Design and Validation of HABTA: Human Attention-Based Task Allocator

Peter-Paul van Maanen[1,2], Lisette de Koning[1], and Kees van Dongen[1]

[1] TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands
{peter-paul.vanmaanen, lisette.dekoning, kees.vandongen}@tno.nl
[2] Vrije Universiteit Amsterdam, De Boelelaan 1081a,
1081 HV Amsterdam, The Netherlands

Abstract. This paper addresses the development of an adaptive cooperative agent in a domain that suffers from human error in the allocation of attention. The design is discussed of a component of this adaptive agent, called Human Attention-Based Task Allocator (HABTA), capable of managing agent and human attention. The HABTA-component reallocates the human's and agent's focus of attention to tasks or objects based on an estimation of the current human allocation of attention and by comparison of this estimation with certain normative rules. The main contribution of the present paper is the description of the combined approach of design and validation for the development of such components. Two complementary experiments of validation of HABTA are described. The first experiment validates the model of human attention that is incorporated in HABTA, comparing estimations of the model with those of humans. The second experiment validates the HABTA-component itself, measuring its effect in terms of human-agent team performance, trust, and reliance. Finally, some intermediary results of the first experiment are shown, using human data in the domain of naval warfare.

1 Introduction

Several challenges can be identified for work on future naval platforms. Information volumes for navigation, system monitoring, and tactical tasks will increase as the complexity of the internal and external environment also increases [1]. The trend of reduced manning is expected to continue as a result of economic pressures and humans will be responsible for more tasks, tasks with increased load, and tasks with which they will have less experience. Problems with attention allocation are more likely to occur when more has to be done with less. To avoid these attention allocation problems, in this paper it is proposed that humans are supported by cooperative agents capable of managing their own and the human's allocation of attention. It is expected that these attention managers have a significant positive impact: when attentional switches between tasks or objects are often solicited, where the human's lack of experience with the environment makes it harder for them to select the appropriate attentional focus, or where an inappropriate selection of attentional focus may cause serious damage.

In domains like air traffic control (ATC) or naval tactical picture compilation these properties are found, even when the people involved are experienced.

The present study discusses the design and validation of a component of an adaptive agent, called *Human Attention-Based Task Allocator (HABTA)*, capable of managing agent and human attention. This component is based on two cognitive models: one that describes the current allocation of a humans attention and one that prescribes the way his attention should be allocated. If there is a discrepancy between the output of the two models, HABTA reallocates the tasks between the human and the agent, for instance depending on certain rules the human and agent agreed upon. Models of attention or situation awareness have already been developed and used to predict faults in attention allocation (e.g., the SEEV model [2]), but less is known about how they can be used to initiate agent adaptation, or automatic task reallocation more specifically. Furthermore, since in many domains (like ATC) it is the tasks altogether rather than mere visual stimuli that eventually require allocation of attention, the design and validation discussed in this paper is more focused on cognitive rather than visual attention. Off course the mentioned tasks also require visual attention, but all the time. Still other applied models mainly focus on visual attention. Finally, the applicability of a HABTA-based agent has not yet been investigated either.

This paper consists of the following sections. In Section 2 the psychological background of human error in the allocation of attention in the domain of naval warfare is shortly described. The understanding of these errors is important for the management of attention allocation. In Section 3 the design requirements of an agent-component *Human Attention-Based Task Allocator (HABTA)* are given. These requirements enable the agent to support the human-agent team by managing attention allocation of the human and the agent.

The main contribution of the present paper is the description of the combined approach of design and validation for the development of applied cooperative agent-components. In Section 4, two complementary methods of experimental validation against the in Section 3 stated design requirements are described. The first experiment validates the model of human attention that is incorporated in a HABTA-component. The validity of the model is determined by comparison of the model's and human's estimation of human attention allocation. The second experiment validates the HABTA-component itself, measuring its effect in terms of human-agent team performance, trust, and reliance. In Section 5 intermediary results of a pilot study are shown as a means to discuss the first experiment described in Section 4, using human data in the domain of naval warfare. In Section 6 the paper ends with concluding remarks and ideas for future research.

2 Human Error in the Allocation of Attention

As is mentioned in the introduction, the domain chosen in this research is naval warfare. One of the important tasks in naval warfare is the continuous compilation of a tactical picture of the situation (see for a description in more detail [3]). In a picture compilation task operators have to classify contacts that are represented on a radar display. The contacts can be classified as hostile, neutral or friendly, based on certain identification criteria (idcrits). Tactical picture compilation is known for its problems in the allocation of attention. To be able to identify contacts, contacts have to be monitored over time. This requires attention, but resources of attention are limited. When a task demands a lot of attention, less attentional resources are available for other tasks (e.g., [4, 5]. In general, two kinds of problems with human attention allocation can be distinguished: underallocation of attention and overallocation of attention.

Underallocation of attention means that tasks or objects that need attention do not receive enough attention from the operator. *Overallocation* of attention is the opposite: tasks or objects that do not need attention do receive attention. Overallocation of attention to one set of tasks may result in underattention to other tasks. Both under- and overallocation of attention can lead to errors. Experience, training, and interface design can improve these limitations, but only to a certain level. Efforts have been done, for example, to fuse tactical information on displays [6]. To be able to investigate whether a support system for attention allocation, like HABTA, can overcome these limitations of attention, it is important to understand these types of errors and more specifically in the domain of naval warfare. In Section 2.1 and 2.2, examples of errors of under- and overallocation when performing a tactical picture compilation task and their possible causes are described.

2.1 Underallocation of Attention

Underallocation of attention means that some objects or tasks receive less attention than they need according to certain normative rules for the task to be performed. Underallocation of attention occurs because of limited resources of attention or because of an incorrect assessment of the task.

When performing a tactical picture compilation task, operators have to monitor a radar screen where the surrounding contacts are represented as icons. The contacts on the screen have to be classified as neutral, hostile or friendly based on observed criteria. This is a complex task and it is essential that attention is allocated to the right objects. Inexperienced operators often allocate too little attention to contacts that they have previously classified as neutral [7]. When the behavior of these contacts changes to that of a hostile contact, this may not be observed because of underallocation of attention to those contacts. One reason for this could be that identity changes are not expected by the operator due to the fact that people are too confident in their identified contacts. Another reason might be that changes in relevant behavior of contacts are not salient enough to be observed without paying direct attention to those objects. Underallocation of attention to objects may also occur because of a lack of anticipatory thinking. This is the cognitive ability to prepare in time for problems and opportunities. In a picture compilation task, classification of contacts that are expected to come close to the own ship have priority over those that are not expected to come close. The reason for this is that there is less need to identify contacts when the own ship is out of sensor and weapon range of those contacts.

Therefore, inexperienced operators often direct their attention only to objects in the direction the ship is currently heading. When unexpected course change is needed because of emerging threats, the ship is sometimes headed toward an area with contacts that are not yet classified [7].

2.2 Overallocation of Attention

Apart from underallocation, overallocation of human attention is also a common problem. Overallocation of attention means that some objects receive more attention than needed according to certain normative rules. Overallocation of attention can occur for example, when operators overestimate the importance of a set of objects or tasks, while underestimating the importance of other objects or tasks. This occurs for example, when some contacts act like distractors and perform salient behavior. Comparable to visual search tasks where objects with salient features generate a pop-out effect (e.g., [8]), those contacts directly attract the attention of the operator (bottom-up). Especially inexperienced operators overrate those salient cues and allocate too much attention to those contacts [7]. Another possibility is that irrelevant behavior of objects is highly salient due to the manner information is presented on the interface. For instance, when a contact's behavior is unexpected, but not threatening, attention is unnecessarily drawn to this contact. In this case, the correct and quick application of identification rules will result in neutral identity and resources become available for the identification of other contacts.

3 Design Requirements

The goal of the efforts described is to come to a generic methodology for developing a component for an agent that supports humans with the appropriate allocation of attention in a domain that suffers from human error in the allocation of attention. As mentioned in Section 2, human attention allocation is prone to two types of errors with several possibilities as causes, such as inexperience and information overload.

In this section the design requirements of an agent-component is described that enables agents to determine whether objects or tasks that are required to receive attention indeed do receive attention, either by the human or the agent, and to intervene accordingly. The component is called an *Human Attention-Based Task Allocator (HABTA)*-component, since it *bases* its decisions to intervene on estimations of *human attention* and intervenes by *reallocating tasks* to either human or agent. It is expected that the combined task performance of the human-agent team will be optimized when the agent consists of such a HABTA-component. This work builds forth on earlier work. In [9] some of the possibilities are already discussed of dynamically triggering task allocation for tasks requiring visual attention, and in [10, 11] the real-time estimation of human attentional processes in the domain of naval warfare is already discussed.

Properly stated design requirements are important for the design of effective agent-systems for a certain purpose and for validating whether the design meets the requirements for that purpose. A HABTA-component has four design requirements, which are the following:

1. It should have a descriptive model, meaning an accurate model of what objects or tasks in the task environment receive the human's attention,
2. It should have a prescriptive or normative model, meaning an accurate model of what objects require attention for optimal task performance,
3. It should be able to reliably determine whether actual attention allocation differs too much from the required attention allocation,
4. It should be able to support by redirecting attention or by taking over tasks such that task performance is improved.

In Fig. 1 the design overview of a HABTA-component is shown that corresponds to the above design requirements. The setting in this particular overview

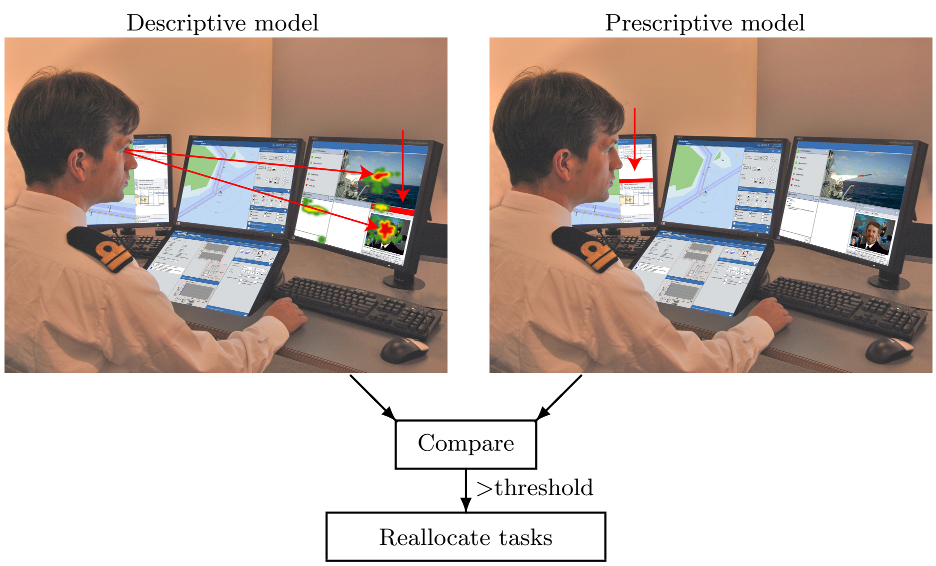

Fig. 1. Design overview of a HABTA-component for a future integrated command and control environment. The discrepancies between the output of the descriptive and prescriptive model result in a reallocation of tasks. The workstation shown in the pictures is the Basic-T [12].

is a naval officer behind an advanced future integrated command and control workstation and compiling a tactical picture of the situation. If the agent cooperatively assists the officer, than the agent should have a descriptive (Requirement 1) and normative model (Requirement 2). When the operator allocates his attention to certain objects or tasks that also require to receive attention, the outcome of both models should be comparable. This means that output of the models should not differ more than a certain threshold. The output of the two models in the example shown in Fig. 1 are clearly different: in the left image, the operator is attending to different objects and corresponding tasks than the right image indicates as being required (see arrows). Because of this discrepancy, which the HABTA-component should be able to determine (Requirement 3), an adaptive reaction by the agent is triggered (Requirement 4). This means that, for instance, the agent either will draw attention to the proper region or task through the workstation, or it will allocate its own attention to this region and starts executing the tasks related to that region, for the given situation.

To prevent that HABTA-based support results in automation surprises, the human-agent team should be able to make and adjust agreements about how they work as a team. It may be, for example, that the human does not want to be disturbed, and the agent is supposed to allocate tasks solely to itself. This option requires a higher form of autonomous task execution by the agent. The other possibility is that the human wants to stay in control as much as possible and therefore only wants to be alerted by the agent to attend to a certain region or execute a certain task. The choice of the agent's autonomy or assertiveness can also depend on a certain estimate of the urgency for reallocating tasks. In the case of tactical picture compilation, human and agent should agree on whether the agent is allowed to take over identification tasks for contacts that are overlooked or not.

On the one hand, the human may be preferred to be dealing with an arbitrary region or task, because the human may have certain relevant background knowledge the agent does not have. But on the other hand, the human is not preferred to be allocated to all objects or tasks at once, because, in a complex scenario, he has limited attentional resources. Hence humans cannot be in complete control, given the fact that both human and agent need each other. Optimal performance is only reached when human and agent work together as a team. Human-agent team work is expected to be effective when the right support is provided at the right time and in the right way. An obvious goal, but there are some potential obstacles in achieving it. Descriptive and prescriptive (normative) models of attention allocation may be inaccurate. Objects that require or receive attention may not be in the output of the descriptive or normative models, respectively. Similarly, objects that do not require or receive attention may be in the output of the models. The agent may conclude that descriptive and normative models differ when they do not, and vice versa. The system may be assertive and wrong, or withholding but right. Attention may be redirected to the wrong region or the wrong set of objects, or tasks are taken over by the agent that should be taken over by the human. Because of the complexity of these consequences of the above design requirements, both the validity of the model and the effectiveness of the agent's HABTA-component should be investigated and iteratively improved. This procedure of investigation and improvement is described in Section 4.

4 Validation

As described in Section 3, HABTA-components require a descriptive and prescriptive model of attention to support attention allocation of humans in complex tasks. Before HABTA-components can be used to support humans, they have to be validated. Validation is the process of determining the degree to which a (cognitive) model is an accurate description of human (cognitive) phenomena from the perspective of the intended use of the model. Again referring to Section 3, for the intended use mentioned in this paper, this means that HABTA-components have to meet the design requirements (1–4) in Section 3.

In the near future two experiments will be carried out to validate a HABTA-component. In Experiment 1 the descriptive model will be validated and optimized (Requirement 1). This experiment aims at determining the sensitivity (d') of the model by comparing it with data retrieved from human subjects executing a complex task that causes problems with attention allocation. Based on the results of the experiment, the d' of the model can be improved by optimizing it off-line against a random part of the same data. It is expected that the higher d', the better the HABTA-component will be able to support the human. If the d' of the descriptive model is not high enough, the HABTA-component will support at the wrong moments and for the wrong reasons, which obviously leads to low performance, trust, and acceptance. In Experiment 2 the applicability of the (improved) descriptive model for attention allocation support is tested (Requirements 2–4). It will be investigated if the support of an agent with the HABTA-component leads to better performance than without HABTA-component.

The remainder of this section is composed of three parts. In Section 4.1 the task that will be used in the above mentioned experiments is described in more detail. After that, the specific experimental design and measurements of the experiments are described in Sections 4.2 and 4.3, respectively. Both experiments still have to be carried out. Preliminary results from a pilot of Experiment 1 will be described in Section 5.

4.1 Task Description

The task used in both experiments is a simple version of the identification task described in [13] that has to be executed in order to buildup a tactical picture of the situation. In Fig. 2 a snapshot of the interface of the task environment is shown.[3] The goal is to identify the five most threatening contacts (ships). In order to do this, participants have to monitor a radar display where contacts in the surrounding areas are displayed. To determine if a contact is a possible threat, different criteria have to be used. These criteria are the identification criteria (idcrits) that are also used in naval warfare, but are simplified in order to let naive participants learn them more easily. These simplified criteria are the speed, heading, bearing, and distance of a contact to the own ship, and whether the contact is in a sea lane or not. When the participant clicks on a contact with

[3] A full color variant of Fig. 2 can be found at [14].

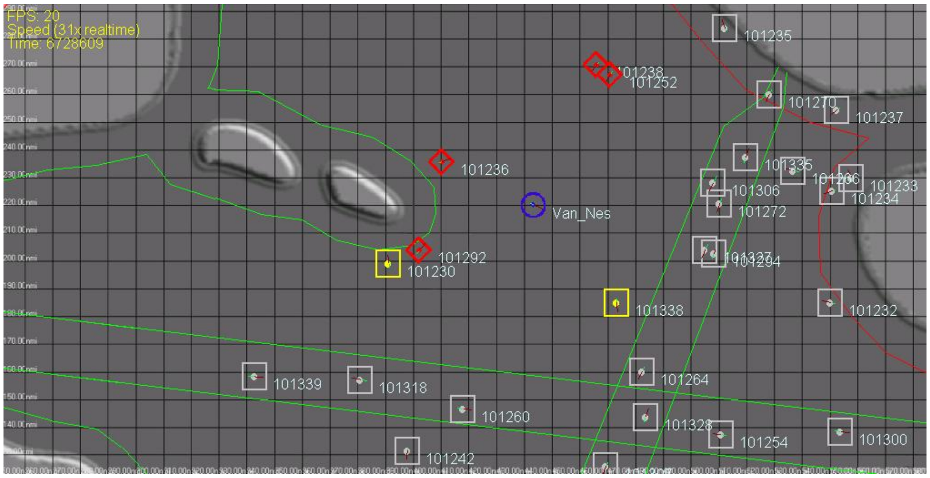

Fig. 2. The interface of the used simplified task environment based on [13]. The green lanes are sea lanes. The blue circle labeled with "Van_Nes" represents the own ship.

the right mouse button this information is displayed. If a participant concludes that a ship is a possible threat or not, he can change the color of the contacts by clicking with the left mouse button on the contact. Contacts can be identified as either a threat (red), possible threat (yellow), or no threat (green). It is not necessary that all contacts are identified. Only the five most threatening have to be identified as a threat (marked as red). The other types of identification (possible threat and no threat) are used to assist the participant in his task. When a contact is marked as green, this means no direct attention is needed. When a contact is marked as yellow, this contact has to be checked regularly to decide if the contact is still no threat. The task has to be performed as accurately as possible. Contacts that are wrongfully identified as a threat will result in a lower score. Performance is determined by the accurateness, averaged over time, of the contacts that are identified as the five most threatening contacts during the task. Behavior of each contact can change during the task and therefore the soundness of classifications (which is not communicated to the participant) may change over time. For instance, a contact can suddenly come closer to the own ship, get out of a sea lane, speedup, or change its bearing or heading.

For Experiment 2 (see Section 4.3) the task is extended to one that includes the support of the HABTA-based agent. The support agent is capable of doing the same as the operator, except with limited background knowledge and therefore limited performance per object. In order to simulate this aspect, for each participant, the measured average performance per contact in Experiment 1 is used in order to set the performance of the agent. The agent can be given a list of objects provided by the its HABTA-component and compile a tactical picture related to those objects.

4.2 Experiment 1: Validation of the Descriptive Model

In Experiment 1 participants perform the task as described in the previous section without support of the agent. The same scenario will be used for all participants. Before the actual task starts, the task will be explained thoroughly to the participants. The task will be illustrated by using different examples to be sure that the participants understand the task and how to decide if a contact is a threat based on the different criteria. All participants have to perform a test to check if they sufficiently understood the rules of classifying the contacts. If they do not perform well, i.e. their score is blow 80%, they receive extra instructions and another test. Also the possible second test has to be performed with a success rate of above or equal to 80%. Then they have to perform a practice trial in which they have to apply the learned rules. After this they get instructions of how to behave when there during the experimental interventions while they are executing their task. This is practiced as well several times, after which the actual experiment begins.

During the task, different variables are measured to determine the d' of the model and to be able to iteratively improve the model afterwards. The following variables will be measured: eye movements, performance, mental workload and at different points in time participants have to mark contacts that received attention according to the participant. The performance and mental workload measures are used as a baseline for comparing the performance and mental workload of the task with and without using the HABTA-component (see Experiment 2). In order to measure the variables, at random moments (varying from 4–6 minutes) the scenario is frozen. During a freeze, the participants have to click on the contacts to which, in their opinion, they had allocated their attention the moment right before the scenario was frozen. The participants also have to motivate why those contacts are selected. Directly after the participant has selected the contacts, mental workload is measured during the same freezes. For this, the mental workload scale from [15] is used (BSMI). On a scale from 0 (not at all strenuous) to 150 (very strenuous) the mental workload of the task has to be indicated. Performance and eye movements of the participants are measured during the task, by calculation according to the rules described in Section 4.1 and by eye-tracker recording, respectively. The patterns of the eye movements (what objects are looked at through time) are compared with the contacts that received attention before the freezes, according to the participant. This is done to be sure that the participants were able to select the objects that received their attention. Those contacts that got a considerate amount of gaze fixations, are expected to have received attention.[4] If the participants do not mention those contacts, it is expected that they are not good at selecting the proper contacts.

After the experiment is performed, the contacts selected by the participants during the freezes are matched with the output of the model in a simulation. The calculation of d' provides information about the sensitivity of the model, i.e. whether the model is able to accurately describe the participant's dynamics

[4] Note that this does not hold vice versa, which would otherwise mean that attention in complex scenarios is easily described using solely fixation data.

of attention allocation. Information about performance, workload, and the description of the participants why contacts are selected, is expected to be valuable for determining in what cases the percentage true positives (hits) is high and percentage false positives (false alarms) is low, which in turn can be used to improve the sensitivity of the model. See Section 5 for the illustration of this process.

4.3 Experiment 2: Validation of the HABTA-Based Support

In Experiment 2 the applicability of the model for supporting attention allocation is tested. The same task as in Experiment 1 has to be performed, except this time the participant is supported by the agent of which the HABTA-component is part of. When there is a discrepancy between the descriptive and prescriptive model, higher than a certain threshold (see Fig. 1), the agent will support the human by either performing the task for the participant or by drawing attention to the contact that should receive attention. Different variables are measured to determine the excess value of the HABTA-based support. Performance and mental workload are measured in the same way as in Experiment 1. Furthermore, trust and acceptance are measured at the end of the scenario.

In order to determine the effectiveness of an agent, it is also important to measure trust and acceptance of that agent and to investigate what factors influence trust and acceptance. Trust and acceptance indicate whether people will actually use the agent. For instance, it says something about whether people will follow the advice of the agent, in the case the agent provides advice. Validated questionnaires are adjusted to be able to measure trust and acceptance in adaptive systems. The trust questionnaire is based on the questionnaire of [16]. An example of a question on this questionnaire is: "Is the agent reliable enough?". The acceptance questionnaire is based on the questionnaire of [17] and [18]. An example of a question on this questionnaire is: "Is the support of the agent useful for me?". The trust and acceptance scores are expected to provide more insight in the results of the experiment. If trust in and acceptance of the agent is low, people will not follow any suggestions made by the agent.

The performance and mental workload without a HABTA-based agent will be compared with those with a HABTA-based agent, using the results of Experiment 1 as a baseline. This is one of the reasons that the same participants are used as in Experiment 1. The other reason is that the measured performance in Experiment 1 is used for setting the performance of the agent. For Experiment 2, it is expected that performance is higher and mental workload is lower when supported with HATBA.

5 Intermediary Results

In this section preliminary results of the experiments described in Section 4 are shown based on a pilot study for Experiment 1, using one arbitrary participant. The actual experiment will be performed with more participants. The pilot is primarily meant to explore the applicability of the experimental method of Experiment 1 to the given task. It is also meant as an illustration of the form and dynamics of the participant's and model's estimation of human allocation of attention. Finally, it is used as a basis for a better understanding of the possibilities of HABTA-based support, which is important for a proper preparation and performing of Experiment 2. This is because this type of support is required in the experimental setup of Experiment 2.

In the pilot study, the participant was required to execute the identification task and to select contacts during the freezes. In contrast with the procedure during the actual experiment, no questions concerning the participant's cognitive workload or motivation for the selected contacts were asked. In Fig. 2 the interface right before a freeze is shown. During a freeze both the participant and the model had to indicate their estimation of what contacts the attention of the participant was allocated to. In the situation presented in Fig.2, the participant selected contacts 101238, 101252, 101236, 101338, 101230, 101292, 101294, and 101327. Between every two freezes certain events can cause the participant to change the allocation of his attention to other attention demanding regions. The preceding course of events of the situation in Fig.2 clearly caused the participant to attend to the contacts close to his own ship "Van_Nes". If the model made a proper estimation of the participant's allocation of attention, the selected contacts by the participant would resemble those selected by the model. Consequently, the performance of the model is best determined by means of the calculation of the overall overlap of the participant's and model's selection of contacts. This calculation is explained below.

There are four possible outcomes when comparing the participant's and model's selection of contacts, namely, a Hit, False Alarm, Correct Rejection, and Miss. The counts of these outcomes can be set out in a 2×2 confusion matrix. Tab. 1 is such a confusion matrix, where T and F are the total amount of the participant's selected and not selected contacts, respectively, and T' and F' are the total amount of the model's selected and not selected contacts, respectively. The ratios of all the possible outcomes are represented by H, FA, CR,

		Participant		
		t	*f*	*total*
Model	t'	Hits	False Alarms	T'
	f'	Misses	Correct Rejections	F'
	total	T	F	

Table 1. Confusion matrix of the participant's and model's estimation of the allocation of attention.

and M, respectively. A higher H and CR, and a lower FA and M, leads to a more appropriate estimation by the model. This is the case because the selected contacts by the model then have a higher resemblance with those selected by the participant. Furthermore, a higher T' leads to a higher H, but, unfortunately, also to a higher FA. Something similar holds for F'. The value of T' therefore should depend on the trade-off between the costs and benefits of these different outcomes.

In Fig. 3 the $15 \times 10 \times 1$ output of the model for the situation presented in Fig. 2 is shown. If the estimated attention on the z-axis, called Attention Value (AV), is higher than a certain threshold, which is in this case set to .035, the contact is selected and otherwise it is not. The different values of AV are normally distributed over the (x, y)-plane. The threshold is dependent on the total amount of contacts the participant is expected to allocate attention to [10]. The AV-distribution in Fig. 3 results in the selection of contacts 101235, 101238, 101252, 101236, 101292, 101230, 101338, and 101260. Using this selection and the selected contacts by the participant, for each contact, the particular outcome can be determined. For each freeze, if one counts the number of the different outcomes, a confusion matrix can be constructed and the respective ratios can be calculated. For Fig. 3, for example, these ratios are $H = \frac{6}{8} = 0.750$, $FA = \frac{2}{19} = 0.105$, $CR = \frac{17}{19} = 0.895$, and $M = \frac{2}{8} = 0.250$, respectively.

To study the performance of models Receiver-Operating Characteristics (ROC) graphs are commonly used. A ROC-space is defined by FA as the x- and H as the y-axis, which depicts relative trade-offs between the costs and benefits of the model. Every (FA, H)-pair of each confusion matrix represents one point in the ROC-space. Since the model is intended to estimate the participant's allocation of attention for each freeze and participant, this means that for N participants and M freezes, there are NM points in the ROC-space.

Once all points have been scatter plotted in the ROC-space, a fit of an isosensitivity curve leads to an estimate of the d' of the model. Isosentitivity corresponds to:

$$d' = z(H) - z(FA)$$

where d' is constant along the curve and $z(x)$ is the z-score of x.[5] Larger absolute values of d' mean that the model is more specific and sensitive to the participant's estimation (and thus has a higher performance). If d' is near or below zero, this indicates the model's performance is equal to or below chance, respectively. If there does not exist a proper fit of a isosensitivity curve, the area under the curve (AUC) can also be used as a model validity estimate. In non-parametric statistics the ROC-graph is determined by the data and not by a predefined curve. If the different values of H and FA appear to be normally distributed, the d' can be obtained from a z-table. In this case, the (FA, H)-pair from Fig. 3 results in $d' = 1.927$. Which is a fairly good score.

[5] The z-score reveals how many units of the standard deviation a case is above or below the mean:

$$z(x_i) = \frac{x_i - \mu_x}{\sigma_x}$$

where μ_x is the mean, σ_x the standard deviation of the variable x, and x_i a raw score.

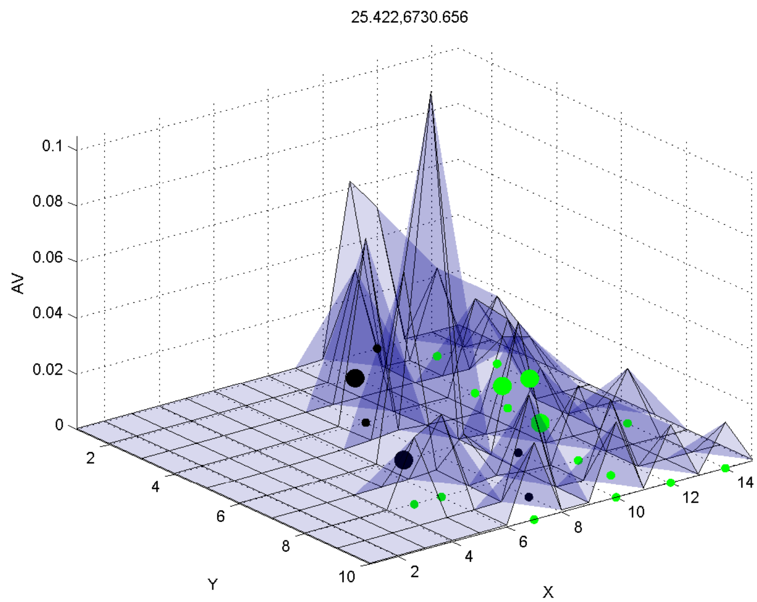

Fig. 3. The output of the model for the situation shown in Fig. 2. The black dots are the selected contacts by the model. Bigger dots mean that there are more contacts on the respective coordinates.

6 Conclusion and discussion

This paper describes the development of an adaptive cooperative agent to support humans while performing tasks where errors in the allocation of attention occur. In general, human attention allocation is prone to two types of errors: over- and underallocation of attention. Several factors may cause over- or underallocation of attention, such as inexperience and information overload. The design is discussed of a component of an agent, called Human Attention-Based Task Allocator (HABTA), that is capable of detecting human error in the allocation of attention and acts accordingly by reallocating tasks between the human and the agent. In this way the HABTA-based agent manages human and agent attention, causing the performance of the human-agent team to increase. The development of such an agent requires extensive and iterative research. The agent's internal structure, i.e. the models describing and prescribing human attention allocation and the support mechanism that is based on those models, has to be validated. In this paper, two experimental designs are described to validate the internal of the agent. The first experiment aims at validating the model of human attention allocation (descriptive model) and the second experiment aims at validating the HABTA-component as a whole, incorporating a prescriptive model and support mechanism.

The results from the pilot of the first experiment presented in this paper have proven to be useful, but the actual experiments still have to be performed. Therefore, future research will focus on the performance and analysis of these experiments. It is expected that the accuracy of the model can be increased hereafter, however 100% accurateness will not be attainable. The results of the first experiment will show if the variables indeed provide enough information to improve the accurateness of the model.

With respect to the second experiment, one might argue to add another variant of support, such as one that is configurated by the participant itself. The participant will then do the same as HABTA does, which might result in him being a fair competitor for HABTA. In this way the effectiveness of HABTA-based support can be studied more convincingly, comparing human-agent performance when either the participant or the agent is managing attention allocation. Deciding on this will be subject in the near future.

If the agent does not support the human at the right time and in the right way, this might influence trust and acceptance of the agent. It is interesting to investigate whether an observable and adjustable internal structure of the agent improves trust and acceptance of the system (see for instance [19] in these proceedings). This also needs further research.

In this paper the development and validation of a normative model (prescriptive model) is not described. Validation of this model is important, as it is also a crucial part of the HABTA-component. Errors in this model will lead to support at the wrong time and this will influence performance, trust, and acceptance. Further research is needed in order to develop and validate normative models.

Finally, in general, agent-components have more value when they can be easily adjusted for other applications. It is therefore interesting to see whether HABTA-based support can be applied in other domains as well.

7 Acknowledgments

This research was funded by the Dutch Ministry of Defense under progr. nr. V524. The authors would like to thank Annerieke Heuvelink, Harmen Lafeber, Jan Treur, Kim Kranenborg, Marc Grootjen, Tibor Bosse, Tjerk de Greef, and Willem van Doesburg for their contribution and helpful comments.

References

1. Grootjen, M., Neerincx, M.: Operator load management during task execution in process control. In: Human Factors Impact on Ship Design. (2005)
2. Wickens, C., McCarley, J., Alexander, A., Thomas, L., Ambinder, M., Zheng, S.: Attention-situation awareness (a-sa) model of pilot error. Technical Report AHFD-04-15/NASA-04-5, University of Illinois Human Factors Division (2005)
3. Chalmers, B.A., Webb, R.D.G., Keeble, R.: Modeling shipboard tactical picture compilation. In: Proceedings of the Fifth International Conference on Information Fusion. Volume 2., Sunnyvale, CA, International Society of Information Fusion (2002) 1292–1299
4. Kahneman, D.: Attention and effort. Prentice Hall, Englewoods Cliffs, NJ (1973)
5. Wickens, C.: Processing resources in attention. Varieties of attention (1984)
6. Steinberg, A.: Standardisation in data fusion. In: Proceedings of Eurofusion'99: International Conference on Data Fusion, UK, Stratford-Upon-Avon (1999) 269–277
7. Verkuijlen, R.P.M., Muller, T.J.: Action speed tactical trainer review. Technical report, TNO Human Factors (2007)
8. Treisman, A.: The perception of features and objects. Attention: Awareness, selection, and control (1993)
9. Bosse, T., Doesburg, W.v., Maanen, P.-P. van, Treur, J.: Augmented metacognition addressing dynamic allocation of tasks requiring visual attention. In Schmorrow, D.D., Reeves, L.M., eds.: Proceedings of the Third International Conference on Augmented Cognition (ACI) and 12th International Conference on Human-Computer Interaction (HCI'07). Volume 4565 of Lecture Notes in Computer Science., Springer Verlag (2007)
10. Bosse, T., Maanen, P.-P. van, Treur, J.: A cognitive model for visual attention and its application. In Nishida, T., ed.: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT-06), IEEE Computer Society Press (2006) 255–262
11. Bosse, T., Maanen, P.-P. van, Treur, J.: Temporal differentiation of attentional processes. In Vosniadou, S., Kayser, D., eds.: Proceedings of the Second European Cognitive Science Conference (EuroCogSci'07), IEEE Computer Society Press (2007) 842–847
12. Arciszewski, H., Delft, J.v.: Automated crew support in the command centre of a naval vessel. In: Proceedings of the 10th International Command and Control Research and Technology Symposium. (2005)

13. Heuvelink, A., Both, F.: Boa: A cognitive tactical picture compilation agent. In: Proceedings of the 2007 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2007), IEEE Computer Society Press (2007) forthcoming
14. http://www.few.vu.nl/∼pp/public/vanmaanendekoningvandongen-AmI07.pdf
15. Zijlstra, F.R.H., Doorn, L.v.: The Construction of a Scale to Measure Perceived Effort. Department of Philosophy and Social Sciences, Delft University of Technology, Delft, The Netherlands (1985)
16. Madson, M., Gregor, S.: Measuring human-computer trust. In: Proceedings of the Australasian Conference on Information Systems. (2000)
17. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: Toward a unified view. MIS Quarterly **27**(3) (2003) 425–478
18. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly **13**(3) (1989) 319–340
19. Mioch, T., Harbers, M., van Doesburg, W.A., van den Bosch, K.: Enhancing human understanding through intelligent explanations. In: Proceedings of the First International Workshop on Human Aspects in Ambient Intelligence, Springer Verlag (2007) forthcoming