenciennes, France
Email : *frederic.vanderhaegen@univ-valenciennes.fr*

[3] *Institute for Transport Studies, University of Leeds, UK*
Email : O.M.J.Carsten@its.leeds.ac.uk

Abstract. The paper discusses dissonance engineering and its application to risk analysis of human-machine systems. Dissonance engineering relates to sciences and technologies relevant to dissonances, defined as conflicts between knowledge. The richness of the concept of dissonance is illustrated by a taxonomy that covers a variety of cognitive and organisational dissonances based on different conflict modes and baselines of their analysis. Knowledge control is discussed and related to strategies for accepting or rejecting dissonances. This acceptability process can be justified by a risk analysis of dissonances which takes into account their positive and negative impacts and several assessment criteria. A risk analysis method is presented and discussed along with practical examples of application. The paper then provides key points to motivate the development of risk analysis methods dedicated to dissonances in order to identify the balance between the positive and negative impacts and to improve the design and use of future human-machine system by reinforcing knowledge.

1. Introduction

Sometimes, engineers or researchers design walking robots directly with two legs. They do not copy the learning process of human walking that begins initially with the legs and the hands, then with the use of supports and finally with both legs after achieving a fairly complete control of equilibrium. Vanderhaegen (2014) wondered if this design process is wrong because, from a control process viewpoint, undesirable events such as "lack of equilibrium" or "breakdown of equilibrium" should be studied in order to design algorithms or other devices that are able to prevent the walking robots from losing their equilibrium by applying for instance the human learning rules. Conflicts between viewpoints related to the manner of designing a system can then occur. They are called dissonances.

Classical risk analysis focuses on the identification and the control of such undesirable events and aims at providing human-machine systems with barriers to protect them from the occurrence or the impact of these events (Vanderhaegen, 2010). Procedures or technical systems are then designed to prevent the system from incidents or accidents. Despite these barriers, accidents remain, and retrospective analyses can help the designers to identify what went wrong. Here again, dissonances can occur between the accident prevention process and reality.

Some accident analysis has demonstrated that the control of an isolated undesirable event is not sufficient and that the associated possible secondary accident also has to be treated. For instance, in the railway domain, undesirable events such as collision, fire or explosion are treated independently, and procedures or technical barriers are built to manage each event separately. What might be wrong with this normal risk analysis process? Suppose for instance that a collision will occur between a train and a truck at a level crossing and that this collision risk is handled well by the staff who apply the correct procedures to prevent the collision. What may be wrong if nonetheless the collision actually occurs? Possible secondary accidents such as a fire or an explosion have to be managed. Indeed, the shock due to the collision may generate sparks and damage to the train and the truck. Then, a fire may occur because of the presence of these sparks and the leakage of diesel fuel from the damaged tank of the truck. Moreover, if the truck cargo is composed of full gas bottles, an explosion may occur due to the presence of sparks, fire, and gas coming from the damaged gas bottles. Such possible secondary accidents may require the definition of their own specific procedures and technical barriers. The procedures for analysing undesirable events such as accidents have to be extended to possible secondary accidents by taking into account both independent and dependent facts. Retrospective analysis may demonstrate that the initial risk analysis is insufficient and usually the human-machine system is a posteriori adapted to cover the newly identified problems. This is a new kind of dissonance, when conflicts occur between the prospective and the retrospective risk analysis processes.

This risk analysis requires the use of a variety of scientific contributions. There are approaches such as (Vanderhaegen et al., 2004; Vanderhaegen, 2012a):

- RAMS based analyses, i.e., Reliability, Availability, Maintainability and Safety based analyses to treat about technical failures.
- Analyses from cindynics, i.e., analyses to handle organizational dangers.
- Human reliability or human error based analyses, i.e., analyses to handle the success or the failure of human behaviours.
- Resilience or vulnerability based analyses, i.e., analyses to handle the success or failure of the control of system stability.
- Dissonance based analyses, i.e., analyses to handle conflicts between different actors' knowledge.

This paper focuses on the topic of the risk analysis of dissonances and it aims at opening future discussion in Cognition Technology & Work. Section 2 of the paper proposes a definition of the concept of dissonance and of dissonance engineering before presenting some strategies to control dissonances in Section 3. A taxonomy of dissonances based on the conflict principle and on different kinds of analysis baselines is detailed in Section 4. Section 5 gives solutions for analysing risks of dissonances and Section 6 illustrates the value of this risk analysis.

2. From cognitive to organisational dissonance

Dissonance engineering relates to the engineering sciences that help to manage dissonance. It focuses then on the dissonance concept developed by cognitive sciences (Festinger, 1957) and cindynics (Kervern, 1995). It consists in treating such dissonances in a practical way in terms of risks. A cognitive dissonance is defined as an incoherency between cognitions. Cindynics dissonance is a collective or an organizational dissonance related to incoherency between persons or between groups of people. Dissonance engineering is a way to analyse risks by using the concept of dissonances that occur when something sounds wrong, i.e. something will be, is, maybe or was wrong (Vanderhaegen, 2012a, 2014). The occurrence of these dissonances will relate to individual and collective knowledge. We can also think about dissonance between humans and robots. One example here might be dissonance between the human and the highly automated road vehicle.

A dissonant cognition relates to contradictory information and a dissonance produces a discomfort state due to the occurrence of conflicting cognitions. Cognition is a cognitive element of knowledge, or relates to knowledge. For instance, it is a behaviour, an attitude, an idea, a belief, a viewpoint, a competence, etc. Globally, a dissonance is associated with the occurrence of incoherent individual or collective knowledge (Festinger, 1957; Kervern, 1995; Brunel and Gallen, 2011).

Concerning organisational dissonance, incoherency between groups may concern several groups with similar goals or different groups such as a group of designers and a group of users. Table 1 gives an example of such organisational dissonances between designers, employers, work teams and users.

Table 1. Example of organisational dissonances in risk analysis

Actors	Designers / Employers	Users / Teams
Objective	Safety evaluation	Multi-risk control
Criterion	Mono-criterion	Multi-criterion
Processing	Off-line	On-line
Validation	Static	Dynamic
Integrity	Homogeneous	Variable
Source	External	Cognitive
Output	Barriers System reorganisation Training programme Users manual	Personal barriers Learning process Experience Violations

Such incoherency can also be a factor at a societal level. Inherent in "Vision Zero" for road safety — the vision that no road user should be killed or seriously injured — is the principle of rule compliance: "Road users are responsible for following the rules for using the road transport system set by the system designers" (Tingvall and Haworth, 1999). Thus the safety philosophy of Vision Zero and subsequently of the Safe Systems approach (OECD, 2008) is that while the designers of the traffic system have the ultimate responsibility of providing the means to safe operation, there is a counterpart responsibility on the users of the system. There is then a kind of social contract between designers and users to deliver overall safe design and operation.

The design of a complex human-machine system requires the application of adequate norms for safety conformity such as those presented in Rouhiainen and Gunnerhed (2002). Therefore, risk analysis concerns mainly an off-line safety evaluation. When the result of this analysis is certified, its validation remains static: there is no recovery of this analysis except in case of an accident or a safety-critical event such as a near-accident. Its integrity is considered as permanent. The source of this validation is made externally, i.e. independently from the viewpoint of future potential users. Residual acceptable risk is achieved after evaluation of system organisation, the design of barriers, the proposal of a specific training program, the production of user manuals, etc. The classical steps in risk control are: risk analysis, risk evaluation and risk reduction and control.

From the point of view of the work teams or the users, the risk analysis process can be rather different. It concerns an on-line and cognitive multi-risk control considering several evaluation criteria such as safety, production, quality or workload. Its validation can evolve with regard to the human operators' or the teams' state. It is then dynamic and its integrity is variable. The identification of residual acceptable risk leads human operators to create personal barriers, to increase their knowledge and refine their experience through confronting unanticipated events for example, and to make violations of rules or technical barriers in order to solve exceptional situations or new problems.

The main goal of the designers or employers is to make the residual risks acceptable, whereas the work teams or the users have to control them when their associated events occur. Table 2 summarizes such a residual risk management process. The designers or the employers modify the structure of the proposed human-machine systems in order to make residual risk acceptable. They provide the human-machine system with barriers that protect it from the occurrence or the consequences of undesirable events (Vanderhaegen, 2010). This makes the residual risks acceptable. However, when the events associated with these acceptable residual risks occur, the work teams or the users have to manage them whatever their probability of occurrence or their consequences. Those residual risks that were considered acceptable by the designers or the employers can become unacceptable to the users or the teams, because no barriers were planned to manage such risks or because they did not receive adapted training to control them. Then, dissonances occur when there is discord, and the users or the teams may be obliged to create procedures to solve these new situations. When the management of these situations is successful, this behaviour makes the system resilient. When it fails, it makes the system vulnerable. Resilience engineering relates then to the concept of dissonances, when dissonances are successfully treated or when they do not affect the system safety.

Moreover, the period of use of a system can modify the frequency of occurrence of undesirable events. Indeed, the occurrence of an event that was considered as incredible at the beginning of the use of a given human-machine system can become probable after several years of operation. This transformation is not controlled and can lead to hazardous dissonances.

Table 2. Residual risk management process

▨ Residual risks management and control processes allocated to users: safety management

⬚ Risk analysis and reduction processes allocated to designers: safety analysis

		Consequences of hazardous events			
		Catastrophic	Critical	Significant	Minor
Frequency of hazardous events	Incredible	Acceptable	Acceptable	Acceptable	Acceptable
	Improbable	Unacceptable	Undesirable	Acceptable	Acceptable
	Rare	Unacceptable	Unacceptable	Undesirable	Acceptable
	Probable	Unacceptable	Unacceptable	Unacceptable	Undesirable

Dissonances occur when there are discordances or divergences between groups of persons such as the designers, the employers, the work teams and the users. They can also appear when something is wrong for a given person when this person has to face an event for which there is no adapted prescription, and this person has to create new knowledge. Several tests of this new knowledge on the human-machine system aim at refining it until it is considered as optimal. This new knowledge can then be satisfactory for a user or a group of users, whereas it can be unacceptable for other persons or groups of persons. The creation or the refining of knowledge is called the knowledge reinforcement process and it can generate other possible dissonances. Other reasons can explain the occurrence of a dissonance and several control strategies are possible.

3. Dissonance control and knowledge reinforcement

The causes of cognitive or organisational dissonance are multiple. Dissonances can be due to the occurrence of important or difficult decisions involving the evaluation of several possible alternatives (Chen, 2011). They can also occur when viewpoints on human behaviours are contradictory (Polet et al., 2003) or when behaviours such as competitive or cooperative ones fail (Vanderhaegen et al., 2006; Vanderhaegen, 2012b). Organisational changes that produce incompatible information are possible sources of dissonance occurrence (Telci et al., 2011; Brunel and Gallen, 2011). Between human and machine, they can occur when there is a lack of transparency, i.e. when the machine does not understand the intention of the human or the human does not sufficiently understand the intention or strategy of the machine, leading to a failure of the joint cognitive system (Lyons, 2013; Hollnagel and Woods, 2005). Then, the updating or the refining of a given cognition due to new feedback from field can also generate dissonance.

Whatever the causes of the dissonance occurrence, several paradigms exist. Human operators aim at reducing any occurrence or the impact of a dissonance because it produces discomfort. This activity leads to maintaining a stable state of knowledge without producing any effort to change it (Festinger, 1954). Despite this reduction, a breakdown of this stability is sometimes useful in order to facilitate the learning process and refine, verify or confirm knowledge (Aïmeur, 1998). Such knowledge adjustment improves learning abilities. Finally, dissonance can also be seen as a feedback of a decision: dissonance occurs after a decision and this requires a modification of knowledge (Telci et al., 2011).

Therefore, a discomfort can be a dissonance or can be due to the production of a dissonance, and the detection or the treatment of a dissonance can also produce discomfort. Discomfort can occur if this dissonance is out of control of the human operators or because the treatment of a detected dissonance increases human workload or human error (Vanderhaegen, 1999a). Such an activity involves a minimum learning process in order to improve human knowledge and to control such a discomfort. There are then positive and negative feedbacks from dissonance management. Negative feedbacks relate to discomfort and positive ones to the learning aspect.

The more difficult is the learning process to handle a dissonance, the less acceptable is this dissonance (Festinger, 1954). Therefore, strategies for dissonance reduction are required in order to minimise knowledge changes or to facilitate the learning process and manage the acceptability of a dissonance. Typical strategies, adapted from Festinger (1954) and extended in order to take into account the learning process to reinforce knowledge, are:
-   The *elimination* or the *inhibition* of the dissonance impact by maintaining the initial knowledge without looking for any explanation. There is no modification of current knowledge and the data from the dissonance are disapproved and not handled. This consists in reinforcing the current content of knowledge independently from the dissonance impact.
-   The *addition* of new cognitions to limit the dissonance impact and justify the initial knowledge. This new knowledge consists in giving more importance to the current knowledge than to the knowledge coming from the dissonance. This consists in producing new rules that reinforce the current content of knowledge.
-   The *attenuation* of the dissonance impact by modifying or reinterpreting knowledge. The knowledge coming from the dissonance is integrated into the current knowledge by degrading its importance. New rules related to this dissonance are then produced but they aim at reinforcing the current content of knowledge.
-   The *integration* of the dissonance impact into the knowledge by refining the current knowledge or by creating new knowledge. This can cancel or refine some knowledge and produce new knowledge. This process is another kind of reinforcement of knowledge that handles the current content of knowledge by integrating rules associated to the controlled dissonance.

For example, regarding the use of an industrial rotary press described in Polet et al. (2003), suppose that the initial knowledge of user A includes the following fact without any explanation: "I intervene on the machine even if the machine is running at high speed". Another user or the designer B of this machine can generate a dissonance by saying to him/her: "Any interaction with the machine is very dangerous when the machine is running". From the first user A, the inhibition-based behaviour consists in producing no new knowledge but in ignoring or rejecting the new incoming dissonant knowledge: "No, it is not proved". The addition-based behaviour consists in attenuating the impact of this dissonance and in

justifying the initial knowledge by producing knowledge such as: "It is true but I like taking risks". The attenuation based behaviour consists in modifying the content of the new incoming knowledge to limit its impact: "There is one chance in a billion to have an accident when interacting with the running machine". Finally, the last behaviour consists in reconfiguring the initial knowledge and changing it radically by creating an opposite knowledge: "I stop interacting with the machine when the machine is running at high speed".

The reduction process of a dissonance implies the reinforcement of knowledge. It can be realized by specific algorithms such as those developed in Vanderhaegen et al. (2009), Vanderhaegen et al. (2011), Polet et al. (2012), and Ouedraogo et al., (2013). A trial-and-error process is applied when no knowledge is available to treat a given dissonance. Therefore, the human operators act on the process and wait for the consequences of these actions until they find a solution (Vanderhaegen and Caulier, 2011). This aims at refining the existing knowledge or at creating new knowledge. These reinforcement strategies aim at making the knowledge evolve when a dissonance is treated. Then, this knowledge reinforcement to reduce dissonance leads to maintaining a stable level of knowledge or aims at transforming an unstable level towards a stable level of knowledge. It aims at consolidating, validating, refining or deleting the existing knowledge or at creating new knowledge.

A dissonance may perturb the stability of a knowledge level by affecting other dissonance dimensions such as the interpreted risk level, and its management aims at returning to a new level of knowledge stability or to the previous one by reinforcing knowledge. The maintenance of the coherence of cognitive systems requires stability (Festinger, 1954). The control of this stability can be facilitated by good management of human workload and performance, integrating different human-machine organisations (Vanderhaegen, 1999b). This aims at reducing the occurrence or the impact of a dissonance. For instance, the control of overloaded situations reduces the occurrence of human errors when tasks are dynamically shared between human and machine (Vanderhaegen, 1999c). Knowledge stability relates to sustainable knowledge equilibrium and any deviation from this stability generates dissonances, or is generated by the occurrence of a dissonance or by the impact of its control.

Facing instability of the human knowledge, if the treatment of this dissonance is successful, human operators contribute to the resilience of the system they control. On the other hand, if this treatment produces serial other dissonances and may fail, then it contributes to the vulnerability of the controlled system. The frequency of perturbations such as dissonances may have an impact of the system resilience or vulnerability (Westrum, 2006; Zieba et al., 2010). The management of a regular dissonance increases knowledge about it and may converge to a high stable knowledge level whereas a new dissonance can provoke instability that needs to modify, refine or create knowledge. The lower the frequency of a dissonance, the smaller is the associated knowledge to manage it and the higher is the discomfort or workload this dissonance may produce.

Dissonance engineering methods are required in order to analyse such dissonances and reduce their possible negative impacts. The next section proposes a taxonomy of dissonances based on the sources of conflicts and on the baselines of prescription.

4. Taxonomy of dissonances

Any breakdown of stability of the human-machine system functioning may lead to the occurrence of dissonance. Table 3 proposes a taxonomy of dissonance based on different types of instability identified as conflicts.

Table 3. Dissonance taxonomy and conflict sources

Baseline	Dissonance	Principle	Examples of references
No baseline of prescription	Lack of autonomy	Lack of knowledge	[Vanderhaegen, Caulier, 11]
Erroneous baseline of prescription	Serendipity	Conflict of goal	[McCay-Peet, 15]
One baseline of prescription	Tunnelling effect	Cognitive blindness	[Dehais et al., 12]
	Erroneous cooperation	Conflict of allocation	[Vanderhaegen, 99b; Zieba et al., 11]
	Organizational change	Conflict of information	[Brunel, Gallen, 11; Telci et al., 11]
	Stability breakdown	Conflict of learning	[Aïmeur, 98]
	Emotional dissonance	Conflict of emotion	[Yozgat et al., 2012]
Several baselines of prescription	Automation surprise	Conflict of intention	[Rushby, 02; Inagaki, 08]
	Difficult decisional compromise	Conflict between alternatives	[Chen et al., 14; Ben Yahia et al., 15]
	Barrier removal	Conflict between viewpoints	[Polet et al., 03; Vanderhaegen, 10]
	Social dissonance	Conflict between designers and users	[Tingvall and Haworth, 1999]
	Competition	Conflict of interest	[Vanderhaegen et al., 06]
	Anamorphosis	Conflict of perception	[Dali, 75; Massironi, Savardi, 91]
	Dispositional, epistemic and ontological dissonance	Conflict of sensemaking	[Hunter, Summerton, 06]
	Affordance	Conflict of use	[Gibson, 86 ; Zieba et al., 10]

Dissonances relate to different sources of conflicts and to different baselines of prescription. A baseline of prescription is what the system is supposed to do or behave or believe for instance. No baseline, an erroneous baseline, one baseline or several baselines can exist (Vanderhaegen, 2016). Usually, an error is a conflict between what the system does and what it is supposed to do. The dissonance concept aims at extending such a limited view of error by considering several kinds of prescription to identify conflicts. Therefore, conflicts exist when the system faces a situation for which there is no baseline, or the initial baseline is incorrect, or relates to a single baseline or to several baselines.

The concept of knowledge discovery can be adapted to dissonance discovery for conflict identification (Vanderhaegen, 2016). A lack of autonomy, and more precisely a lack of knowledge is a typical dissonance discovery due to the inexistence of baseline. Thus, the system has to apply trial-and-error and wait-and-see based behaviours to solve the new problem (Vanderhaegen and Caulier, 2011). Serendipity is a conflict of goal that relates to unexpected discovery that demonstrates that the initial baseline is wrong (McCay-Peet, 2015): what is obtained has nothing to do with what it was expected. Cognitive blindness such as perseveration or the tunnelling effect is a conflict of perception when human experts with high levels of knowledge do not hear alarms even though the latter are functioning correctly (Dehais et al., 2012). Erroneous cooperation is a dissonance due to an error of task allocation (Vanderhaegen, 1999a; Zieba et al., 2011). Another kind of

dissonance relates to inconsistency between rules, data, beliefs, intentions, perceptions, interpretations or decisions for instance, due to organisational change (Brunel, Gallen, 2011; Telci et al., 2011). An outcome of the stability breakdown of the learning process can be the reinforcing of the initial knowledge (Aïmeur, 1998). Emotional dissonance occurs when a conflict appears between the self-perceived emotion and the expressed one. Such emotional dissonance can have impacts on human behaviours by affecting emotional exhaustion and job satisfaction (Yozgat et al., 2012).

Automation surprises, difficult decisional compromises between alternatives, or barrier removals are other examples of inconsistency. Automation surprise is a conflict of intention between an automated system and its user (Rushby, 2002; Inagaki, 2008), which can occur as a result of a number of factors, one of which is the lack of transparency. Relaxing safety constraints can lead to the discovery of new alternative action plans (Ben Yahia et al., 2015), or to the discovery of the best compromise between performance criteria (Chen et al., 2014). The discovered alternative generates several baselines of analysis. Barrier removal is an inconsistency between viewpoints on the same situation involving the use of a safety barrier (Polet et al., 2003; Vanderhaegen, 2010). Such conflicts can also be interpreted in terms of social dissonances (Tingvall and Haworth, 1999). Competition relates to conflicts of interest between groups of persons (Vanderhaegen et al., 2006). Anamorphosis consists in having different perceptions of the same object or view (Dali, 1975; Massironi, Savardi, 1991). Then, dispositional dissonance relates to opposite knowledge about the same facts, epistemic dissonance concerns different beliefs about the sources of knowledge, and ontological dissonance is different or opposite meanings of the same knowledge (Hunter, Summerton, 2006). The last example of dissonance concerns the affordances that are based on relations between objects and possible new actions by using these objects (Gibson, 1986; Zieba et al., 2010). Therefore, the dissonance discovery process consists in creating new relationships between objects and actions and this process can concern several groups of users. Conflicts may occur between some of the discovered relationships.

5. Risk analysis process of dissonances

Speed management is central to the Safe Systems approach to road safety, since it is the duty of the road designers and road operators to design roads and set speed limits such that all road users, including pedestrians and cyclists, can use those roads without risk of serious injury or fatality. But, of course, road users have to obey those limits, either voluntarily or through enforcement. So compliance with speed limits is crucial. Here cognitive dissonance can have a positive effect in terms of safety. Users who are pressured into behaviour change or rule compliance may adjust their attitudes to conform to their new behaviour, so that rather than resisting, they grow to accept the new reality and become conformists.

One can observe such changes in attitude with driver assistance systems that restrict rule violation such as Intelligent Speed Assistance (ISA), the system that discourages driving above the speed limit. In the ISA trials conducted in the UK, it could be noted that the attitudes of the participants, who all had four months of driving with vehicles equipped with a soft speed limiter (i.e. one that defaulted to limiting speed to the prevailing speed limit, but which could nevertheless be overridden) went through a change. Mean intention to speed was −0.90 in the baseline situation before the ISA system was enabled, −1.14 at the end of the period of driving with ISA and −1.28 in the after period when ISA had been disabled (Chorlton and Conner, 2012). Negative intention to speed here indicates intention to comply, so that in this instance the driver became increasingly willing to comply.

Attitudes to speed compliance and speed enforcement are not just formed at the individual level; they also have a strong social element. In France, prior to rollout of automatic speed cameras as an enforcement tool by the Chirac government in 2003, there was a culture among French drivers and society at large that it was acceptable to speed. In addition to its highly effective deployment of automatic enforcement, the French government also conducted the LAVIA project, using very similar technology to that used in the UK ISA trials to examine the attitudinal, behavioural and safety implications of driving with ISA. The attitudes of French drivers who lived in the area in which the trial was conducted were examined by Pianelli et al. (2007). They applied the *Social Representation Theory* of Jean-Claude Abric, which holds that attitudes tend to be held in common, i.e. have a very strong social element, from which it follows that to change attitudes it is necessary to a change the shared representations that the group or groups hold. "[S]ocial representations can be defined as 'systems of opinions, knowledge, and beliefs' particular to a culture, a social category, or a group" (Rateau et al., 2011).

Pianelli et al. (2007) found that attitudes towards the LAVIA system were strongly conditioned by general attitudes towards speed and speeding. They identified four different groups in the population of drivers: *prudent drivers* who saw excessive speed as dangerous; *defiant drivers* who enjoyed danger and obtained pleasure from speed; *hedonists* who gained pleasure from moving fast and from saving time; and *pragmatists* who, while they also valued moving fast and saving time, were also concerned about enforcement. Attitudes about the LAVIA system were in line with those representations of speed, so that for the prudent drivers LAVIA signified safety, peace of mind, compliance with speed limits, vigilance and assistance,

while for the defiant drivers it was seen as a constraint. Here we can see strong dissonances between groups, dissonances that will have to be overcome to secure general voluntary use of a system such as ISA. Measures to ensure acceptance and compliance with ISA and similar systems will have to be tailored to the attitudes and preferences of the various groups of drivers.

Dissonance can then occur between current knowledge and additional knowledge related to the use of new technology and new system for instance. Figure 1 gives then an example of the risk analysis process of dissonances by taking into account these two sets of knowledge related to experiences, tests, feedback from human-machine systems. When dissonances occur, a risk analysis process is required in order to decide if their resulting risks are acceptable. Risk analysis does not only focus on safety because the analysis can be a compromise between several criteria of performance. If the risk of a dissonance is considered as unacceptable then the dissonance may be rejected. Taking into account several compromises, the rejection consists in eliminating, adding or attenuating the impact of the dissonance into the current knowledge. The acceptance of the dissonance relates to its integration into the current knowledge. The rejection and the acceptation required a possible reinforcement of the content of the current knowledge.
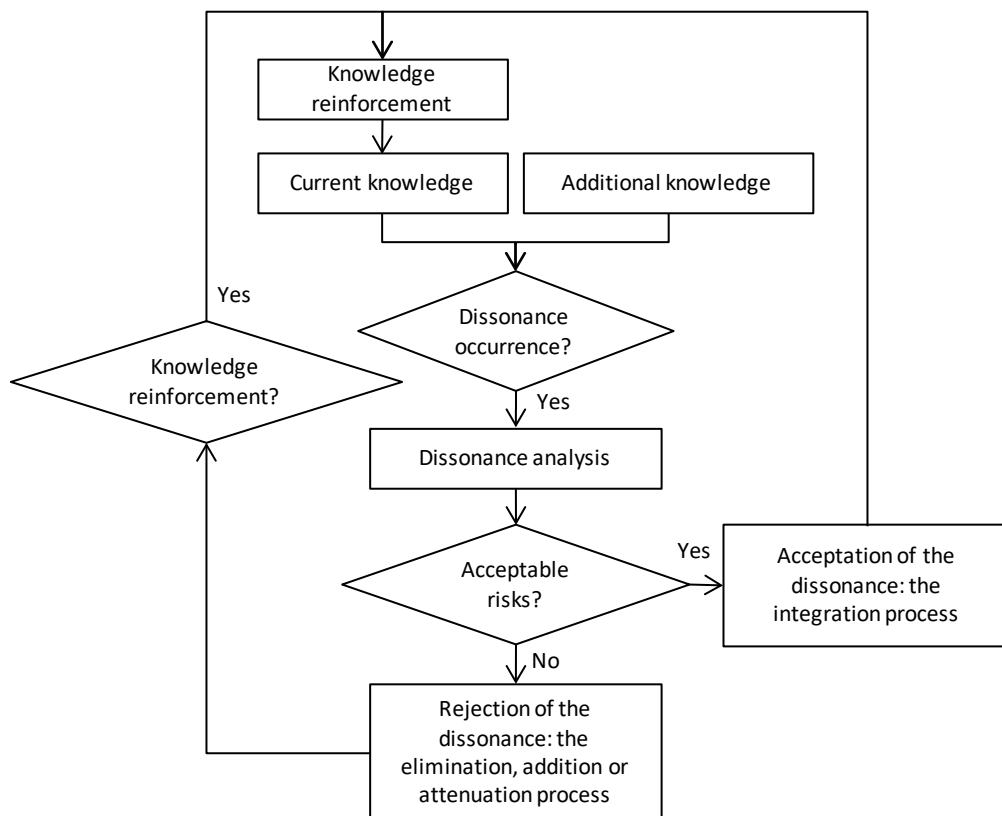

Figure 1. Risk analysis process of dissonances

A control of dissonance can relate to its negative or positive consequence perception. Indeed, the first control of an unprecedented dissonance may generate negative perception because of the induced workload or discomfort required for recovering it. After a couple of similar dissonance processing, positive consequence can be perceived and the corresponding knowledge can become the new norms to be followed. The so-called Benefit-Cost-Deficit or Danger model (BCD model) is then useful for analysing positive and negative consequences of a dissonance in terms of several criteria such as preference, workload, safety, security, economy and quality of human activity (Vanderhaegen et al., 2011; Sedki et al., 2013). The BCD model consists in analysing a given behaviour related to another one in terms of gains, i.e. benefits, and of losses, i.e. costs and deficits. Costs are acceptable losses whereas deficits are unacceptable. The classical risk analysis focuses on the probability of occurrence of a given event combined with its consequences. The BCD model aims at extending this approach by taking into account the positive impacts and the possible acceptable but negative ones of events such as dissonances.

Therefore, even if a dissonance occurs when something sounds wrong, its analysis may identify its positive and negative impacts in order to handle dynamically the possible evolution of its risk analysis and the current knowledge associated to the

functioning rules of a human-machine system. Next section gives some examples of application of this risk analysis process of dissonances.

6. A case of affordance, automation surprise and social dissonance

This examplar concerns the evolution of knowledge in the car driving domain.

Suppose that at a given time, the knowledge of a driver is composed of simple rules related to manual car speed control and to manual aquaplaning control:
- R1: To increase the current car speed manually implies to push the gas pedal
- R2: To decrease the current car speed manually implies to release the gas pedal
- R3: To control aquaplaning implies not to brake
- R4: To control aquaplaning implies not to accelerate

Few months or years after, suppose that the car is equipped with a Cruise Control system (CC). New knowledge of the driver might be developed related to the use of the CC (i.e. the control of the setpoint of the CC) and the delegation of tasks (i.e. the delegation of the speed control) from the driver to the CC or reverse:
- R5: To increase the car speed setpoint when the CC is activated implies to push on the "+" button of the activated CC
- R6: To decrease the car speed setpoint when the CC is activated implies to push on the "−" button of the activated CC
- R7: To turn on the CC implies to push on the "on" button of the CC
- R8: To turn off the CC implies to push on the "off" button on of the CC
- R9: To deactivate the CC implies to brake

Moreover, the driver may develop a model of the CC behaviour by building rules such as:
- R10: To increase the current car speed when it is under the CC setpoint and when the CC is activated implies that the CC will increase engine speed
- R11: To decrease the car speed when it is over the CC setpoint and when the CC is activated implies that the CC will reduce engine speed

Applying methods such those developed in (Vanderhaegen, 2014, 2016), it is possible to identify dissonances such as affordances or inconsistencies of a knowledge base composed by rules. Risk analysis of the evolution of knowledge based on the driver experience can then be done.

Figure 2 illustrates the possible occurrence of conflicts of use, i.e. affordances, and of delegation of task, i.e. inconsistencies.
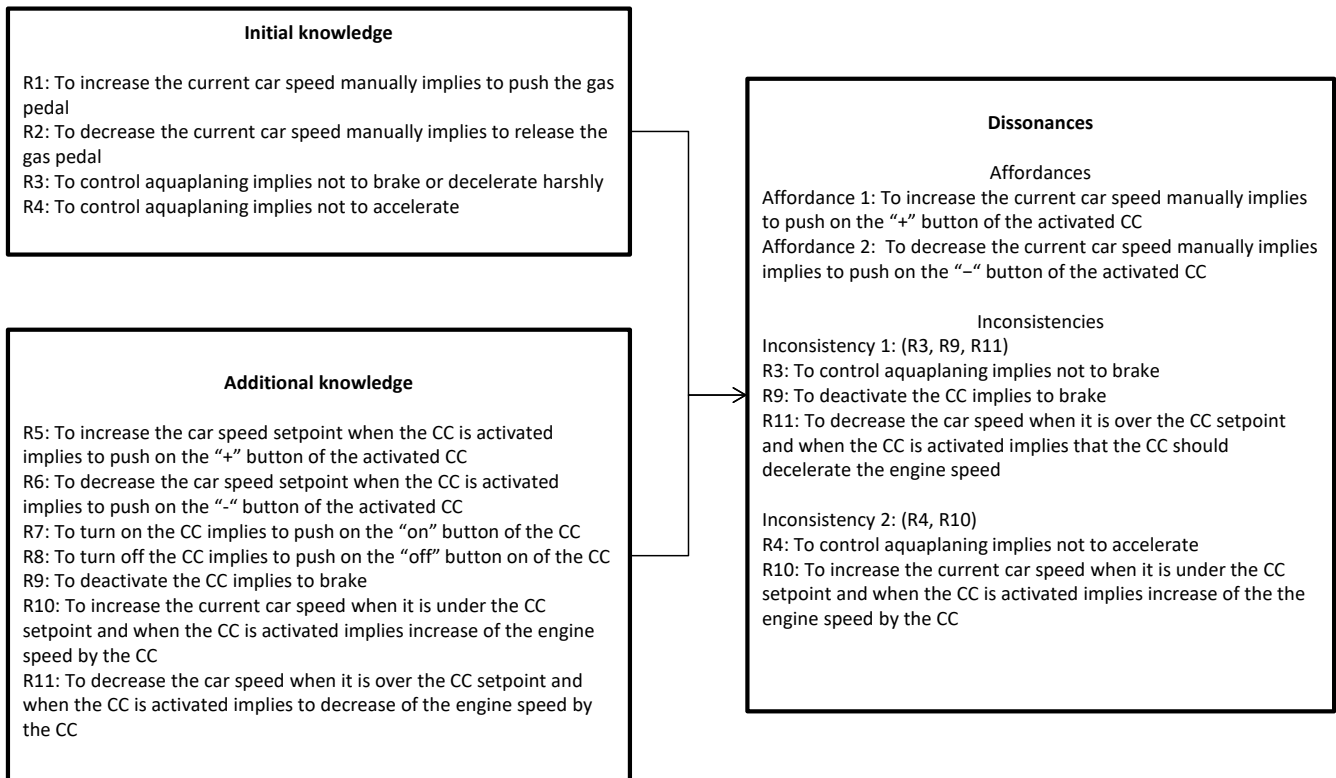
Initial knowledge

R1: To increase the current car speed manually implies to push the gas pedal
R2: To decrease the current car speed manually implies to release the gas pedal
R3: To control aquaplaning implies not to brake or decelerate harshly
R4: To control aquaplaning implies not to accelerate

Additional knowledge

R5: To increase the car speed setpoint when the CC is activated implies to push on the "+" button of the activated CC
R6: To decrease the car speed setpoint when the CC is activated implies to push on the "-" button of the activated CC
R7: To turn on the CC implies to push on the "on" button of the CC
R8: To turn off the CC implies to push on the "off" button on of the CC
R9: To deactivate the CC implies to brake
R10: To increase the current car speed when it is under the CC setpoint and when the CC is activated implies increase of the engine speed by the CC
R11: To decrease the car speed when it is over the CC setpoint and when the CC is activated implies to decrease of the engine speed by the CC

Dissonances

Affordances
Affordance 1: To increase the current car speed manually implies to push on the "+" button of the activated CC
Affordance 2: To decrease the current car speed manually implies implies to push on the "–" button of the activated CC

Inconsistencies
Inconsistency 1: (R3, R9, R11)
R3: To control aquaplaning implies not to brake
R9: To deactivate the CC implies to brake
R11: To decrease the car speed when it is over the CC setpoint and when the CC is activated implies that the CC should decelerate the engine speed

Inconsistency 2: (R4, R10)
R4: To control aquaplaning implies not to accelerate
R10: To increase the current car speed when it is under the CC setpoint and when the CC is activated implies increase of the the engine speed by the CC

Figure 2. Identification of dissonances between an initial knowledge and an additional one.

The affordances 1 and 2 relate to the use of the "+" and "-" buttons of the CC as an accelerator and braking systems respectively. The benefits of such new behaviours concerns the decreasing of the workload related to the management of the pedals. No direct cost can be identified but possible danger of such new functions of the CC interfaces might be a failed control due to the increasing of the reaction time in case for example of emergency stop.

Inconsistencies 1 and 2 relate to opposite actions related to the knowledge of a driver or to knowledge of the driver and the CC. Rules R3 and R9 concern the driver who has to brake to deactivate the CC and not to brake in case of an aquaplaning occurrence. Additionally, rule R11 represents a CC behavioural model related to speed reduction when the current speed car is over the speed setpoint, whereas the driver may decide not to brake in case of an aquaplaning occurrence. Even if specific conditions have to be gathered to observe such contradictions, it is important to analyse their associated risks and avoid possible loss of control of the car. Rules R4 and R10 concern the driver and the CC respectively regarding of engine speed. Benefits of the use of automated systems such as an CC can then decrease when hazardous situations associated to their use are discovered.

Such evolution of knowledge requires the identification of dissonance and the risk analysis process of these dissonances in order to modify the human practices and the associated knowledge or to modify the human-machine system organization or structure. Regarding the automation of car driving as defined by SAE International (2016), the levels of automation have to be studied in this sense. Indeed, for example, Levels 2 or 3 entitled partial automation or conditional automation respectively have to guaranty that there are no possible dissonances related to the use of automated systems and related to the capacity of the automated system to detect and treat dangerous dissonances. This capacity of the automated system is much more relevant and obvious for Levels 4 and 5, i.e. high automation and full automation. This means that the risk analysis process requires new methods based on dissonance engineering.

In the case of automated driving at the intermediate levels defined by SAE International (2016), i.e. at Levels 2 and 3, the human and the machine constitute a joint cognitive system. Here there is ample opportunity for dissonances between both human and machine. The machine can misinterpret human intention: maybe the human would prefer the vehicle to drive more slowly or faster; maybe the human wants a higher level of automated support, so that he/she can engage in infotainment, while the automation senses that road markings are fading and therefore wants more human attention. Similarly the human needs to understand the amount of support currently being given by the automation and the capability of that automated support. If that understanding is not properly calibrated, the human may over-trust the machine as may

well have occurred prior to the fatal crash of the Tesla being driven with "Autopilot" in Florida in May 2017. In that instance, the driver was reportedly watching a video on the approach to an intersection with a potential for turning traffic that the vehicle was not capable of handling. Equally, the human may distrust the automation and might intervene disastrously in the middle of a time-critical manoeuvre, thus creating a crash in a situation that the vehicle on its own would have been able to manage safely.

Social dissonance may also be an issue with highly automated driving. The vehicle is likely to choose to drive at a safe time headway in car following. But very many drivers choose to drive at time headways that are too small to be safe and that are well below the time headways that are recommended by the authorities or, in some countries, stipulated in law. If the automated vehicle choses to drive at say a time headway of 2 seconds, other vehicles are likely to cut in ahead, giving the human the feeling that the vehicle is receding in the traffic stream. So the human may want the vehicle to select an unsafe time headway. This is actually allowed in many Adaptive Cruise Control systems, but there the driver is supposedly still fully engaged in the driving task, so that the vehicle is not responsible for its actions. In automated driving, especially at SAE Level 3 and above, the vehicle *is* responsible, but the human may not understand or respect the system's behaviour.

Traffic is a social system, involving the interplay of multiple actors — vehicle drivers, motorcycle riders, pedestrians, cyclists, horse riders and others. That system works fairly well, albeit there are breakdowns and misunderstandings resulting in near misses and collisions. There is also the problem of rule violations, which can lead to severe events. Crucial to that normal operation is communications, typically by means of informal cues but also by such means as vehicle indicators or cyclists' hand signals, between road users. Adding automated vehicles to the current mix poses the challenge of whether they will have their own distinct rule sets and behaviours and of how the human participants in traffic will understand the behaviours and intended actions of the automated vehicles. Such questions arise as whether the fact that a vehicle is driving itself will have to be indicated to the outside, whether an external HMI on the automated vehicle is needed to indicate intention, and, if the answer to those questions is positive, how do we achieve consensus on the form those indications should take. A world in which we had hundreds of different communication strategies would be totally confusing and dangerous — full of dissonance.

7. Conclusion

This paper extended one of the challenging points developed in Cacciabue et al. (2014) related to the added value of dissonance engineering for risk analysis. Both positive and negative impacts have to be considered in future risk analysis process of dissonances. Future risk analyses have to consider a dissonance as an undesirable event to assess its probability of occurrence, and then its analysis has to interpret its consequences in terms of positive and negative impacts. This paper has focused on the identification of dissonances and on possible ways to analyse their associated risks.

Designers may consider the possible dissonance discovery and control capacity of human-machine systems instead of limiting knowledge development to all the possible situations the system may face. It is obvious that the first alternative related to dissonance discovery and control capacity has the advantage of taking into account the knowledge discovery process because the second one, i.e. the development of systems that are capable of solving any situation, cannot guaranty the completeness of the implemented knowledge. As a matter of fact, the corresponding risk analysis has to evolve from a static process based on the current knowledge of human-machine systems to a dynamic one by taking into account the assessment of dissonances and the possible evolution of the resulting knowledge. Two main challenges in risk analysis have to be considered regarding this dissonance discovery and control capacity. The first one concerns the autonomy of future automated systems and the associated risks of their uses when dissonances occur — we may even face systems that cannot be "designed" in the traditional sense. Even if autonomous systems are capable of learning on their own and creating new knowledge, there may be risks from this new knowledge. And because the systems have self-learning capacity, it may not be possible to identify those risks by formal methods. Indeed the risk may only emerge after the fact, and it may not be possible to identify what has caused the new behaviour. The second challenge relates to the possible evolution of the risk analysis of dissonances by the users of a human-machine system. This analysis is not static but dynamic. Therefore, prospective analysis is not sufficient and has to be combined with on-line and retrospective analysis.

References

Aïmeur E (1998) Application and assessment of cognitive dissonance—theory in the learning process. J Univers Comput Sci 4(3):216–247

Ben Yahia W, Vanderhaegen F, Polet P, Tricot N (2015) A2PG: alternative action plan generator. Cogn Technol Work 17:95–109

Brunel O, Gallen C (2011). Just like cognitive dissonance. 27th International Congress of French Association of Marketing, 18–20 May 2011, Brussels

Cacciabue C, Carsten O, Vanderhaegen F (2014) Is there still a need for CTW? Cogn Technol Work 16:311–317

Chen T-Y (2011) Optimistic and pessimistic decision making with dissonance reduction using interval-valued fuzzy sets. Inf Sci 181(3):479–502

Chen C-H, Khoo LP, Chong YT, Yin XF (2014) Knowledge discovery using genetic algorithm for maritime situational awareness. Expert Syst Appl 41(6):2742–2753

Chorlton K, Conner M (2012) Can enforced behaviour change attitudes: exploring the influence of intelligent speed adaptation. Accid Anal Prev 48:49–56

Dali S (1975) Gala Contemplating the Mediterranean Sea. Dalí Theatre-Museum, Figueres, Spain

Dehais F, Causse M, Vachon F, Tremblay S (2012) Cognitive conflict in human-automation interactions: a psychophysiological study. Applied Ergonomics 43(3):588–595

Festinger L (1957) A theory of cognitive dissonance. Stanford University Press, Stanford

Gibson JJ (1986) The ecological approach to visual perception (Originally published in 1979). Lawrence Erlbaum Associates, Hillsdale

Hollnagel E, Woods DD (2005) Joint cognitive systems: foundations of cognitive systems engineering. CRC Press, Boca Raton

Hunter A, Summerton R (2006) A knowledge-based approach to merging information. Knowl-Based Syst 19:647–674

Inagaki T (2008) Smart collaboration between humans and machines based on mutual understanding. Annu Rev Control 32:253–261

Kervern G-Y (1995) Eléments fondamentaux des cindyniques (Fundamental elements of cindynics). Economica Editions, Paris

Lyons JN (2013). Being transparent about transparency: a model for human-robot interaction. In: Trust and autonomous systems: papers from the 2013 AAAI spring symposium, pp. 48–53

Massironi M, Savardi U (1991) Why anamorphoses look as they do: an experimental study. Acta Psychol 76(3):213–239

McCay-Peet L, Toms EG, Kelloway EK (2015) Examination of relationships among serendipity, the environment, and individual differences. Inf Process Manage 51(4):391–412

OECD International Transport Forum (2008) Towards zero: ambitious road safety targets and the safe system approach. OECD, Paris

Ouedraogo A, Enjalbert S, Vanderhaegen F (2013) How to learn from the resilience of Human–Machine Systems? Eng Appl Artif Intell 26(1):24–34

Pianelli C, Saad F, Abric, J-C (2007) Social representations and acceptability of LAVIA (French ISA system). In: Proceedings of the 14th world congress and exhibition on intelligent transport systems and services, 8–12 October 2007, Beijing

Polet P, Vanderhaegen F, Amalberti R (2003) Modelling border-line tolerated conditions of use (BTCU) and associated risks. Saf Sci 41(2–3):111–136

Polet P, Vanderhaegen F, Zieba S (2012) Iterative learning control based tools to learn from human error. Eng Appl Artif Intell 25(7):1515–1522

Rateau P, Moliner P, Guimelli C, Abric J-C (2011) Social representation theory. In: Van Lange PAM, Kruglanski AW, Higgins ET (eds) Handbook of theories of social psychology, vol 2. Sage Publications, Los Angeles, pp 477–497

Rouhiainen V, Gunnerhed M (2002) Development of international risk analysis standards. Saf Sci 40:57–67

Rushby J (2002) Using model checking to help discover mode confusions and other automation surprises. Reliab Eng Syst Saf 75(2):167–177

SAE International (2016) Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. Society of Automotive Engineers report, SAE J3016

Sedki K, Polet P, Vanderhaegen F (2013) Using the BCD model for risk analysis: an influence diagram based approach. Eng Appl Artif Intell 26(9):2172–2183

Telci EE, Maden C, Kantur D (2011) The theory of cognitive dissonance: a marketing and management perspective. Procedia Soc Behav Sci 24:378–386

Tingvall C, Haworth N (1999) Vision zero: an ethical approach to safety and mobility. In: 6th ITE international conference road safety & traffic enforcement: beyond 2000, 6–7 September, Melbourne, Australia

Vanderhaegen F (1999a) Multilevel allocation modes—allocator control policies to share tasks between human and computer. Syst Anal Model Simul 35(191–213):1999

Vanderhaegen F (1999b) Cooperative system organisation and task allocation: illustration of task allocation in air traffic control. Le Travail Humain 63(3):197–222

Vanderhaegen F (1999c) Toward a model of unreliability to study error prevention supports. Interact Comput 11:575–595

Vanderhaegen F (2010) Human-error-based design of barriers and analysis of their uses. Cogn Technol Work 12:133–142

Vanderhaegen F (2012a). Dissonance engineering for risk analysis. Workshop: Risk Management in Life Critical Systems, Human-Centered Design Institute, Florida Institute of Technology, Melbourne, FL, USA, March

Vanderhaegen F (2012b) Cooperation and learning to increase the autonomy of ADAS. Cogn Technol Work 14(1):61–69

Vanderhaegen F (2014) Dissonance Engineering: a New Challenge to Analyse Risky Knowledge When using a System. Int J Comput Commun Control 9(6):776–785

Vanderhaegen F (2016) A rule-based support system for dissonance discovery and control applied to car driving. Expert Syst Appl 65:361–371

Vanderhaegen F, Caulier P (2011) A multi-viewpoint system to support abductive reasoning. Inf Sci 181(24):5349–5363

Vanderhaegen F, Jouglet D, Piechowiak S (2004) Human-reliability analysis of cooperative redundancy to support diagnosis. IEEE Trans Reliab 53(4):458–464

Vanderhaegen F, Chalme´ S, Anceaux F, Millot P (2006) Principles of cooperation and competition—application to car driver behavior analysis. Cogn Technol Work 8(3):183–192

Vanderhaegen F, Zieba S, Polet P (2009) A reinforced iterative formalism to learn from human errors and uncertainty. Eng Appl Artif Intell 22:654–659

Vanderhaegen F, Zieba S, Enjalbert S, Polet P (2011) A Benefit/Cost/ Deficit (BCD) model for learning from human errors. Reliab Eng Syst Saf 96(7):757–766

Westrum R (2006) A typology of resilience situations. In: Hollnagel E, Woods D, Leveson N (eds) Resilience engineering: concepts and precepts. Ashgate, Aldershot

Yozgat U, CAliskan SC, Uru FO (2012) Exploring emotional dissonance: on doing what you feel and feeling what you do. Procedia Soc Behav Sci 58:673–682

Zieba S, Polet P, Vanderhaegen F, Debernard S (2010) Principles of adjustable autonomy: a framework for resilient human machine cooperation. Cogn Technol Work 12(3):193–203

Zieba S, Polet P, Vanderhaegen F (2011) Using adjustable autonomy and human-machine cooperation for the resilience of a human-machine system—application to a ground robotic system. Inf Sci 181:379–397