Cogn Tech Work (2012) 14:3–18
DOI 10.1007/s10111-011-0191-6

ORIGINAL ARTICLE

# Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations

**Frank Flemisch · Matthias Heesen · Tobias Hesse · Johann Kelsch · Anna Schieben · Johannes Beller**

Received: 18 July 2011 / Accepted: 14 September 2011 / Published online: 18 November 2011


**Abstract** Progress enables the creation of more automated and intelligent machines with increasing abilities that open up new roles between humans and machines. Only with a proper design for the resulting cooperative human–machine systems, these advances will make our lives easier, safer and enjoyable rather than harder and miserable. Starting from examples of natural cooperative systems, the paper investigates four cornerstone concepts for the design of such systems: ability, authority, control and responsibility, as well as their relationship to each other and to concepts like levels of automation and autonomy. Consistency in the relations between these concepts is identified as an important quality for the system design. A simple graphical tool is introduced that can help to visualize the cornerstone concepts and their relations in a single diagram. Examples from the automotive domain, where a cooperative guidance and control of highly automated vehicles is under investigation, demonstrate the application of the concepts and the tool. Transitions in authority and control, e.g. initiated by changes in the ability of human or machine, are identified as key challenges. A sufficient consistency of the mental models of human and machines, not only in the system use but also in the design and evaluation, can be a key enabler for a successful dynamic balance between humans and machines.

F. Flemisch (✉)
(DLR-ITS), RWTH Aachen University, Institute of Industrial Engineering and Ergonomics IAW, Fraunhofer Institute for Communication, Information Processing and Ergonomics FKIE, Bonn, Germany
e-mail: frank.flemisch@fkie.fraunhofer.de

M. Heesen · T. Hesse · J. Kelsch · A. Schieben · J. Beller
DLR Institute of Transportation Systems, Braunschweig, Germany

**Keywords** Assistant systems · Automation · Human-machine cooperation · Adaptive automation · Levels of automation · Balanced automation

## 1 Introduction: The fragile balance between humans and automation

**In general**, scientific and technological progress, in close coupling with cultural achievements, offers benefits that our ancestors could only dream of. Properly applied, machines can make our lives easier, and improperly applied, machines can make our lives really miserable. Advances in hardware and software power hold promise for the creation of more and more intelligent and automated machines.

How do we design these complex human machine systems? How do we balance between exploiting increasingly powerful technologies and retaining authority for the human? How can we define clear, safe, efficient and enjoyable roles between humans and automated machines? Which of the subsystems of future human–machine systems should have which ability, which authority and which responsibility? Can authority, responsibility and control be traded dynamically between human and automation? What other concepts besides authority and responsibility do we need to describe and shape a dynamic but stable balance between humans and automation?

**Applied to movement**, vehicles, a special kind of machines, can help us to move further, faster, safer and more efficient. These moving machines become more capable and autonomous as well: At the beginning of the twenty-first century, vehicles like modern airplanes are already so sophisticated that they can operate autonomously for extended periods. Prototype cars utilizing

machine vision can, under limited circumstances, drive fully autonomously on public highways (Dickmanns 2002), deserts (e.g. Thrun et al. 2006) or urban environments (Montemerlo et al. 2008; Wille et al. 2010).

But advances in hardware and software do not automatically guarantee more intelligent vehicles. More importantly, intelligent or autonomous vehicles do not necessarily mean progress from which humans can really benefit. In aviation, a forerunner in technology through the twentieth century, the development towards highly automated and intelligent aircraft led not only to a reduction of physical workload but also to problems like mode confusion, human-out-of-the-loop and many more (Billings 1997; FAA 1996; Wiener 1989). This could create what Bainbridge calls the "ironies of automation," where "by taking away the easy parts of human tasks, automation can make the difficult parts … more difficult" (Bainbridge 1983). If more and more assistance and automation subsystems are possible for vehicles, how do they cooperate with the human, what abilities do they have, what authority for the control of which aspects of the driving task and who bears which responsibility?

In an effort to foster the understanding of underlying principles and facilitate the answers to some of these open questions, this paper starts with a brief look into natural cooperative systems and then investigates four cornerstone concepts for the design of human–machine systems: ability, authority, control and responsibility. An ontology of these cornerstone concepts is developed to provide a framework of consistent relations between the four as basis for further analysis and design. The cornerstone concepts are linked to other important concepts like level of automation or autonomy. Consistency between ability, authority, control and responsibility is identified as an important quality of a human–machine system. Additionally, a graphical tool is developed that can help to simplify the design and analysis of human machine systems by visualizing the cornerstone concepts and their relations in a single diagram. The use of the devised framework and its visualization are demonstrated by the application to the human–machine interaction in existing prototypes of highly automated vehicles.

## 2 Inspiration for ability, authority, control and responsibility in cooperative situations from non-technical life

**In general**, if machines become more and more intelligent, what role should they play together with humans? The interplay of intelligent entities is historically not new, but as old as intelligence itself. In nature and everyday life, there are many examples for this: flocks or herds of animals living and moving together, or people interacting with each other and the environment. Acting together does not necessarily mean acting towards common goals: Competitive behaviour like hunting for the same food source or in the extreme killing each other is quite common in nature. Competitive behaviour in the form of market competition might be a necessary part of human life, and competitive behaviour in the form of war is clearly an undesirable part of human life. In contrast to the competition, cooperation as a means to successfully compete together against other groups or against challenging circumstances seems to be a historically newer, but quite successful, concept.

**Applied to movement**, cooperation is also a common concept in the non-technical world. Imagine a crowd of people moving along a street, including a parent and a child walking hand-in-hand. Another example would be a driver and a horse both influencing the course of a horse cart, or a pilot and a co-pilot alternatively controlling an airplane. Differences and interplay of abilities, authority, control and responsibility shape out different characteristics of those cooperative movement systems.

A young child on the hand of the parent will have a different authority than her parent, e.g., to determine the crossing of a busy road. The decision when and how to cross the road will depend here mainly on the weaker ability of the child (and the ability of the parent to carry the child quickly out of danger if necessary). If something goes wrong, the parent will be held completely responsible.

Imagine the situation of a rider or coach driver and a horse: The horse has much stronger and faster abilities in movement, but the human usually has a higher authority except in emergency situations where the horse already reacts before the human might even be aware of a danger. The human can control the horse quite directly with a tight rein, or more indirectly with a loose rein. Even with a loose rein, the human will keep a majority of the responsibility. The breeder (or owner) will only be held responsible, if the horse behaves outside of the range of accepted behaviour.

Imagine the situation of a pilot and co-pilot: Only one of the two pilots is actually flying the aircraft (often called the pilot flying), while the other pilot is assisting. Regarding the authority, there is a clear seniority where the senior pilot or captain (who usually also has the higher experience, but not necessarily the higher abilities in a particular situation) can take over control at any time. When control is interchanged between the two pilots, this is usually done in a schematic way with the wording "I take control," with the other pilot responding "You have it." Regarding the responsibility, the pilot flying has a high responsibility for the flying task within his or her ability, but the captain will usually be held responsible as well if the other pilot who was not so experienced caused an accident (Fig. 1).

These natural examples of cooperative behaviour, here especially cooperative movement, can also be helpful to understand and design human–machine systems. The


**Fig. 1** Cooperative situations in nature and in human–machine systems

metaphor of an electronic co-pilot is used in aviation (e.g. Flemisch and Onken 1999) and in car and truck safety systems, e.g. Holzmann et al. 2006. While the co-pilot metaphor is also raising anthropomorphic expectations, the metaphor of horse and rider (or horse and cart driver) describes a more asymmetric relationship of cooperative control of movement (Flemisch et al. 2003). The examples have influenced both the framework of ability, authority, control and responsibility and the example, e.g., of highly automated vehicles in the EU project HAVEit, described further down. The examples can also be an inspiration for any kind of human–machine system dealing with ability, authority, control and responsibility issues.

## 3 Ontology: human–machine systems, ability, authority, control and responsibility between humans and machines

To have a chance to grab the essence of cooperation in human machine systems in general, and especially of authority, ability, control and responsibility, let's apply a rather abstract perspective for a moment and describe the concepts more precisely.

**In general**, and in an abstract perspective, the world including natural systems and human–machine systems embedded in their environment (Fig. 2) is not static, but changes over time from one state or situation to another. A substantial part of this change is not incidental but follows the actions of acting subsystems or actors (sometimes called agents), which can be natural (e.g. humans) and/or artificial (e.g. machines), and their interplay with the environment. Based on (explicit or implicit) understanding of good or bad situations (e.g. with the help of goals and/or motivations), actors perceive the world and influence the situation by using their abilities to act, thereby forming (open or closed) control loops.

For human–machine systems, the behaviour of the machine (i.e. its abilities, the amount of control it exercises and the distribution of authority and responsibility between human and machine) is determined outside in the meta-system. The meta-system includes, among others, the equipment and people responsible for the development and the evaluation, see Fig. 2. This determination is done usually before and after the operation phase, e.g. during the development phase or in an after-the-fact evaluation phase, e.g. in case of an accident. An important feedback loop is running via the meta-system, where experience is used to enhance the system design for future systems.

*Control* means having "the power to influence […] the course of events" (Oxford Dictionary), Applied to human machine systems, to have control means to influence the

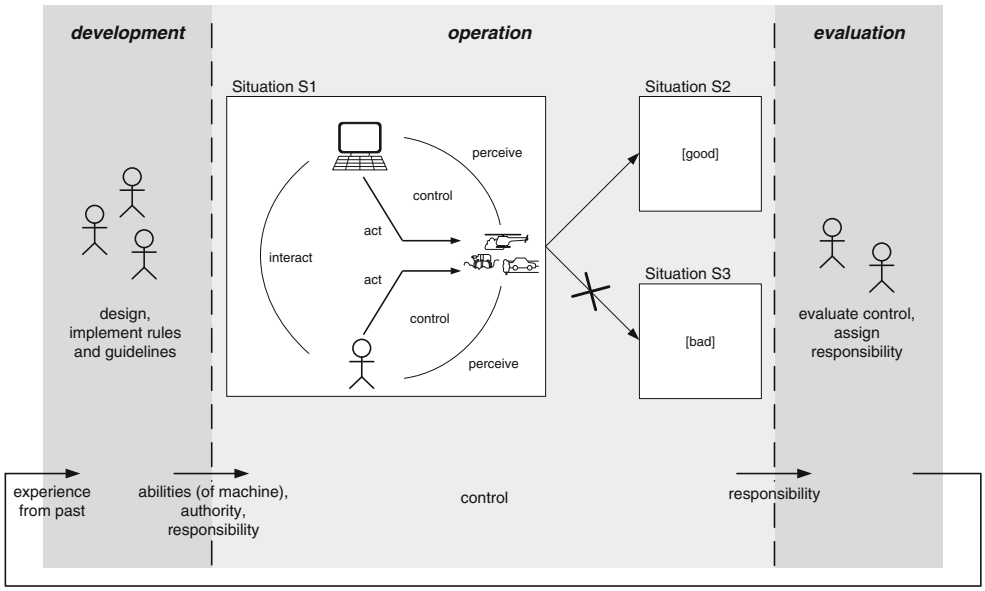


**Fig. 2** Ability, authority, responsibility and control in the three phases development, operation and evaluation

situation so that it develops or keeps in a way preferred by the controlling entity. Usually for the control of a situation, there has to be a loop of *perception*, *action selection* and *action* that can stabilize the situation and/or change it towards certain *aims or goals* (Fig. 3). If necessary, the concept of control can be linked to the concept of (control) *tasks and subtasks*, where the completion of (control) tasks contributes to the general goal of control.

While the action and action selection should always exist, especially the perception could be missing, e.g. if the actor does not receive certain sensor information (e.g. a human taking his eyes from the situation). From a control theory perspective, the "closed-loop control" changes to what is called "open-loop control," in case it is not closed by the perception of the outcome of the control action, thereby altering the overall system dynamics. From a human factors perspective, missing perception might cause an "out-of-the-loop" problem (e.g. Endsley and Kiris 1995) that refers to the fact that necessary parts of the control loop are not present or activated enough so that control cannot be asserted.

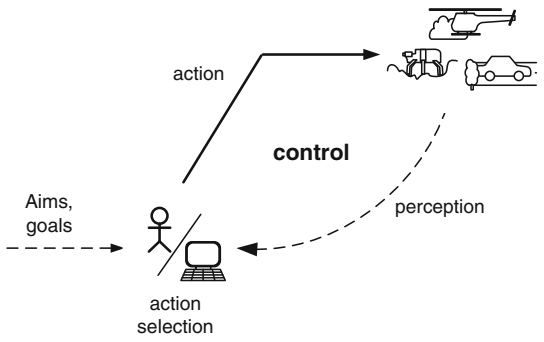


**Fig. 3** Single control loop

*Ability* in general is the "possession of the means or skill to do something" (Oxford Dictionary). Applied to human machine systems, ability can be defined as the possession of the means or skill to perceive and/or select an adequate action and/or act appropriately.

Related in meaning and also frequently used is the term competency that "refers to correct behaviour in context" (Miller and Parasuraman 2007). In many cases, a control task requires not only skills but also the use of certain resources. Therefore, the term ability used in the following text includes having the necessary competence, skills and resources (e.g. time, tools or personnel) to execute control, including perception, action selection and action.

*Authority* in general signifies "the power or right to give orders, make decisions, and enforce obedience" (Oxford Dictionary). Applied to human machine systems, the authority of an actor can be defined by what the actor is allowed to do or not to do. Usually, authority is given to an actor beforehand by the system designer and has an impact on evaluations after the use, e.g. in the case of an abuse of authority. Of main interest in this context are two levels of authority:

> Control authority: This is the authority of the actors to execute a certain control, or as described more precisely further down, a certain control distribution.
> (Control) Change authority: This is the authority to change the control authority to another control distribution giving more or less control to one of the actors.

Authority could even be abstracted or broken down further to relate to any part of the control loop or

interaction between actors, such as the authority to perceive, to act, to change the aim or to inform or warn the other actor (Miller and Parasuraman 2007).

*Responsibility* describes "a moral obligation to behave correctly," "the state or fact of having a duty to deal with something" or "the state or fact of being accountable or to blame for something" (Oxford Dictionary). Applied to human machine systems, responsibility is assigned beforehand to motivate certain actions and evaluated afterwards, where the actor is held accountable or to blame for a state or action of the human machine system and consequences resulting thereof. It can make sense to differentiate between a *subjective responsibility* that an actor feels regarding his actions, which can differ from the *objective responsibility* mostly defined by other entities and by which the actor is then judged.

Before we proceed with the four cornerstones ability, authority, control and responsibility, a brief look is taken into some of the many more related or connected concepts. One example is autonomy as a quality how much actors depend on each other in their actions (described in further detail, e.g., by Miller 2005). Autonomy is used, e.g., in the job demand-control model (Karasek 1979), stating that high demand without sufficient autonomy leads to stress. Autonomy and the fragile balance with its antipodal quality cooperativeness can be an important aspect to explain why certain task combinations work better than others.

Another example is the concept of levels of automation (Parasuraman et al. 2000) which e.g. (Miller 2005) describes as follows: "A 'Level of Automation' is, therefore, a combination of tasks delegated at some level of abstraction with some level of authority and resources delegated with some level of authority to be used to perform that (and perhaps other) task(s). The 'level of automation' in a human–machine system increases if the level of abstraction, level of aggregation or level of authority […] increases". In this paper, levels of (assistance and) automation corresponds to the distribution of control. A high level of automation is a control distribution with a high percentage of control done by the machine and a low level of automation with a low percentage of control done by the machine.

Now back to the four cornerstones of this paper, how do the concepts ability, authority, control and responsibility relate to one another?

The most evident relationship is between ability and control: Ability enables control, or in other words, no successful control is possible without sufficient ability. Second, the appropriate authority is needed to be allowed to control. Note, however, that control does not occur automatically once the ability and authority exist; the actor still needs to execute control. A certain subjective or objective responsibility might motivate him to do so. Depending on the a priori responsibility and the control actions, a final responsibility results, leaving the actor accountable for his actions.

Responsibility, authority and ability are not independent. Woods and Cook (2002) and Dekker (2002), for example, propose a double bind between authority and responsibility. Figure 4 displays an extension of this relationship to triple binds between ability, authority and responsibility: Ability should not be smaller than authority; authority should not be smaller than responsibility.

In other words, responsibility should not be bigger than ability and should not be bigger than authority.

More precisely, the portion of control for which (a priori) responsibility is assigned should be less or equal to the portion of control for which authority is granted and ability is available. Authority to control should only be granted to less or equal the extent that can be covered by the given ability. Remember that as defined above, the ability does not only include the skills and competence of each actor but also the resources at his disposal and therefore subsumes even their abilities. Responsibility without sufficient authority and ability would not be fair. The actor should have authority or responsibility only for (control) tasks that he or his resources are able to perform. It would not be wise to give authority to actors who do not have the appropriate ability.

Often, there is a tendency that authority should not be smaller than ability: Especially humans who estimate their abilities high also want to have an appropriate authority. In addition, sometimes, the existence of sufficient ability and authority to control constitutes also the responsibility to control. An example is a situation where a person had the ability to help another person in danger, did not help and is held responsible afterwards. This is brought to the point in the phrase from the movie "Spiderman," "With great power comes great responsibility," which originally can be attributed to Voltaire. In the context of this publication, power means having the ability and the control authority. Hence, the extent of given ability and authority may hint a certain responsibility, as indicated in Fig. 4.

## 4 Visualization of ability, authority, control and responsibility in A2CR diagrams

Let's get back to the focus point, where ability and authority come together to form control. How can this be structured if more than one actor can contribute to the control, e.g. if a human and a machine can both contribute to the control? The simplest way to distribute a control task between a human and a machine is that either the human or the machine is in control. However, if several actors such as a human and a machine act in a cooperative manner,

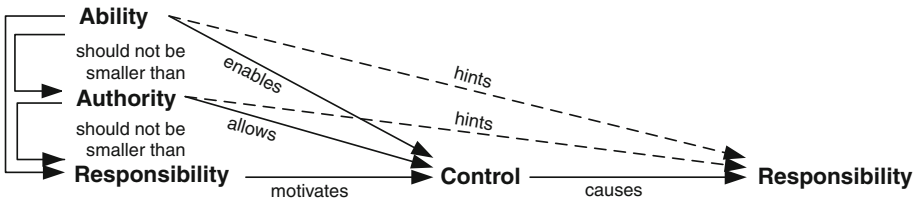


**Fig. 4** Relations between ability, authority, control and responsibility

they can share control. Then, the simple "switch" between human and machine is extended to a more complex relationship, which can be simplified into a spectrum or scale ranging from manual control (human has complete control) to fully automated (machine has complete control), see Fig. 5. On this continuous assistance and automation scale, different regions of control distributions can be identified such as assisted/lowly automated, where the human is doing most of the control task, semi-automated, where both human and machine contribute about half of the control, or highly automated, where the machine has taken over the majority of the control and the human still contributes to a smaller extent.

Each actor (human and machine) has certain abilities. Therefore, not every control distribution might be possible. More precisely, it is of importance whether human and machine have the ability to handle a certain control distribution, which might also depend on the situation. An example would be an emergency situation where an imminent action is necessary and the human cannot perform it due to his limited reaction time.

The range of possible control distributions can be visualized by bars on top of the assistance and automation spectrum, see Fig. 6. The top bar shows the control distributions on the spectrum that the human is able to handle, while the bottom bar shows the ones the machine is able to handle. In the first example of Fig. 6 (top), the human can handle all control distributions, but the machine cannot handle situations completely alone; it needs the human in the control loop at least to a minimum, here of 20%. Figure 6 (bottom) also shows a second example of a different situation, which the human cannot handle without a substantial amount of control by the machine, e.g., in

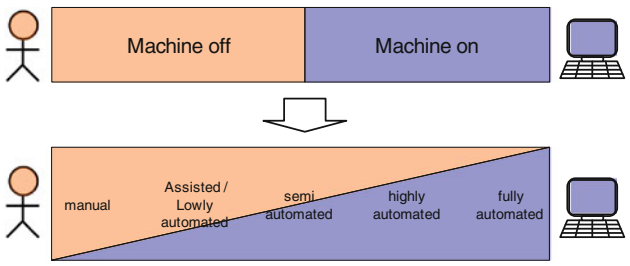


**Fig. 5** Assistance and automation spectrum (adapted from Flemisch et al. 2003, 2008)

difficult driving environments. Here, control distributions that are possible lie between 40 and 20%, and 60–80% of human and automation control, respectively.

Analogously to the abilities that enable certain control distributions, authority is required to allow them. The allowed control distributions can also be visualized together with the assistance and automation spectrum, as shown in the example in Fig. 7. In a human machine system, often only small areas within the range of all theoretically possible control distributions are realized, corresponding to the levels of automation implemented by the system designers. Corresponding to the example in Fig. 7, two small areas of control distribution are allowed for both the human and the machine. These areas on the control spectrum resemble levels of automation. Only within this specified areas, human and/or machine can have the control authority. In this example, we chose two areas, but there are other and also more areas imaginable, depending on which and how many levels of automation are implemented by the system designers.

Within a level of automation, the control distribution is usually not very precise, but can have a certain variety; therefore, these areas are visualized by small bars in the diagram (e.g. Fig. 7). Furthermore, only one level of automation can be active; the so-called current control authority is indicated by a solid border around the bars. The non-active levels of automation resemble potential control authority and are indicated by a dashed border around the bars.

The authority to change the control distribution is indicated by arrows that symbolize the scope and direction in which human (top arrow) and machine (bottom arrow) are allowed to change the control distribution. In this example (Fig. 7), the human is allowed to change the control distribution (for both human and machine) in both directions (indicated by solid arrow), while the machine is only allowed to propose a change in control (indicated by dashed arrow), but not to change the control distribution directly.

In the example of Fig. 8, a situation is shown where the human has no ability to cope with a situation, for example, due to limited resources. An example would be a suddenly occurring situation in which the human cannot react quickly enough. Here, the machine may have the control change authority to higher levels of automation (blue

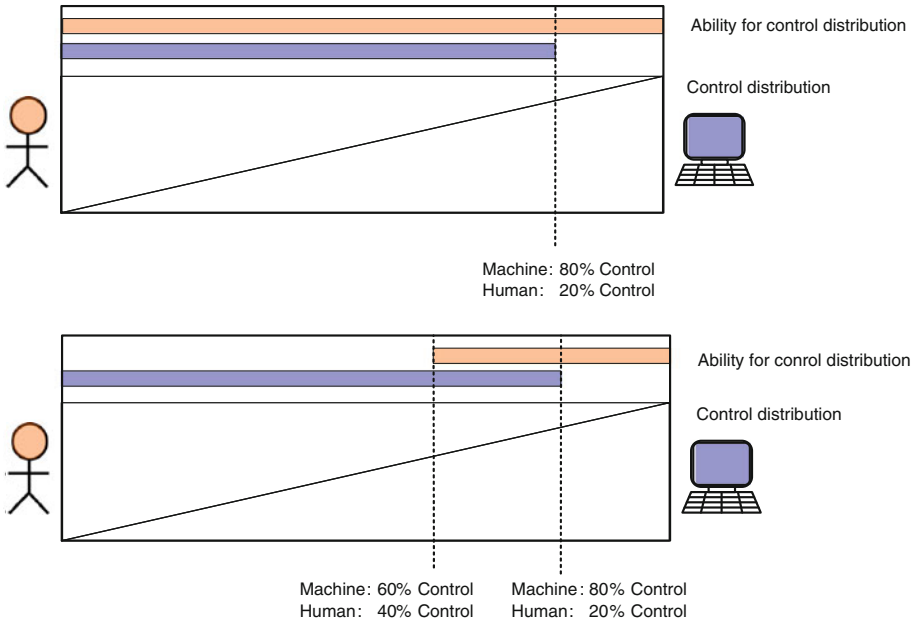


**Fig. 6** Abilities (to handle certain control distributions) in assistance and automation spectrum. The bars on the top resemble the area of possible control distributions on the spectrum

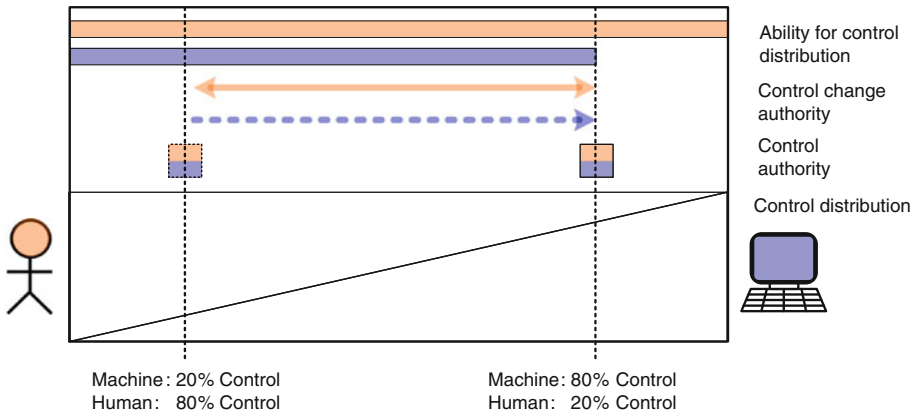


**Fig. 7** Authorities to change control distribution

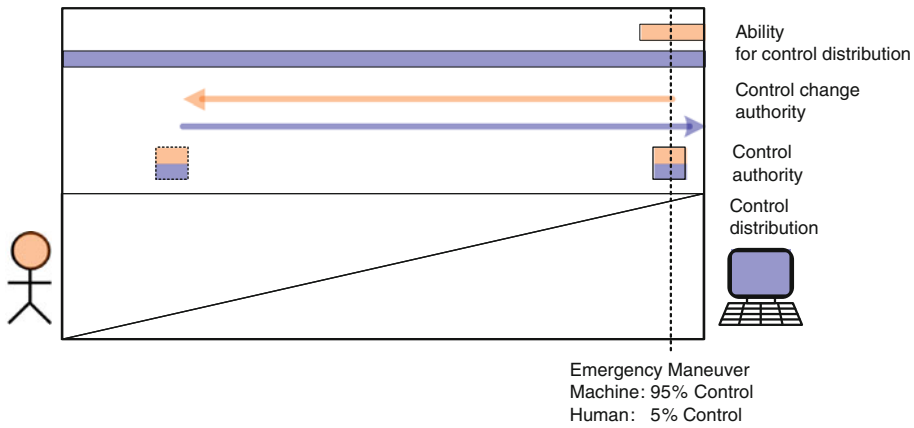


**Fig. 8** Authorities to change control distributions, for example emergency situation

arrow), whereas the human has only the control change authority downwards to lower levels of automation.

The actual distribution of control can be visualized by vertical lines in the assistance and automation spectrum. Ideally, the actual control distributions meet at the border of the two diagonals of human and machine and thus add up to 100%, as shown in Fig. 9. However, for example, a lack of ability (by human and/or machine) could cause a

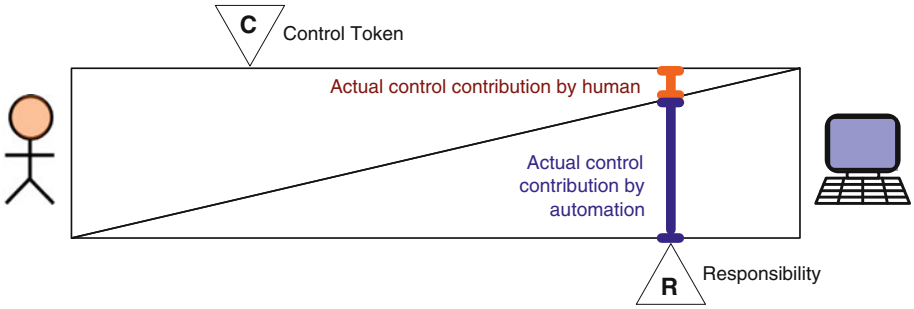


**Fig. 9** Responsibility, control token and actual control in the assistance and automation scale, here with an inconsistency between control token and actual control

smaller or larger actual control than is desired and/or necessary.

It can be helpful to distinguish between the actual control, which a most objective observer from outside would determine, and notional control, which is yet to be established. In this tension field between actual and notional (a term that goes back to a concept by Schutte and Goodrich 2007), a control token can be a representation of the notional control. Just like in mediaeval times a crown or a sceptre indicated authority, responsibility and power, a control token can be understood as symbol for the notional or desired distribution of control between two actors. Control tokens are not the control itself, but are a representation of the notional control that points towards the actual control. An example for a control token is the graphical marker of who is in control in an automation display. The location of the control token can be applied in the diagram as well. It is symbolized by the "C" marker. In certain control situations, it can make sense to split up the control token and differentiate between an explicit display of control and an action for the exchange of control. An example for this would be a situation, where the human does an action for the exchange of control, like pressing a button, and takes this already for the actual exchange of control, without realizing that the machine might not be able to actually accept and execute the control.

The responsibilities of human and automation can also be visualized in the assistance and automation scale, see Fig. 9, where a marker "R" indicates the responsibility distribution or shared responsibility. In this instantiation, the people and/or organizations behind the machine carry a majority of the responsibility, while the human operator carries a minority. It is important to note here that after the fact, it is quite common to use a numerical description of responsibility (e.g. 20–80%) such as in law suites regarding the sharing of the penalty between operator, operator's organization and manufacturer of the machine. However, a priori, the distribution of responsibility is hardly a crisp number, but often described in linguistic terms. A quite common distribution of responsibility is that (the humans behind) the machines (e.g. the developers) are responsible for a correct behaviour within the state of the art described, e.g., in standards, while the human operator is responsible for a correct use of the machine, e.g., as described in the manual. Even if the a priori responsibility might be fuzzy, (it makes sense) it makes nevertheless sense to think about this already in the design phase of the human machine system.

All the elements discussed above can now be combined to an ability–authority–control–responsibility diagram or A2CR diagram, which can be used as a tool to analyse and design human–machine systems with consistent relations between the cornerstone concepts of ability, authority, responsibility and control (Fig. 10—top). This diagram can be merged to a more compacted diagram (Fig. 10—bottom).

## 5 Consistency between ability, authority, control and responsibility (A2CR consistency)

The distribution of responsibility and authority and the control changes over times can be designed in many different ways, but it is highly desirable to ensure certain principles. Miller and Parasuraman (2007), for example, demands that "human–machine systems must be designed for an appropriate relationship, allowing both parties to share responsibility, authority, and autonomy in a safe, efficient, and reliable fashion." This relates to other interaction guidelines such as "the human must be at the locus of control" (Inagaki 2003) or "the human must be maintained as the final authority over the automation" (e.g. Inagaki 2003).

In the context of authority, ability, control and responsibility, we would like to emphasize a quality that connects these four cornerstone concepts, which we call consistency of authority, ability, control and responsibility in a human–machine system, or if an abbreviation is needed, A2CR consistency. A2CR consistency means that the double and triple binds between ability, authority, responsibility and control are respected, e.g., that there is not more responsibility than would be feasible with the authority and ability, that there is enough ability for a given authority, that the control is done by the partner with enough ability

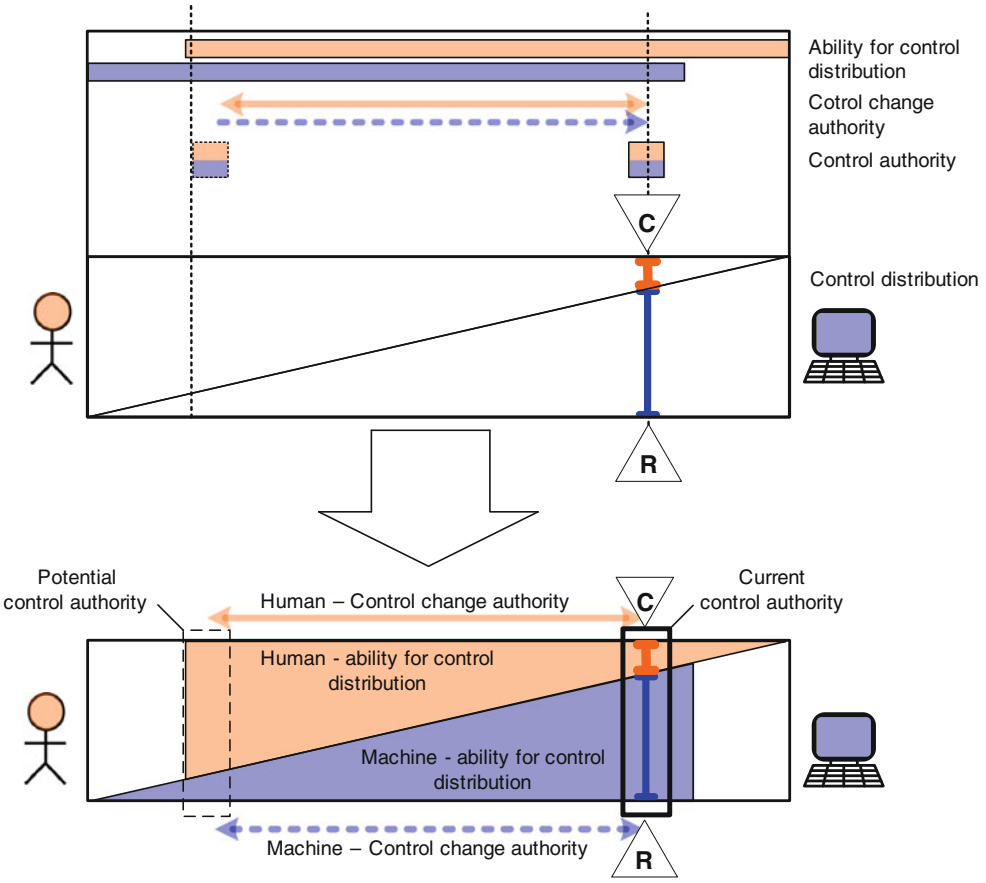


**Fig. 10** Evolution of a merged A2CR diagram

and authority and that more responsibility is carried by the actor or his representatives who had more control.

The goal of consistency is not achieved automatically, but rather constitutes a design paradigm for the system design including the interaction design. The chance for a high A2CR consistency can be ensured by a proper interaction design process in the development phase of the technical system, see Fig. 2. If this consistency is violated, tension fields might build up that could lead to negative results. An extreme would be an automation that does the control task completely, but where the human would keep all the responsibility.

The concepts of ability, authority, responsibility and control are major cornerstones to understand the operation of an automated and/or cooperative human–machine system. It is important to stress again that the most critical aspects of the double, triple and quadruple binds, which are subsummized here as A2CR consistency, are determined outside of the human–machine system in the meta-system. This is done usually before and after the operations, e.g. during the development or in an after-the-fact evaluation, e.g. in the case of an accident, as already shown in Fig. 2 at the beginning of this paper. An important feedback loop is running via the meta-system, where experience is used to change the ability, authority, control and responsibility configuration in a human–machine system.

## 6 Ability, authority, control and responsibility applied to cooperative control of (highly automated) vehicles

In the following text, the analysis of the relationship between ability, authority, responsibility and control as introduced above is exemplified with two driver assistance and automation systems that were developed in the project HAVEit that is heavily influenced by the base-research project H(orse)-Mode.

In the H-Mode projects, which originated at NASA Langley and span from DLR, Technical University of Munich and RWTH Technical University Aachen, a haptic-multimodal interaction for highly automated air and ground vehicles (H-Mode) is developed and applied to test vehicles (e.g. Kelsch et al. 2006; Goodrich et al. 2006; Heesen et al. 2010). Based on these base-research activities, EU projects like HAVEit (Highly Automated Vehicles for Intelligent Transport) bring these concepts closer to the application in serial cars and trucks (see e.g. Hoeger et al. 2008 or Flemisch et al 2008). Together with other research activities like

Conduct-by-wire (Winner et al. 2006), general concept of cooperative (guidance and) control can be formulated and applied to all moving machines like cars, trucks, airplanes, helicopters or even the teleoperation of robots (Fig. 1).

In HAVEit, the basic idea that vehicle control can be shared between human and a co-automation was applied as a dynamic task repartition (see e.g. Flemisch et al. 2010; Flemisch and Schieben 2010). Three distinct modes of different control distributions, lowly automated (or assisted), semi-automated (here: ACC) and highly automated, have been implemented.

The example in Fig. 11 resembles a normal driving situation with the control distribution of the automation level highly automated. In general, both driver and automation have full ability to handle all possible control distributions between 100% driver (manual driving) and 100% automation (fully automated driving). Three areas of control distribution have been defined by the system designers. In this example, only the driver has the control change authority between the three possible areas of control authority. Here, the chosen automation level is highly automated as indicated in the automation display on the right and indicated by the control token. The co-automation has no control change authority but has the authority to suggest other control distributions.

In the second example (Fig. 12), due to a sensor/environment degradation, the ability of the automation does not cover the whole spectrum, so that the control distribution of the highly automated mode is not available. This is also indicated in the automation display (highly automated is not highlighted). Here, the driver has only the control change authority between the two remaining modes, semi-automated driving and driver assisted. In this example, semi-automated is activated, and driver assisted is still available.

Figure 13 visualizes an emergency situation to exemplify a possible change in authorities depending on the abilities to handle the given control task (of driving the vehicle) in the current situation. The situation is critical such that the ability of the human to control the vehicle has decreased dramatically because his reaction time would be too long. A similar situation occurs in case the driver falls asleep or is otherwise impaired. As a consequence, the co-automation has received a higher control authority and also, in this emergency case only, the control change authority. In the example shown in Fig. 13, the automation has shifted the control token to emergency, i.e. fully automated, and has taken over control to resolve the situation. The human still has the authority to take over control again.

Note that in this example, some A2CR inconsistency is consciously accepted: The human driver retains the control change authority, even though his ability has diminished in this situation. This design choice was made to abide by current liability and regulatory legislation, which requires that the driver can always override interventions by the automation.

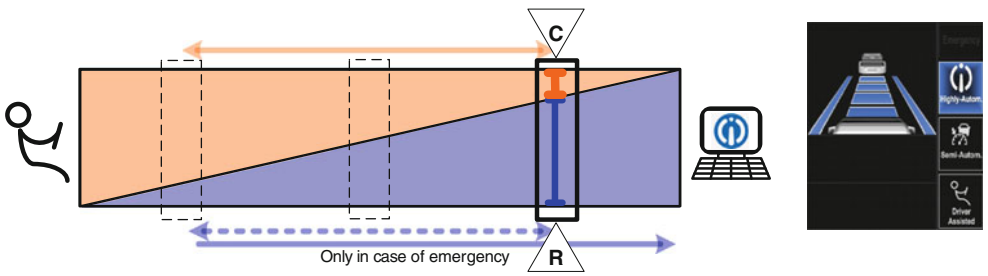


**Fig. 11** *Left* Ability, authority, responsibility and control in highly automated driving. Example HAVEit. *Right* Corresponding automation display in the research vehicle FAS Car

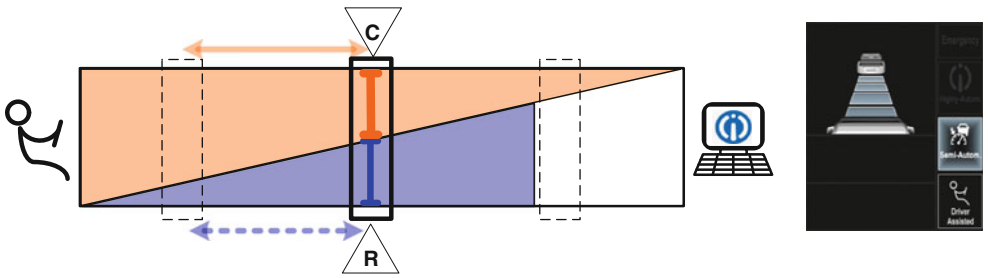


**Fig. 12** Ability, authority, responsibility and control for semi-automated driving while highly automated driving is not available (example HAVEit). *Right* Corresponding automation display in the research vehicle FAS Car

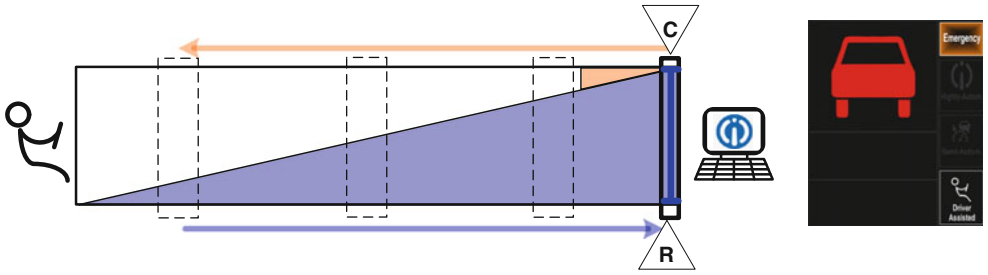


**Fig. 13** Emergency situation in HAVEit. *Right* Corresponding automation display in the research vehicle FAS Car

## 7 Consistency of mental models and transitions of control

**In general**, the information about authority, ability, responsibility and control is usually embedded in the system itself. Humans as subsystems of the system have an implicit or explicit mental model (or system image as Norman (1990) calls it) of the human–machine system, including authority, ability, responsibility and control. Summarizing several definitions of mental models, Wilson and Rutherford (1989) stated that a mental model can be seen as "a representation formed by a user of a system and/or a task based on previous experience as well as current observation, which provides most (if not all) of their subsequent system understanding and consequently dictates the level of task performance." Part of this mental model is already present when humans enter a control situation; other parts are built up and maintained in the flow of control situations.

Machines as subsystems also have information about authority, ability, responsibility and control embedded in them. This can be implicitly, e.g. in the way how these machines are constructed or designed, or explicitly, as internal "mental" models. In the following text, "mental" is used also for machines without quotation marks, even if machines are quite different regarding their mental capacities and characteristics. The explicit mental model of the machine can be as simple as a variable in a computer program "who is in control" or" is an ability available or degraded," or it can be more complex like an explicit storybook embedded in artificial players in computer games.

Figure 14 shows the example of a control distribution between one human and one computer, where each of the two partners has an understanding of where on the control scale the human–machine system is in the moment. The figure shows a specific situation of inconsistent mental models that occurs because the human thinks that the

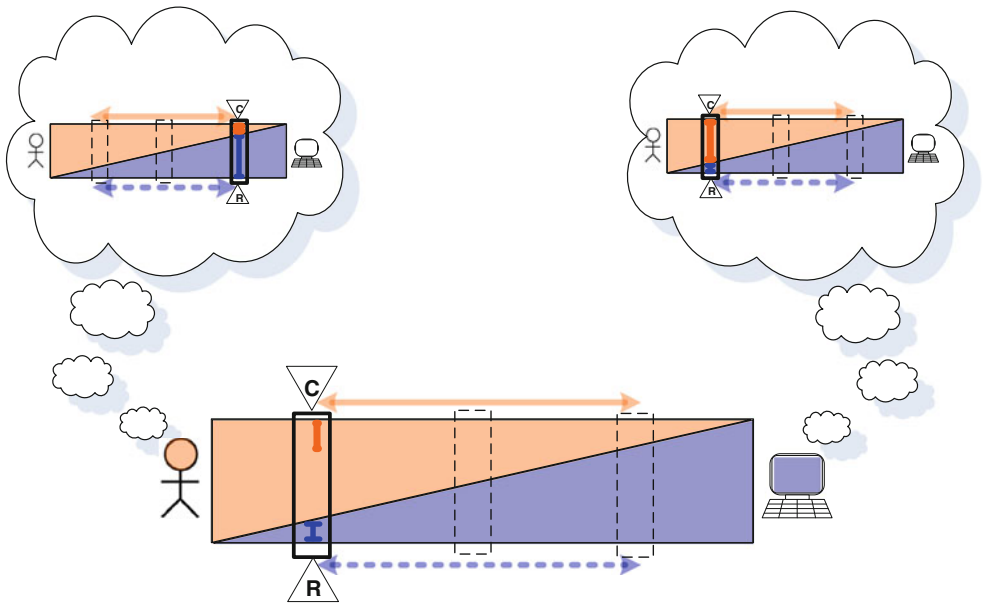


**Fig. 14** Mental models of human and automation, here an inconsistent example

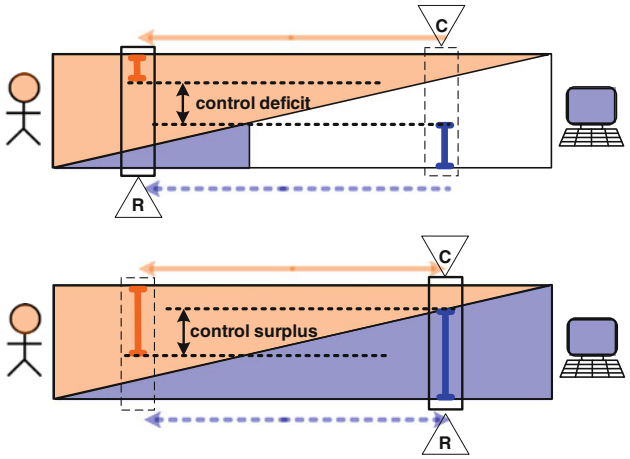


**Fig. 15** *Top* Deficit of actual control, e.g. in case of a refused transition. *Bottom* Surplus of actual control, e.g. in case of a missed transition

automation is in stronger control, while the automation "thinks" that the human is in stronger control (see also Fig. 15). This can be interpreted as a lack in mode awareness, which might lead to a critical system state due to the control deficit that is present (see Fig. 15).

The model of the machine that the human builds up is influenced by written manuals documenting the range of ability, authority and responsibility of the other actors on the control and is influenced by the human's experience with the system in different situations. The model of the human in the machine is mainly predefined by the programmer of the machine by setting the parameters of human's authority, ability and responsibility.

One of the keys to successful combinations of humans and machines is the consistency and compatibility of mental models about the ability, authority, control and responsibility of the partner. Control is one of the most prominent factors, a proper understanding or situation awareness about who is in control (control SA) is important for a proper functioning of a cooperative control situation.

Figure 15 top shows a situation where the human thinks that the machine is in control, while the machine thinks that the human is in control. If both act on their mental model, a lack of control or control deficit results. Figure 15 bottom shows the other extreme: Both actors think that they have control and act based on this belief, causing a control surplus that can result in undesired effects like conflicts between human and automation.

Similar aspects can be true for ability, authority and responsibility: A proper implicit or explicit mental model of the actors in a system about who can do and is allowed to do what, and who has to take which responsibility, can make a real difference between success and failure. Besides the necessity to respect the authority, ability and responsibility of the human in the design of the machine subsystem (implicit "mental model"), it becomes increasingly possible to give machines an explicit "mental" model about their human partners in the system. The proper ways to use this "mental" model, e.g., for an adaptivity of the machine subsystem are yet to be explored.

## 8 From mental models to transitions in control

The cooperation within the system is not static, but can lead to dynamic changes, e.g., of qualities like authority, ability, responsibility and control between the actors. States and transitions are mental constructs to differentiate between phases of a system with more changes and phases of a system with fewer changes. A system is usually called to be in a certain state, if chosen parameters of the system do not change beyond a chosen threshold. A transition is the period of a system between two different states. Applied to the key qualities of a cooperative human–machine system, authority, ability, responsibility and control, it is the transitions in these qualities in which the system might be especially vulnerable. As described above, any change in the system state has also to be reflected in the mental model of the actors, and if this update of the mental model fails, this inconsistency can lead to undesirable situations. This applies especially to control and ability.

**In general**, transitions in control can be requested or initiated by any actor in the system if he has the appropriate *change control authority*. Transitions can be successful if the actors have the appropriate ability and control authority for the new control distribution. If this is not the case and either the ability or the control authority is not adequate, the transition is rejected by one of the partners. For the system stability, it can make a big difference whether an actor looses or drops control "silently" and does not check whether the transition can be accomplished successfully, or whether an actor explicitly requests another actor to take over control in time. Whenever there is a change in the ability of one actor, e.g. an actor is in control, degrades in its ability and cannot control the situation anymore, it is essential that other actors take over control in time before the system gets into an undesirable state (classified as mandatory transition by Goodrich and Boer (1999)). Another starting point for a transition in control can be when one of the actors wants to take control because the own ability is rated as more expedient and/or safe (classified as discretionary transition (Goodrich and Boer (1999)).

The concepts of authority, ability and responsibility also apply to transitions. Authority and ability to initiate, accept or refuse certain transitions, e.g. in the modes of an automation, can be given or embedded to an actor before the fact, responsibility about the transition can be asked after the fact.

**Applied to vehicles**, due to the increasing number, complexity and ability of assistance and automation, the consistency and compatibility between the mental models of human(s) and assistance/automation subsystems about ability, authority, control and responsibility becomes increasingly critical. Critical situations might occur especially during and shortly after transitions of control between the driver and the vehicle automation. In highly automated driving, a control surplus where both the driver and the automation influence the vehicle strongly mainly leads to a decreasing acceptance by the driver and can be handled relatively easy by an explicit transition towards a control distribution with higher control for the driver. Because without sufficient control the vehicle might crash, a control deficit, however, is more critical and has to be addressed with extra safeguards, in HAVEit described as interlocked transitions (Schieben et al. 2011).

In the EU project HAVEit, the *change control authority* of the co-system is restricted to specific situations. The co-system has the authority to initiate a transition of control towards the driver only in the case of environment changes that cannot be handled by the co-system (decrease in ability of the automation) and in case of detected driver drowsiness and distraction (due to responsibility issues). In addition, the co-system has the control change authority to initiate a transition to a higher level of automation in the case of an emergency braking situation (non-adequate ability of the driver) and in case the driver does not react to a takeover request after escalation alarms. In any case, the co-system does not just drop control, but in case the co-system cannot hand over control to the driver in time, a so-called Minimum Risk Manoeuvre is initiated, brings the vehicle to a safe stop and stays there until the driver takes over again. In all other cases, the co-system's change control authority is restricted to propose another control distribution but not to actively change it.

To avoid mode confusion and mode error, all transitions in HAVEit follow general interaction schemes. For all transitions, the concept of interlocked transitions of control is applied. Interlocked means that transitions in control are only regarded successful, when there is clear information for the actor initiating the transition that the other actor has incorporated the transition as well. Applied to the transition of control from the co-system to the driver, this means that the co-system is only withdrawing from the control loop, if there are clear signs that the driver has taken over. In HAVEit, these signs were the information that the driver has his hands on the steering wheel, is applying a force to the steering wheel and/or one of the pedals or pushes a button for a lower level of automation.

In the example of the highly automated HAVEit (Fig. 16), the system will soon enter a situation where the ability of the automation decreases due to system limits (Figs. 2, 16). A takeover request is started to bring the driver back in the control loop before the ability of the automation decreases. In a first step, the automation informs the driver via HMI, so that the driver is prepared to take over more control over the vehicle (Figs. 3, 16). In Fig. 16, this is indicated by a shift of the control token. As soon as the driver reacts to the takeover request, the automation transfers control to the driver, and the actual control as well as the responsibility is shifted to the new control distribution.

The transitions of control were investigated during the course of the HAVEit project. Automation-initiated transitions of control towards the driver in the case of drowsiness and detection were well understood and well accepted by the drivers. Different design variants of driver-initiated transitions triggered by inputs on the accelerator pedal, brake pedal or steering wheel were tested according to the mental model that the drivers could build up. All design variants for the transitions were well understood, but the in-depth analysis of the data showed that some transition designs were closer to the expectation of the drivers than others and revealed potential for improvement (Schieben et al. 2011). After the investigation in research simulators and vehicles, the general transition schemes were applied to the demonstrator vehicles of HAVEit (e.g. Flemisch et al. 2010), e.g. to Volkwagen and Volvo (Fig. 17).

## 9 Outlook: challenges for the future balance of authority, ability, responsibility and control in human machine systems

**Applied to vehicles**, the examples from HAVEit shown in this paper are just one of a couple of projects in the vehicle domain in 2011, where assistance and automation systems have the ability to take over major parts of the driving task, and where increasingly questions arise about the proper balance of abilities, authority, control and responsibility between the human driver and the automation represented by it's human engineers. First prototypes of driver–automation systems exist where a dynamic balance of abilities, authority, control and responsibility between the driver and vehicle assistance and automation systems can be experienced and investigated, with already promising results with respect to good performance and acceptance. However, many questions are still open regarding the proper balance, especially about the authority of the assistance and automation systems, e.g. in emergency situations. The transitions of control seem to be a hot spot of this dynamic balance and need further structuring and investigation, see e.g. (Schieben et al. 2011). When drivers and automation share abilities and authority and have different opinions about the proper behaviour, the negotiation and arbitration

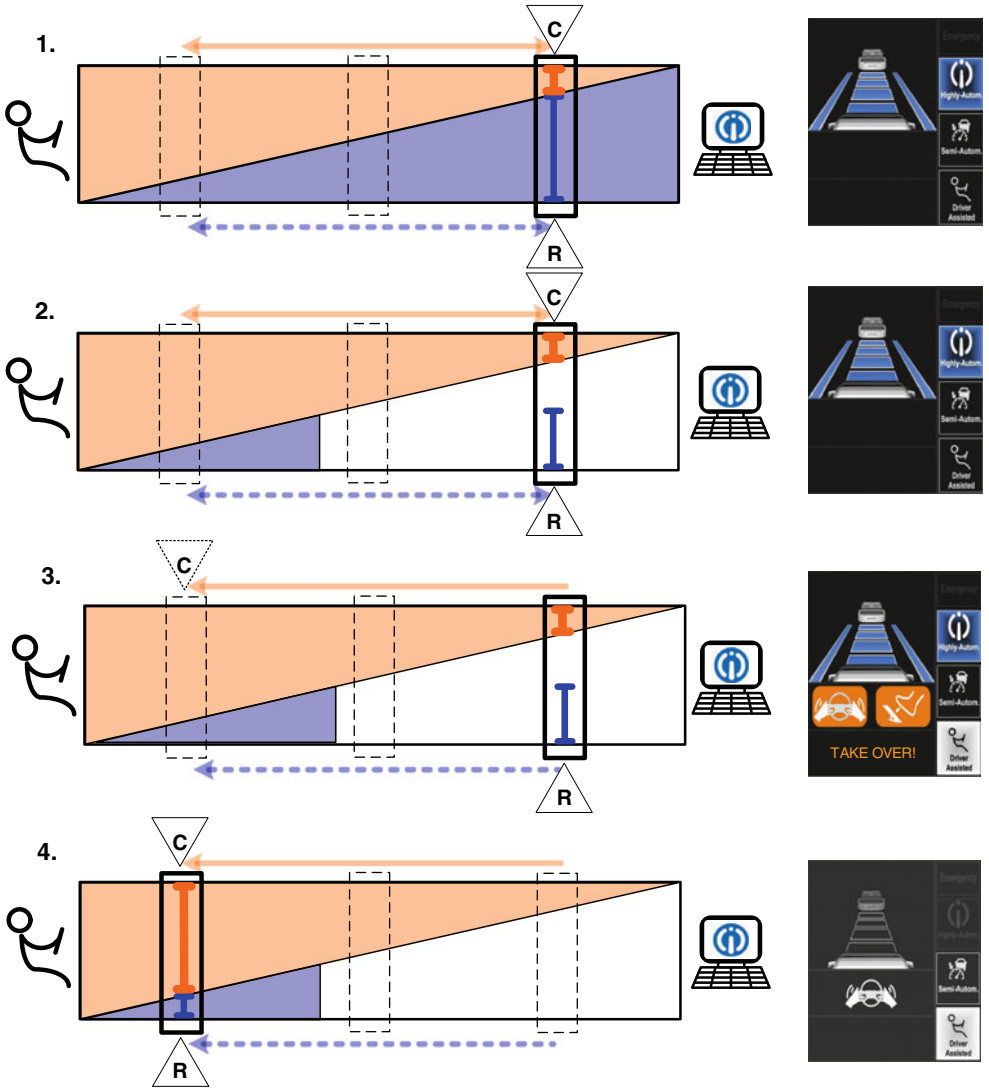


**Fig. 16** Example for a transition in automation mode due to a system limit of the co-automation (from the HAVEit project). In the steps 3 and 4, the "Driver Assisted" symbol in the automation display is flashing. On the right, the corresponding automation display in the research vehicle FAS Car

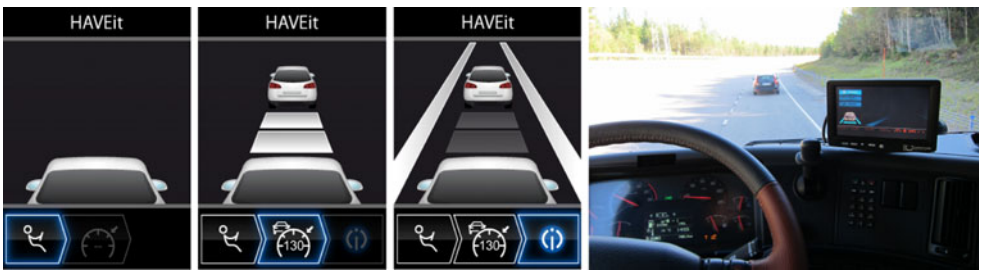


**Fig. 17** Assistance and automation modes in the Volkswagen HAVEit TAP (Temorary autopilot), adapted from Petermann and Schlag 2009, and Volvo

between the two partners becomes a critical aspect in the dynamic balance, see e.g. (Kelsch et al. 2006). In situations where the ability of a partner, e.g., of the automation can change dynamically, a preview of the ability into the future might be able to improve a successful dynamic balance, see e.g. (Heesen et al. 2010).

Only one part of these questions on the proper balance can be addressed with technical, cognitive and ergonomics sciences; other parts of these questions can be addressed with legal or ethical discussions including the society as a whole. In 2011, an increasingly intense discussion about these factors is being led in interdisciplinary working

groups, e.g. in Germany (Gasser et al. 2011, in preparation), or internationally, e.g. Burns 2011, in preparation. These questions do not only apply to vehicles, but to a much broader range of human–machine systems.

**In general**, the question on the proper dynamic balance of abilities, authority, control and responsibility between humans and increasingly capable technology is one of the core questions for any future human–machine system, and for any society. A consistent ontology of human–automation and easy-to-use techniques and tools are important prerequisites, for which this paper might be able to contribute some small pieces of the puzzle. If we take arguments like in Arthur (2009) serious that technology develops a dynamics of its own, the proper balance between humans and machines is not yet decided. On the one hand, this dynamic situation contains the risk of an imbalance of ability, authority, control and responsibility, which would leave the human with low abilities, low authority and insufficient control, but still with the full responsibility.

On the other hand, it contains the chance to combine the individual strength of the different partners, creating a fruitful symbiosis between humans and technology. Technology can play an important role, but still has to serve the human and should, as long as no other important, societally agreed values like human health or environmental aspects are too much at risk, leave the choice and the final authority to the human.

**Acknowledgments** The ideas presented in this paper were inspired by the projects on the H-Metaphor and H-Mode, which were funded by the US-National Research Council, NASA Langley and the Deutsche Forschungsgemeinschaft DFG. Extremely valuable support came from DLR, while the lead author was a research team leader on system ergonomics and design at DLR-ITS in Braunschweig. Further inspiration came from a Round Table on Human Automation Coagency for Collabarative Control during the IFAC HMS 2010 in Valenciennes. HAVEit was co-funded by the EU in FP7.



## References

Arthur B (2009) The nature of technology. What it is and how it evolves. Penguin, London

Bainbridge L (1983) Ironies of automation. Automatica 19(6):775–779

Billings CE (1997) Aviation automation: the search for a human-centered approach. Lawrence Erlbaum Associates, Mahwah

Dekker SWA (2002) Reconstructing the human contribution to accidents: the new view of human error and performance. J Safe Res 33(3):371–385

Dickmanns ED (2002) Vision for ground vehicles: history and prospects. Int J Veh Auton Syst 1(1):1–44

Endsley MR, Kiris EO (1995) The out-of-the-loop performance problem and level of control in automation. Hum Factors 37:381–394

FAA (1996) The interface between flightcrews and modern flight deck systems. Federal Aviation Administration, Washington, DC

Flemisch F, Schindler J, Kelsch J, Schieben A, Damböck D (2008) Some bridging methods towards a balanced design of human-machine systems, applied to highly automated vehicles. Applied ergonomics international conference, Las Vegas, USA

Flemisch F, Onken R (1999) The search for pilot's ideal complement, experimental results with the crew assistant military aircraft CAMA. Paper presented at the HCI International, Munich

Flemisch F, Schieben A (eds) (2010) Validation of preliminary design of HAVEit systems by simulation (Del. 33.3). Public deliverable to the EU-commission, Brussels, Belgium

Flemisch F, Adams A, Conway SR, Goodrich KH, Palmer MT, Schutte PC (2003) The H-metaphor as a guideline for vehicle automation and interaction. Technical Report NASA/TM–2003-212672. NASA, Hampton, VA

Flemisch F, Nashashibi F, Glaser S, Rauch N, Temme T, Resende P, Vanholme B, Schieben A, Löper C, Thomaidis G, Kaussner A (2010) Towards a highly automated driving: intermediate report on the HAVEIt-joint system. Transport Research Arena, Brussels

Goodrich MA, Boer ER (1999) Multiple mental models, automation strategies, and intelligent vehicle systems. In: Proceedings IEEE/IEEJ/JSAI conference on intelligent transportation systems: October 5–8, 1999, Tokio Japan: Tokyo, Japan

Goodrich K, Flemisch F, Schutte P, Williams R (2006) A design and interaction concept for aircraft with variable autonomy: application of the H-Mode. Paper presented at the 25th digital avionics systems conference. Portland, USA

Heesen M, Kelsch J, Löper C, Flemisch F (2010) Haptisch-multimodale Interaktion für hochautomatisierte, kooperative Fahrzeugführung bei Fahrstreifenwechsel-, Brems- und Ausweichmanövern 11. Braunschweiger Symposium Automatisierungs-, Assistenzsysteme und eingebettete Systeme für Transportmittel (AAET), Braunschweig

Hoeger R, Amditis A, Kunert M, Hoess A, Flemisch F, Krueger H-P, Bartels A (2008) Highly automated vehicles for intelligent transport: HAVEit approach. ITS World Congress, NY

Holzmann F, Flemisch F, Siegwart R, Bubb H (2006) From aviation down to vehicles—integration of a motions-envelope as safety technology. Paper presented at the SAE 2006 automotive dynamics stability and controls conference, Michigan, USA

Inagaki T (2003) Automation and the cost of authority. Int J Ind Ergon 31(3):169–174

Karasek R (1979) Job demands, job decision latitude, and mental strain: Implications for job redesign. Adm Sci Q 24:285–306

Kelsch J, Flemisch F, Löper C, Schieben A, Schindler J (2006) Links oder rechts, schneller oder langsamer? Grundlegende Fragestellungen beim cognitive systems engineering von hochautomatisierter Fahrzeugführung. Paper presented at the DGLR Fachauschusssitzung Anthropotechnik, Karlsruhe

Miller C (2005) Using delegation as an architecture for adaptive automation. Technical Report AFRL-HE-WP-TP-2005-0029. Airforce

Miller C, Parasuraman R (2007) Designing for flexible interaction between humans and automation: delegation interfaces for supervisory control. Hum Factors 49:57–75

Montemerlo M, Becker J et al (2008) Junior: the Stanford entry in the urban challenge. J Field Robot 25(9):569–597

Norman DA (1990) The design of everyday things. Currency, Doubleday, New York

Parasuraman R, Sheridan TB, Wickens CD (2000) A model for the types and levels of human interaction with automation. IEEE Trans Syst Man Cybern Part A: Syst Human 30(3):286–297

Petermann I, Schlag B (2009) Auswirkungen der Synthese von Assistenz und Automation auf das Fahrer-Fahrzeug System. Presented at the 11. Braunschweiger Symposium Automatisierungs-, Assistenzsysteme und eingebettete Systeme fuer Transportmittel (AAET), Braunschweig, Germany

Schieben A, Temme G, Koester F, Flemisch F (2011) How to interact with a highly automated vehicle—generic interaction design schemes and test results of a usability assessment. In: de Waard D, Gerard N, Onnasch L, Wiczorek R, Manzey D (eds) Human centred automation. Shaker Publishing, Maastricht, pp 251–266

Schutte PC, Goodrich KH, Cox DE, Jackson EB, Palmer MT, Pope AT, Schlecht RW, Tedjojuwono KK, Trujillo AC (2007) The naturalistic flight deck system: an integrated system concept for improved single-pilot operations. NASA/TM-2007-215090. NASA, Hampton, VA

Thrun S, Montemerlo M et al (2006) Stanley: the robot that won the DARPA Grand Challenge. J Field Robot 23(9):661–692

Wiener EL (1989) Human factors of advanced technology "Glass Flightdeck" transport aircraft. Technical Report NASA-CR-117528. NASA, Moffett Field

Wille JM, Saust F, Maurer M (2010). Stadtpilot: driving autonomously on Braunschweig's inner ring road. Intelligent vehicles symposium IV 2010 IEEE, pp 506–511

Wilson JR, Rutherford A (1989) Mental models: theory and application in human factors. Hum Factors 31(6):617–634

Winner H, Hakuli S, Bruder R, Konigorski U, Schiele B (2006) Conduct-by-Wire—ein neues Paradigma für die Weiterentwicklung der Fahrerassistenz. Workshop Fahrerassistenzsysteme 2006:112–125

Woods DD, Cook RI (2002) Nine steps to move forward from error. Cogn Tech Work 4(2):137–144