# Chapter 16

# Research Overview and General Discussion

As was stated in the introduction of this dissertation, the objective of the reported research was to investigate means for integrating knowledge of the human factor in human-computer cooperation into the reasoning capabilities of support systems. This is done to reduce the amount of problems caused by insufficient mutual understanding of the capabilities and limitations of humans and of support systems. The goal to increase reasoning capabilities of support systems was reached by incorporating executable cognitive models, which describe human cognition as accurately as possible, including its limitations, into these systems. Subsequently, these cognitive models are used to detect occurrences of limitations. Limitations were detected by the comparison of the output of two types of cognitive models: one that describes the current cognitive state (i.e., a *descriptive cognitive model*) and one that prescribes the desired cognitive state (i.e., a *prescriptive cognitive model*). Such limitation detections were then used as triggers for adaptation of the support to the human need for assistance, ideally resulting in an increase, or prevention of a decrease, of human-computer team performance. The specific adaptive support explored in this thesis focused on *adaptive autonomy* and *decision support*. The specific cognitive models explored in this thesis focused on *trust* and *attention*.

This chapter is composed of two sections: Section 1 is an overview of the research reported in this thesis. This is done by going through all methodological phases introduced in Chapter 1 and the chapters in which those phases were used. In this section also the most important conclusions and possible future research is discussed per chapter. Section 2 is a short general discussion of this thesis.

## 1 Research Overview

The research methodology outlined in Chapter 1 was used for pursuing the above stated research objective. In Table 1 it is shown which phases of the methodology were used in which chapters. The different phases are on the vertical axis. The different chapters of the

| | | Part II (Trust) | | | | | | Part III (Attention) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| a | Determination of domain and related Human Factors issues | + | + | − | − | − | − | + | − | + | − | − | − | − |
| b | Development of informal cognitive models | + | − | + | − | − | − | − | + | − | + | + | + | − |
| c | Psychological experimentation | +* | − | + | − | − | − | − | − | − | + | − | − | + |
| d | Formalization of cognitive models | + | − | − | + | + | − | − | + | − | +** | + | + | − |
| e | Verification of cognitive models | − | − | − | − | + | − | − | + | − | − | − | + | − |
| f | Validation and tuning of cognitive models | +* | − | − | − | + | − | − | − | + | + | + | − | − |
| g | Development of adaptive support system | +* | +* | − | + | − | + | +* | − | + | − | − | + | + |
| h | Verification of adaptive support system | − | − | − | + | − | − | − | − | − | − | − | + | − |
| i | Evaluation of adaptive support system | +* | +* | − | + | − | + | − | − | + | − | − | + | + |

* only the plans or preliminary results from this methodological phase were reported in the chapter
** methodological phase was used in the study, but was not reported in the chapter

Table 1: Overview of the methodological phases described in the different chapters of this thesis.

thesis (except the chapters in Part I, IV and V) are on the horizontal axis. A "+" indicates that a certain methodological phase was used in the study described in a corresponding chapter and a "−" that it was not used. Note that the reason for the table not being completely filled with pluses is not that the corresponding phases were impossible to be used, but rather that the focus in the particular chapters was not on those phases.

Since the application of the research methodology used in this thesis was yet to be explored, the exact content in Table 1 could not have been determined before the studies in the different chapters were performed. Now that these studies *have* been performed, the application of the method can be further evaluated and described. This is why the application of the method is further discussed in this chapter rather than in the introduction, where it was first introduced. Below the above mentioned description and evaluation is given, together with an overview of the main conclusions and future research.

(II) *Trust*

In Part II, two chapters used methodological phase a, two chapters phase b, two chapters phase c, three chapters phase d, one chapter phase e, two chapters phase f, four chapters phase g, one chapter phase h and four chapters phase i. Two experimental environments have been developed (a pattern learning and a classification task environment in Chapter 3 and 7, respectively) for which four types of support have been developed and evaluated (two in Chapter 6 and two in Chapter 8), using four different variants of cognitive models of trust (one in Chapter 3, one in Chapter 6 and two in Chapter 7).

(3) **Towards Task Allocation Decision Support by means of Cognitive Modeling of Trust**

*Research methodology:* First, in Chapter 3 the Human Factors issues related to trust and automation reliance were explored and discussed (a). Then informal descriptive and prescriptive cognitive models of attention were described (b) and formalized (d). In this chapter, descriptive trust was formalized as estimated trust of an agent $i$ in another agent $j$ concerning the execution of a certain action $\alpha$. Prescriptive trust was formalized as the estimated trust of an 'infallible agent $*$'. A first description was given of an adaptive support system that was able to reallocate tasks dynamically using cognitive models of trust (g$^*$)[1]. A design was described of an experiment with an implemented task environment (a *pattern learning task*, where people had to predict the next number out of 1, 2 and 3, given a certain pattern of past correct answers) to gain more insight into the Human Factors issues (c$^*$) and to validate the above mentioned cognitive model (f$^*$) and to evaluate the above mentioned support system (i$^*$).

*Main conclusions:* The results were of the exploratory kind and no definite conclusions could be drawn.

*Future research:* The described experimental environment should be used for further research on extensions of the proposed cognitive model of trust and dynamical task allocation, such as on indirect acquisition of knowledge (e.g., reputation, gossip), analogical judgments, allocation engagement costs (e.g., waiting, cooperation, and overhead costs), allocation implementation errors, level of autonomy, the allocation decision inhibitory bound, quantity and seriality of tasks and time pressure. It also is suggested that future research on cognitive modeling of trust should aim at support in the four stages of information processing (Parasuraman et al., 2000): the acquisition of information relevant for trust, its integration to trust concepts, task allocation decision making based on trust concepts and the implementation of the allocation decision.

(4) **Closed-Loop Adaptive Decision Support Based on Automated Trust Assessment**

*Research methodology:* In Chapter 4 the implemented task environment and Human Factors issues related to trust and automation reliance in Chapter 3 were further developed and explored, respectively (a). First descriptions of different support systems (g$^*$) and some preliminary evaluation results were presented and discussed (i$^*$). The support systems were variants of the in

[1]The marks "*" and "**" are related to the same footnotes as in Table 1.

Chapter 3 first described support system and augmented human cognition with respect to the human's cognitive task to calibrate trust and make reliance decisions. The goal of augmented cognition is to extend the human's cognitive performance via the development and use of computational technology, such as the envisioned support systems. The support systems had different autonomy settings: minimal autonomy, maximal autonomy and adaptive autonomy support. The *minimal autonomy support* assisted the human by giving advice related to trust and reliance decision making (called Operator Reliance Decision Making (Operator-RDM)), the *maximal autonomy support* took over reliance decision making (called the RDM Model (RDMM)) and *adaptive autonomy support* could dynamically decide between the two former support types (called Meta-RDMM).

*Main conclusions:* First results showed that human reliance decision making was not perfect and could be augmented by computational decision making. Maximal autonomy support (RDMM) turned out to be the best with respect to the human-computer team performance as compared to the other support types (Operator-RDM and Meta-RDMM).

*Future research:* It has been recommended that future research should focus on the investigation of how human-machine cooperation can be augmented in more complex and more realistic situations. It should be further explored whether models of trust and reliance can be practically used to adjust the level of autonomy of adaptive systems and in what domains this kind of support has an impact on the effectiveness of task performance, and how the magnitude of the impact depends on the task's and the domain's characteristics.

(5) **Reliance on Advice of Decision Aids: Order of Advice and Causes of Under-Reliance**

*Research methodology:* In Chapter 5 different established cognitive psychological theories and (informal) models about trust and reliance behavior were discussed (b) and several hypotheses related to the order of advice and the causes of mis-calibration of trust were tested in psychological experiments (c), using two further developed versions of the experimental environment introduced in Chapter 3 (the pattern learning task). The two versions were different with respect to the order of the advice given (i.e., either the advice of the human first or that of the support system).

*Main conclusions:* Several main conclusions could be drawn based on the results from this chapter. First of all, the results showed that a 'self bias' (i.e., an *a priori* tendency to trust oneself more than another, and the support system more specifically) can be observed. The results also showed that people disagree more with a support system when they express their decision before rather than after receiving advice from the support system. The results furthermore showed that this is only the case when decision makers trust themselves more than the support system. No self bias was found when trust in the support system exceeded trust in oneself. It was therefore argued that in existing frameworks of automation use, the notion of automation bias needs to be complemented with that of the self bias. Whether self biases lead to desirable outcomes or not, depends on whether perceptions of reliability of one's own performance and that of the support system are appropriate. When people wrongly think they perform better than the support system, self reliance can result in undesirable outcomes. The results showed that decision makers rely less on the support system than what would be expected based on relative trust in performance reliability (difference between trust in oneself and the other) alone: The participants did not rely more often on the support system, although they perceived it to be 30% more reliable. The results further suggested that decision makers rely less on conflicting advice because they perceive the advisor's reasoning to be cognitively less available and understandable than their own reasoning. The results showed that people who felt more responsible for the task outcome relied more on conflicting advice than people who feel less responsible. And finally, perceived reliability of both oneself and the support system was underestimated when feedback about performance was provided and it was found that negative experiences have a greater influence than positive experiences.

*Future research:* It was argued that appropriate reliance on support systems is not guaranteed when only focusing on optimizing the reliability of these systems. Several other things should also be done during the design phase: One of the important recommendations was that it might help when future support systems are able to give feedback about performance of humans and their support systems, but correct for the bias that negative information is given more weight. This feedback can improve the calibration of trust in oneself and the support system and therefore stimulate appropriate reliance and trust calibration. Secondly, by providing advice after, rather than before, more knowledge is brought to the task. Such a design would not be focused on reducing workload by automation, but focused on human-computer collaboration with the goal of increasing accuracy and resilience. Also, it was recommended to make people feel accountable for the outcomes of the human-computer team. That is, hold people responsible for the quality of outcome of the human-computer team. Finally, it was argued that one should control for the attribution of errors. For instance by making sources of error transparent or by making operators aware of their biases in attribution. The idea was that providing information regarding why the automation might be mistaken reduces inappropriate distrust (Dzindolet et al., 2003).

(6) **Aiding Human Reliance Decision Making Using Computational Models of Trust**

*Research methodology:* In Chapter 6 a more elaborate variant of the prescriptive cognitive model of trust introduced in Chapter 3 was formalized and tailored to the in Chapter 4 described support types (d). The second and third support types from Chapter 4 (maximal (RDMM) and adaptive autonomy support (Meta-RDMM)) were further developed and implemented (g). The dynamics of the support system were simulated for the purpose of verification (h) and validation (i). The general goal of the developed support system was to improve performance of human-computer teams either by taking over reliance decision making using trust models calibrated by the support system itself (RDMM), or by deciding adaptively when the human or the system makes the reliance decision (Meta-RDMM).

*Main conclusions:* Overall, the results showed that indeed calibration of trust and intervention by the computer can lead to an increase of human-computer team performance. The participants may have performed worse than (Meta-)RDMM because of limited attentional and memory resources and biases in weighing successes and failures of both themselves and the support system. The results showed a substantial amount of occurrences (above chance) in which humans made better reliance decisions than the support system. It was suggested that this could mean that reliance decision making completely done by the support system does not result in an optimal performance. This could be explained by the asymmetric availability of the underlying reasons for possible decreases of performance (i.e., human compared to support system performance) and the possibility of applying these reasons to the current situation. Meta-RDMM tried to take advantage of this, but without result.

*Future research:* Further extension of the model and exploration of the above mentioned principle behind Meta-RDMM was said to belong to the possible future research. Since the support systems have been simulated, one possibility to indeed find a significant effect of Meta-RDMM is to apply the support with a 'human in the loop', which might imply lower human performance degradation due to less problems with complacency as compared to RDMM when a large part of the task is taken over by the system.

(7) **Validation and Verification of Agent Models for Trust: Independent Compared to Relative Trust**

*Research methodology:* In Chapter 7 two variants of descriptive cognitive models of trust (the independent and relative trust model) were formalized (d), verified (e), validated (f) and compared to each other. A different experimental environment was used (a *classification task environment*). The *independent trust model* was inspired on, but different from, the formalized model in Chapter 6: the model now could estimate trust in three trustees and was used as a descriptive instead of a prescriptive model (the human was assumed to be similar as what the system would think an infallible agent would do). The difference between the independent and the *relative trust model* was that for the relative trust model the estimated human's trust in a certain trustee also depended on estimations for trustees that are considered competitors of that trustee (an additional modeled psychological phenomenon). The used experimental environment was contextually more rich than the environment introduced in Chapter 3 (the pattern learning task environment) and required cooperation between two humans and two computers. The task was now comparable to tasks in specific areas related to target identification based on video footage (but therefore also most probably less comparable to other specific areas not related to that).

*Main conclusions:* The results showed that both an independent as well as a relative trust model can predict reliance behavior with a high accuracy (72% and 80%, respectively). Furthermore, the results also showed that underlying assumptions of the trust models were found in the data of the participants (s.a. the underlying assumption that if on average more positive experiences of a trustee are identified, the advice of that trustee is also more often relied upon).

*Future research:* It was argued that future research should aim at exploring or extending other parameter adaptation methods for the purpose of real-time adaptation. Furthermore, it was mentioned that future research will focus on the development of support systems that monitor and balance the functional state of the human for optimal performance for all kinds of tasks in different domains, such as the military, aviation or air traffic control domain.

(8) **Effects of Reliance Support on Team Performance by Advising and Adaptive Autonomy**

*Research methodology:* In Chapter 8 two types of support systems (graphical and adaptive autonomy support) based on the second type of cognitive model of trust from Chapter 7 (for descriptive trust) and a variant of the model from Chapter 6 (for prescriptive trust) were developed (g), evaluated and compared to no support (i). The idea behind the *graphical support* was that trust calibration and reliance decision making was supported by an advice from the support system, whereas *adaptive autonomy support* could take over reliance decision making, using its own trust models.

*Main conclusions:* The results showed that team performance in the different support conditions was somewhat higher compared to no support. However, these differences were not significant. A significant increased effect was found for participants that performed less well. The results also showed significantly less satisfaction when applying adaptive autonomy compared to advising through the graphical support.

*Future research:* Future efforts should aim at investigating what precisely can go wrong when humans make reliance decisions, why this is such a difficult task for humans and how to provide leverage for exactly that. It was stressed that possible future variants of reliance decision support should aim at making the usage of the support less intrusive. Future improvement of the cognitive models of trust should also improve support systems based on those models. Research should also aim at investigating new efforts for taking away reasons for possible human intolerance for increased machine autonomy in making (important) decisions. Finally, it is mentioned that further research should investigate whether it is of benefit for adaptive team support to also include other psychological and environmental influences, such as analogical judgments and allocation engagement costs (as was already mentioned in Chapter 3).

(III) *Attention*

In Part III, two chapters used methodological phase a, four chapters phase b, two chapters phase c, four chapters phase d, two chapter phase e, three chapters phase f, four chapters phase g, one chapter phase h and three chapters phase i. Three experimental tasks have been developed (an air traffic control, naval tactical picture compilation and shooting game task environment in Chapter 9 (first two) and 13 (last one)) for which five types of support have been developed and evaluated (one in Chapter 11, one in Chapter 14 and three in Chapter 15), using seven different variants of cognitive models of attention (one in Chapter 10, three in Chapter 12, two in Chapter 13 and one in Chapter 14).

(9) **Augmented Meta-Cognition Addressing Dynamic Allocation of Tasks Requiring Visual Attention**

*Research methodology:* In Chapter 9 the different Human Factors issues related to the dynamic allocation of attention were explored and discussed (a). Furthermore, two preliminary descriptions of applications of attention model-based adaptive support were given ($g^*$). These descriptions were applied to two introduced experimental environments (i.e., an *air traffic control task* and a *tactical picture compilation task*). The envisioned support systems were able to dynamically allocate tasks based on the comparison between the estimation of the human's current (descriptive) and desired (prescriptive) attentional state.

*Main conclusions:* The results were of the exploratory kind and no definite conclusions could be drawn.

*Future research:* In this chapter it was stated that the Augmented Cognition Society defined 'Augmented Cognition' as "an emerging field of science that seeks to extend a user's abilities via computational technologies, which are explicitly designed to address bottlenecks, limitations, and biases in cognition and to improve decision making capabilities." It was furthermore mentioned that Augmented Cognition research is a wide area, that is applicable to various types of cognitive processes. As the area develops further, it may be useful to differentiate the field a bit more, for example, by distinguishing augmented cognition focusing on *task content* versus augmented cognition focusing on *task coordination*. As the latter is considered a form of meta-cognition (coordination of cognitive tasks), this suggests augmented meta-cognition as an interesting sub-area of future augmented cognition support systems. Especially in tasks involving multiple stimuli that require fast responses, this concept is expected to provide a substantial gain in effectiveness of system support systems.

(10) **Simulation and Formal Analysis of Visual Attention**

*Research methodology:* In Chapter 10 an informal (descriptive) model of attention and of different attentional states was described (b) and formalized (d). The model was described as being part of the design of an agent-based system (g*) that is able to monitor a human in the execution of the first task introduced in Chapter 9 (the air traffic control task). The output of the model was simulated using eye-tracker data from humans executing a the task and different expected properties of the model were verified against the simulation data (e).

*Main conclusions:* The model was specifically tailored to domain-dependent properties retrieved from a task environment; nevertheless it was expected that the method presented in the chapter remains generic enough to be easily applied to other domains and task environments. Furthermore, although the work reported focused on a practical application context, as a main contribution, also a formal analysis was given for attentional states and processes. Using this analysis, it has been proven that it is possible to identify different attentional states and processes, which can be used as additional triggers for adaptivity in support systems.

*Future research:* The study focused on formal analysis. Although in this formal analysis also empirical data were involved, a more systematic validation of the models put forward in the intended application context should be a next step. Future studies should further focus on the use of estimates of different attentional states and processes for dynamically allocating tasks as a means for assisting humans, as this kind of adaptive human-computer team support may turn out to be fruitful. Open questions are related to modeling both endogenous and exogenous triggers and their relation in one model. Finally, the attention model may be improved and refined by incorporating more attributes within saliency maps, for example based on literature (e.g., Itti and Koch, 2001; Itti et al., 1998; Sun, 2003).

(11) **Design and Validation of HABTA: Human Attention-Based Task Allocator**

*Research methodology:* In Chapter 11 it was further explored what kind of applications are needed given the found Human Factors issues related to over- and under-allocation of attention (a). Moreover, two experiments were described in which the cognitive model of attention from Chapter 10 was validated (f) and in which a developed adaptive attention allocation support system (g) was evaluated (i). The used experimental environment was based on the second environment introduced in Chapter 9. The support system was described as an adaptive cooperative agent assisting humans by managing its own and the human's attention. The component involved in the agent's attention management was called *HABTA: The Human-Attention-Based Task Allocator*.

*Main conclusions:* The results were of the exploratory kind and no definite conclusions could be drawn.

*Future research:* The results of both experiments presented in this chapter could be seen as a 'proof of concept' and large-scale experiments with multiple participants still needed to be performed. Furthermore, an idea was to compare the HABTA-component to the attention management capabilities of humans, where it is the human who allocates attention of himself or the support agent to different subtasks. In this way the effectiveness of HABTA-based support could be studied more convincingly. It is also stressed that future research should also focus on the development and validation of *prescriptive* cognitive models and not only on descriptive models: what would the system do when it were in the shoes of the human? Finally, in general, agent-components have more value when they can be easily adjusted for other applications. It was therefore said that it would be interesting to see whether HABTA-based support could be applied in other domains as well.

(12) **Effects of Task Performance and Task Complexity on the Validity of Computational Models of Attention**

*Research methodology:* In Chapter 12 three variants (the gaze-based, task-based and the combined model) of the attention model from Chapter 10 were described informally in relation to task complexity and performance (b), based on which several psychological hypotheses (c) as well as hypotheses related to the validity of models (f) were formed and experimentally tested. Before these models could be tested they had to be formalized, but this was not reported in this chapter because this was not the focus of the chapter ($d^{**}$). The *gaze-based model* only used the human's gaze data as input for the estimation of the human's attentional state, the *task-based model* only used information from the task and the *combined model* was a combination of the former two. The models were applied to the second task introduced in Chapter 9 (the tactical picture compilation task).

*Main conclusions:* The results showed that overall, the estimation of the combined model was better than that of the other two models. Contrary to what was expected, the performance of the models was not different for good and bad performers and was not different for simple and complex scenarios. The difference in complexity and performance might not have been strong enough.

*Future research:* It was mentioned that further research is needed to determine if improvement of the combined model is possible with additional features, such as the interpretation of mouse behavior or the inclusion of a more elaborate task model. This could be done using a similar validation methodology as was presented in this chapter. To enhance the performance of the models, optimal parameter values need to be determined. Furthermore, since the Area Under the Curve (AUC) performance measure is decision criterion-independent, it was mentioned that it needs to be determined whether liberal or conservative criterion settings are more effective for the estimation or prediction of human attentional states or whether this criterion should be determined dynamically.

(13) **Personalization of Computational Models of Attention by Simulated Annealing Parameter Tuning**

*Research methodology:* In Chapter 13 the cognitive model of attention from Chapter 10 was personalized. First, this personalization was described and motivated (b), after which the personalization process was formalized (d). The personalized models were tuned and validated using data from humans executing a *shooting game task* and compared to non-personalized models (f). Similar as in Chapter 7 about trust, the usage of other environments for experimentation was expected to lead to better understanding of the scalability and the further possibilities of using cognitive models in adaptive support systems. The personalization of the cognitive model of attention was done by tuning specific model parameters (using simulated annealing (SA)) that were related to certain human personality characteristics.

*Main conclusions:* Results showed that the attention model with personalization results in a more accurate estimation of an individual's attention as compared to the model without personalization.

*Future research:* The validation was subjective in the sense that a participant's own estimation was measured by asking to which objects they had directed their attention before certain freezes during the task execution. Future research should also focus on using objective measures. A possible way of measuring objective attention is by looking at mouse clicks at a location. It should be noted that SA is a probabilistic procedure and therefore is sub-optimal, specifically as the necessary computing capacity becomes relatively smaller compared to the problem space. In the future, personalization of attention models can be extended. In the personalized model presented in this chapter, parameters were tuned that are known to differ per individual. However, in future research personalization can be done by using collected data on personality to improve the attention model. Furthermore, in the current personalized model, parameters like the attention threshold and the total amount of attention were static. These could be coupled to a individual's functional state (e.g., experienced pressure, exhaustion), making the model fit for each individual, but also in different conditions (high or low workload). Such adjustments were expected to result in again an increase of the model's validity.

(14) **A System to Support Attention Allocation: Development and Application**

*Research methodology:* In Chapter 14 a description of a more elaborate version of the in Chapter 11 described attention allocation support system was given (g). This support system was based on four different models related to attention and the manipulation of human attention, which were first described (b) and then formalized (d). The first model described the human's current attentional state (as described in Chapter 10), the second was a model for beliefs about the human's attentional state, the third was a model to determine the discrepancy between the estimated current (descriptive) and normative (prescriptive) attentional state and the last was a model for the manipulation of the human's attention. Based on a simulation, several expected model properties of the above mentioned models (e) as well as the attention allocation support system were verified (h). Also, the support system was evaluated using performance data of humans executing a task (i). Like in Chapter 11, the task used was the tactical picture compilation task that was first introduced in Chapter 9.

*Main conclusions:* The participants reported to be confident that the agent's manipulation indeed was helpful. The results of the validation study with respect to performance improvement were positive. A detailed analysis and verification of the behavior of the agent also provided positive results: First, checking of the traces of the experiment confirmed that the agent was able to adapt the features of different objects in the task in such a way that they attracted human attention. The results furthermore showed that when there was a discrepancy between the prescriptive and the descriptive model of attention, the agent indeed was able to attract the human's attention.

*Future research:* Further investigation was needed to rule out possible order effects in the results of the described experiment, which suggests more research with more participants. It was also expected that future improvements of the agent's four sub-models, based on the gained knowledge from automated verification will also contribute to the improved success of such validation experiments. Top-down influences were not taken into account in the current models, but previous research shows that it is possible to extend such models based on saliency maps with top-down features of attention (see e.g., Elazari and Itty, 2010; Navalpakkam and Itti, 2002). As the presented attention model was based on the generic notion of features of a location, it could be easily extended with top-down features as well. In the future, these possibilities need to be explored in detail.

(15) **Adaptive Attention Allocation Support: Effects of System Conservativeness and Human Competence**

*Research methodology:* Finally, in Chapter 15 three variants (the fixed, liberal and conservative support) of the in Chapter 14 developed support were described (g), based on which several psychological hypotheses (c) as well as hypotheses related to the effectiveness of the support (i) were formed and experimentally tested. As in Chapter 14, all support types assisted humans in their allocation of attention. The variants of support were different with respect to their conservativeness (i.e., tendency to support). In *fixed support*, the system calculated an estimated optimal decision and suggested this to the human. In the other two support types, the system estimated the important information in the problem space in order to make a correct decision and directed the human's attention to this information. In *liberal support*, the system attempted to direct the human's attention using only the assessed task requirements, whereas in *conservative support*, the this attempt was done provided that it was estimated that the human was not already paying attention (more conservative).

*Main conclusions:* Overall results did not confirm our hypothesis that adaptive conservative support leads to the best performances. Furthermore, especially high-competent humans showed more trust in a system when delivered support was adapted to their specific needs.

*Future research:* Working with complex (support) systems can raise the cognitive load on the human, leaving less capacity to focus on the actual monitoring of contacts. Future design of adaptive support systems should therefore aim at keeping the system as simple as possible, though preserving the expected advantages of adaptivity. For the adaptive support investigated in this study, it was not possible for the human to simply follow suggestions of the support system. This was because, instead of suggesting a possible answer to a problem, only areas of interest were indicated by the system. This meant that, in any case, the proposed adaptive support must have eliminated inappropriate reliance on the support. It was therefore believed that the found results in the study were not a reason for rejecting this principle and therefore more research on adaptive attention allocation support was suggested, focusing on the *requirements* in which such a system can help to gain task performance.

As can be concluded from the above descriptions of the relation between the different chapters and the methodological phases described in Chapter 1, indeed all phases of the proposed methodology have successfully been used at least once for both trust and attention. This would suggest that the used research methodology indeed was usable given the stated research objective at the beginning of this thesis.

The general discussion about the implications emerging from the in this section summarized main conclusions and future research is held in Section 2.

# 2 General Discussion

As the collection of the main conclusions summarized in the previous section might suggest: one research question can generate multiple answers. In this thesis several examples have been explored of adaptive human-computer team support based on cognitive models of trust and attention. For this reason, one could argue that indeed the objective stated in the beginning of this chapter has been reached. But as the collection of future research summarized in the previous section might also suggest: one answer can generate multiple research questions. And for that reason, one could also argue that there is still a very long way to go.

The cognitive models explored in this thesis focused on *trust* and *attention*. But as was mentioned in the introduction, there are many more cognitive functions, concepts or processes that would be very good candidates for the purpose of adapting automated support to the human state and capabilities. Future research might as well aim at the development and use of cognitive models that can closely predict situation awareness, vigilance, mode awareness, automation-induced complacency, mental load, boredom, emotion, skill, experience, stress, self-confidence and commitment (to name but a few), and determine their characteristics in terms of for example demand for transparency, system autonomy, task switching costs, responsibility, 'human in the loop'-ness, delegation strategy and organization characteristics. Further investigation might also imply alternatives for on-line parameter tuning (s.a. usage of profiles), eye-trackers and mouse devices (s.a. pupil size (for detecting timing of decisions), EEG (s.a. usage of the P300), skin response (arousal, lying detection) and ECG (workload)). The use of such objective measures as input for cognitive models is expected to be very useful, but one should keep in mind that these models easily result in low construct validity (i.e., the degree to which one is indeed estimating the actual psychological phenomenon). Furthermore, one could presume that the discrimination of different more detailed cognitive states are the way to go: these more detailed states can help fine-tune the adaptations to the human need for assistance. But there is, of course, a limit to the value of adding more detail to cognitive models, given the fact that eventually one is estimating the state of a black box, as our knowledge of the underpinnings of the human mind is still limited. Finally, the models used in this thesis are used for adaptive decision support, but they might very well be useful for other kinds of applications, such as for the simulation of human cognition for, for instance, testing new interfaces or displays in expensive machines, such as aircraft (usability testing). In this example, cognitive models can be a cheap alternative for using the 'think aloud protocol' on well-paid pilots in simulators, which is also much more intrusive and time consuming.

The general advantage of the usage of cognitive models, as compared to behavioral or environmental models, as a basis for adaptive support systems, is that the detection of potential performance degradation or dangers can be done in an earlier stage. Behavioral or environmental models can only detect errors after the first signs of the underlying mistakes are observable, because no inference is made of what possible cognitive states might be causing these mistakes. An example is the pilot who relies on his automatic pilot while the current weather conditions are very bad. A support system is more likely to prevent an accident from happening when it infers that the pilot in fact is over-relying on his support system then when the first sign of a decreased altitude is detected. A disadvantage of cognitive models is that these models do not have a direct data source that can help in the inference of cognitive states (apart from for instance EEG, but still such sources are indirect). These sources do exist for behavioral and environmental models. For this reason, experimentally verified rules need to be identified that can substitute a direct data source for cognitive models. These rules are based on the fact that certain changes in the world can be antecedents for cognition and that cognition itself can be an antecedent of behavior. These two facts can be used to search for more specific behavioral and environmental data which help in the estimation of cognitive states and thereafter in the detection of limitations in human cognition.

A note on the scalability of this research. The reason for using different laboratory tasks was that the experiments can be controlled very well, participants could easily be measured (s.a. when using sensors like eye-trackers) and the experiment could be set up more easily, especially when multiple participants and computers were involved. But more realistic scenarios in which the results of these studies can scale up to real applications on, for instance, frigates or air traffic control towers, still need to be proven realizable. However, it is expected that the described studies and studies alike are a necessity when it comes to proper preparation for the further development of such systems.

A final note on the ethical implications of the research. It should be noted that the application of systems that are able to adapt to humans also need to *monitor* humans, *influence* their cognitive state and will *take over* tasks that formerly human beings were responsible for. These tasks can also be tasks that are about life and death. It is evident that such adaptive systems can have tremendous impact on society and, as a consequence, this should be subject for future ethical and political debates. Before technological advances can lead to the use of these adaptive support systems, both humans and systems should be ready for this: humans need to be ready on how to use and get used to such systems; and the systems need to be socially capable enough to take the human factor in human-computer cooperation into account, just like humans would do if they would stand in the shoes of the system (or even better than that).