ATGENTIVE

IST-4-027529-STP

DELIVERABLE D4.2

Status: PUBLIC
Version: 1.0 (Final)

Formative Evaluation

OXFORD BROOKES UNIVERSITY

Oxford, UK, 29th September 2006
Paul Rudman and Mary Zajicek

The following people have contributed to this deliverable:

AUP Damien Clauzel Eugeni Gentchev Claudia Roda Georgi Stojanov Julie Thomas	**Cantoche** Laurent Ach Benoit Morel	**CELN** Jaroslav Cech Ivana Malá Barbora Parrakova	**INSEAD** May Liem *(Administrative Coordinator)* Pradeep Kumar Mittal Thierry Nabeth *(Technical Coordinator)* Ye Deng
OBU **Paul Rudman** *(Contributing Editor)* **Mary Zajicek** *(Contributing Editor)*	**Ontdeknet** David Kingma Inge Molenaar Koen Molenaar Maurice Vereecken	**UTA** Kari-Jouko Raiha Harri Siirtola Veikko Surakka Toni Vanhala	**STC** Gunilla Andersson Hasse Karlsson

Executive Summary

This document describes the formative evaluation (evaluation aimed at guiding and informing the design process) carried out within AtGentive to date. Formative evaluation for AtGentive primarily comprises the generation of feedback from experts and potential users, or their representatives, as to the effectiveness and desirability of proposed interventions. The proposal and design of such interventions begins with the concepts and models developed within the AtGentive Conceptual Framework and is continued and elaborated as part of the design process. Formative evaluation is integral to the development of these concepts and models, and involves the evaluation of simple scenarios of use with experts and potential users, or their representatives, in paper-based form or very simply implemented. Feedback informs researchers working on the Conceptual Framework and the Design as to the usefulness of their approaches. The process is iterative, with new or changed ideas and scenarios undergoing further tests.

The formative evaluation has four main functions, to validate and enhance the Conceptual Framework itself (in particular the scenarios of use), to create a profile for embodied agents suitable for AtGentSchool and AtGentNet, and to propose, validate and enhance design elements of AtGentive interventions for each of the specific platforms AtGentSchool and AtGentNet.

The Conceptual Framework has been extensively influenced by formative evaluation in its content, such as the addition of "temporal context", and in its scenarios of use. These scenarios have undergone evaluation by experts and user representatives, including a dedicated experiment to investigate in depth the impact of selected scenarios.

Experiments have been undertaken to investigate the emotional effects of different types of avatars. These have been measured both subjectively, in questionnaire format, and objectively, using direct physiological measurements. In addition, specially-designed questionnaires have been administered to children aged 11-13 years, to identify specific traits appropriate for an avatar addressing this age group.

Formative evaluation specific to AtGentSchool has comprised a mock-up-style prototype, with the experimenter performing the reasoning, which has been trialled with experts and user representatives. This has enabled the development and validation of event and intervention models, and has shed light on the design of the reasoning module.

Formative evaluation specific to AtGentNet has comprised extensive use of the existing software by the partners as user representatives, leading to informal feedback and formal questionnaire-based feedback. This will enable the design for AtGentNet to interpret the Conceptual Framework in a manner appropriate for the situation of use and intended users.

Finally, the process of formative evaluation within AtGentive is ongoing, and will continue throughout the design phase. This ongoing formative evaluation will continue to guide and inform the AtGentive design process, maximising the effectiveness of the overall AtGentive interventions.

Contents

1. Introduction

1.1 General introduction to AtGentive

Attention appears to represent one of the key factors of learning performance. The most effective learners are not necessary the most intelligent or the brightest ones, but those who are able to (1) organise efficiently their time; (2) sustain concentrating on their key activities and that are able to complete them and (3) have the psychological strength to mobilise all their energy for the last miles that will really make a difference.

This situation is aggravated in an online setting, where learners are left on their own, have fewer points of reference to situate themselves, do not receive any direct pressure from a tutor or from their peers, and can more easily procrastinate or engage in learning activities that are very ineffective.

The objective of this project is to investigate the use of artificial agents for supporting the management of the attention of young or adult learners in the context of individual and collaborative learning environments. Practically, this project comprises the modification of existing learning-support software to incorporate attention-enhancing features identified within AtGentive's conceptual framework (see deliverable D1.3 – "Attention Framework Report") and found to be desirable by the formative evaluation. Such features will range from relatively simple enhancements to facilitate perception of more relevant information, to direct intervention with the user by embodied agents. Overall, the aim is to enhance the learners' effectiveness by directing their attention in more appropriate directions. This will be achieved in three broad ways: implicitly, by direct intervention, and by interventions to proactively coach the learners in the management of their own attention (assessment, guidance, stimulation, etc.).

Interventions will be controlled by agents that will profile the (short or long term) state of the attention of the learners by observing their actions, to assess, to analyse and to reason on these states of attention and to intervene as suggested by the conceptual framework. Where agents need to communicate directly with the learners, these agents will be able to appear as cartoon-style characters, embedded in the application and its interface.

These interventions will be designed and delivered as part of two different learning infrastructures / contexts. One context, AtGentSchool, will be primarily designed for eventual use by selected schools in the Czech Republic. It will support students aged between 7 and 15 years old collaborating with pedagogical experts. AtGentSchool will be built on the Ontdeknet eLearning platform, created by the Dutch company Ontdeknet. This platform is an electronic learning environment that makes knowledge and skills in society accessible to educational institutions in general and individual students in particular. Virtual learning relationships between "experts" and students are established in this virtual learning environment. The Ontdeknet environment provides guidance to support individuals to learn together based upon common interests. This platform already uses an embodied agent ("Onty") to guide learners around the learning environment. The software will be adapted to incorporate the AtGentive interventions, and the existing embodied agent will be changed as necessary.

The second context, AtGentNet, will be primarily designed for eventual use by adult learners enrolled in business-related courses organised by the Swedish Trade Council (STC). It will support adult learners, located individually but collaborating using an internet-based system. AtGentNet will be built on the ICDT virtual community platform, created by the INSEAD Business School's Centre for Advanced Learning Technologies. This platform is a web-based virtual environment aimed at supporting distributed groups and communities. In terms of functionality, the ICDT Platform integrates features aimed at providing efficient Information, Communication, Distribution and Transaction channels used by the community of users. This platform will be adapted to incorporate the AtGentive interventions including, where deemed necessary, an embodied agent.

1.2 Introduction to formative evaluation

The purpose of formative evaluation is to contribute to and inform the design process by providing early and ongoing feedback on the efficacy, suitability and efficiency of the ideas, models, and methods being developed. The result is to guide and inform the design process, maximising the effectiveness and efficiency of the final system.

Practically, the general process of formative evaluation begins with the collation of ideas, models, methods and simple prototypes relating to the system under development. These are then presented to people who will be involved in the final system's use, or who have the necessary experience to act as system users (e.g. usability experts or people with experience of the situation where the system will be used). Their reactions, views and reflections are elicited and collected for the final stage, which is to use these responses as input to the design process, modifying existing ideas, models, methods and prototypes to improve the design. The process is iterative, with modified designs undergoing further evaluation.

This document describes the formative evaluation carried out within AtGentive to date. Formative evaluation for AtGentive primarily comprises the generation of feedback from experts and potential users, or their representatives, as to the effectiveness and desirability of proposed interventions. The proposal and design of such interventions begins with the concepts and models developed within the AtGentive Conceptual Framework (Work Package one—WP1) and is continued and elaborated as part of the Design (WP2). Formative evaluation is integral to the development of these concepts and models, and involves the evaluation of simple scenarios of use with experts and potential users, or their representatives, in paper-based form or very simply implemented. Feedback informs researchers working on the Conceptual Framework and the Design as to the usefulness of their approaches. The process is iterative, with new or changed ideas and scenarios undergoing further tests.

Formative evaluation commenced in January 2006 as the learning models, user characteristics, models of attention, methods of registering attention and models of emotional states began to emerge with early versions of the Conceptual Framework. Formative evaluation is an ongoing, iterative process. It will continue during the design phase as design of the AtGentive modules both generates early prototypes and feeds back to the Conceptual Framework.

The formative evaluation has four main functions by which it contributes to and informs the design:

- to validate and enhance the Conceptual Framework itself, in particular the scenarios as exemplars of the Conceptual Framework in practice.
- to create a profile for embodied agents, suitable for AtGentSchool and AtGentNet.
- to propose, validate and enhance design elements of AtGentive interventions, based upon the Conceptual Framework and its scenarios, related to the specific platform AtGentSchool
- to propose, validate and enhance design elements of AtGentive interventions based upon the Conceptual Framework and its scenarios, related to the specific platform AtGentNet

Finally, formative evaluation contrasts with the later summative and strategic evaluations. The objective of the summative evaluation will be to evaluate the success of the attention-enabled collaborative e-learning platforms, as implemented in the two pilot systems. The strategic evaluation will assess the substantive value of knowledge (technical components or approaches) that has been generated in this project, in order to identify and assess the value of the unique knowledge assets that have been generated, as well as their future potential. Detailed plans for the summative and strategic evaluations will be published in deliverable D4.3 ("Evaluation Plan"), due month 12.

1.3 Approach

The starting point for design of AtGentive interventions is the Conceptual Framework. This framework is grounded in relevant theories from areas such as cognitive psychology and learning, and provides the background to AtGentive interventions and thus their formative evaluations. In approaching these evaluations, then, the first area of interest is the framework itself. Since the framework is well illustrated by scenarios for each of the main themes, an experiment was performed to investigate the efficacy of a selection of these scenarios. This investigation is separate from, but able to influence, other areas of investigation.

One aspect of AtGentive is the use of embodied agents, in the form of a cartoon-style character on the screen, which may be used by AtGentive to communicate with the user. The appearance, style and general behaviour of this character will be important to acceptance of the interventions themselves. Therefore, a second independent strand of evaluation has been to define appropriate characteristics for such an agent. This needed to be done separately for adult and child (around 12 years) users.

The two collaborative learning platforms themselves provide the other two areas for formative evaluations. The approach for the Ontdeknet system has been to match system events against potential interventions suggested by the Conceptual Framework to create an intervention model, and to test this model with experts and potential users using a system mock-up. The approach for the ICDT platform has been to use the platform as the major conduit for communication within AtGentive. This has allowed all partners to evaluate proposed interventions, discussed with reference to the Conceptual Framework.

Thus, the formative evaluation for AtGentive may be broken down into four major sections:

- Conceptual Framework
- Agent embodiment
- AtGentSchool
- AtGentNet

Taken as a whole, the results of the formative evaluation provide a guiding force for the design process of AtGentive.

1.4 Methodology

As stated earlier, the starting point for design of AtGentive is the Conceptual Framework. The relationship between the Conceptual Framework and the main aspects of formative evaluation is illustrated in Figure 1. Support is broadly categorised as that of the user's immediate focus of attention (what they are concentrating on at this moment) and their broader voluntary attentional choices (what they may choose to attend to next). More specifically, the framework breaks attention-supporting interventions into four main forms:

- Perceptual – "Bottom-up" processes (e.g. a flashing image attracts attention)
- Deliberative – "Top-down" processes (e.g. the user may decide to check their email every hour)
- Operational – Managing interruptions (e.g. the user may disconnect a telephone)
- Meta-cognitive – Self-support (e.g. the user may learn which emails are "junk" and can be ignored)

Interventions may be seen as relating to three types of problem:

- Procedural interventions for Regulative problems
- Content interventions for Cognitive problems
- Process interventions for Meta-cognitive problems

A key aspect of the Conceptual Framework is that such interventions may be driven by the monitoring of events. Such events will be discerned from as wide a variety of sources as possible. The main categories of event are as follows:

- Application events (e.g. The user has started a task in the application, new information is available for the user)
- User events (e.g. The user indicates that (s)he wants to be notified about certain events, or that a task should have a high priority)
- Tracking events (e.g. The user has been idle for some time, a resource has been used by other users)

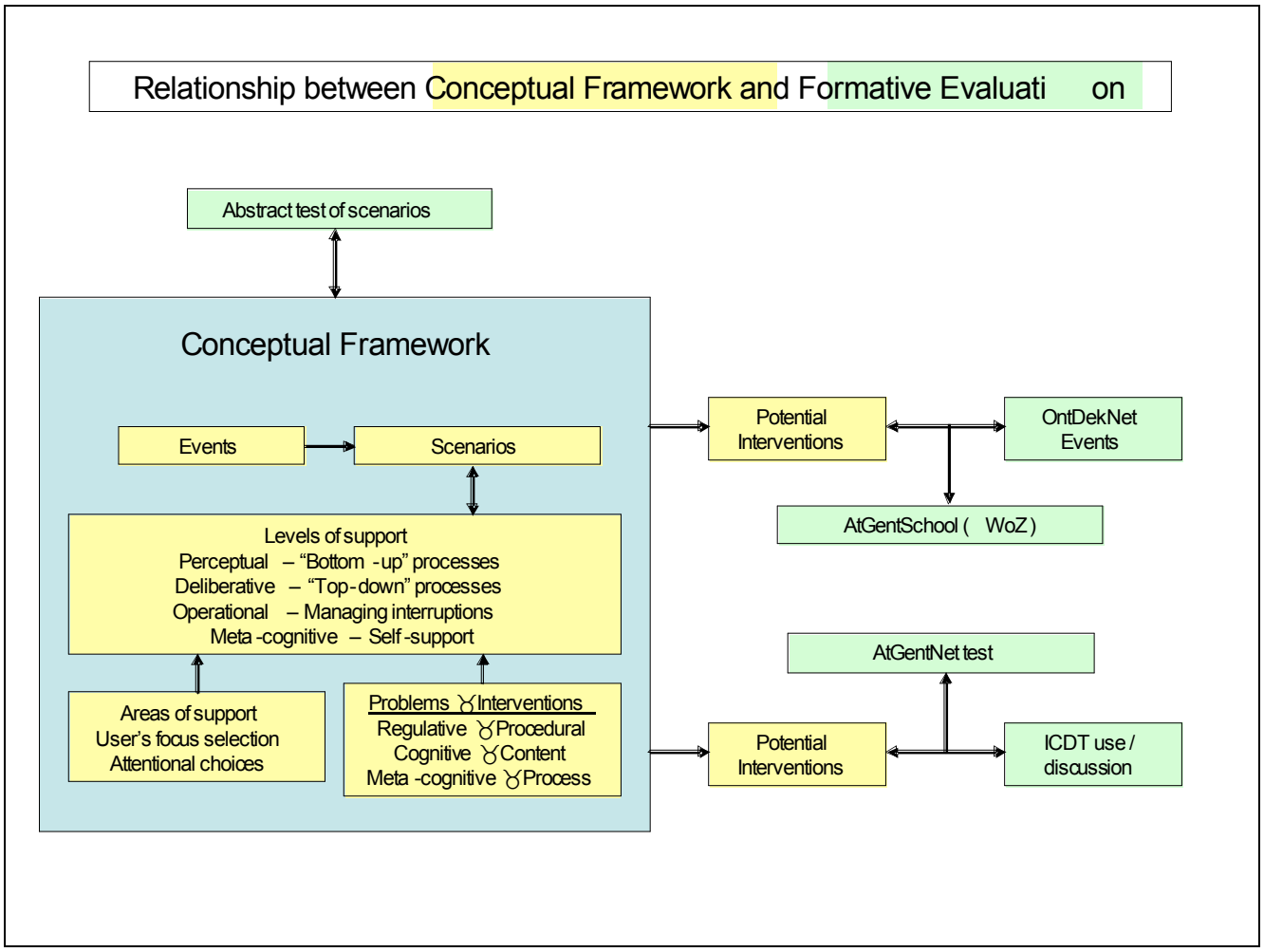


Figure 1 - Relationship between the Conceptual Framework and AtGentive Formative Evaluation

It is hypothesised that these events, which may be captured and analysed by AtGentive agents, will reveal the interaction paths between the essential elements of User(s), Application (Ontdeknet, ICDT), Environment (external events) and the AtGentive agent(s). Further, this observation will reveal the user(s)' attentional choices, preferences, and possible future foci. It is this analysis that results in the agent's interventions. (See deliverable D1.3 ("Attention Framework Report") for further details.)

1.4.1 Formative evaluation of the Conceptual Framework

In approaching these evaluations, then, the first area of interest is the framework itself. In particular, the Conceptual Framework defines a number of user scenarios that typify the anticipated agent interventions based upon the framework. For example, "The user is visiting a knowledge area and the application evaluates that the user should also visit another knowledge area, which he / she has not explored…" (See Appendix A for a full list of the scenarios from the Conceptual Framework). It is the Conceptual Framework, exemplified by these scenarios, that forms the basis for the design of interventions for the two AtGentive applications, AtGentSchool and AtGentNet.

The creation of these scenarios was greatly influenced by the partners' scenario presentations at an AtGentive workshop (held in Paris, January 23rd and 24th 2006 and detailed in Deliverable D1.1 – "Advanced support for attention in collaborative learning settings"). This event allowed each partner to interpret AtGentive from their own experience, generating a rich source of ideas and directions, leading ultimately to the scenarios present in the Conceptual Framework today. For example, Ontdeknet analysed problems that students had experienced and the (teacher's) intervention solutions to create typical evidence-based scenarios for that environment.

These scenarios allow evaluation of the principles underlying the Conceptual Framework by evaluation of the scenarios generated. This allowed an experimental evaluation to be performed of the efficacy of scenarios in isolation (i.e. outside the context of any specific software package) (see Section 2.3). Low-fidelity tools (pen and paper) were used to create a situation in which participants were placed in the situation defined by each of the scenarios. The experiment examined the participants' reactions to and views of selected interventions, allowing conclusions to be drawn about their suitability. This investigation was separate from, but able to influence, other areas of investigation.

1.4.2 Formative evaluation for Embodied Agents

One way in which AtGentive may communicate with the user is by use of an embodied agent, in the form of a cartoon-style character on the screen. The appearance, style and general behaviour of this character will be important to acceptance of the interventions themselves. Therefore, a second independent strand of evaluation has been to define appropriate characteristics for such an agent. This needed to be done separately for adult and child (around 12 years) users.

1.4.3 Formative evaluation of intervention designs for AtGentSchool

The two collaborative learning platforms themselves provide the other two areas of formative evaluations. AtGentSchool is to be based upon the Ontdeknet guided collaborative learning system. The methodology enlisted for AtGentSchool was to match Ontdeknet system events against potential interventions suggested by the Conceptual Framework. This created an intervention model of events / interventions. Testing of this model with experts and representative users was achieved using a system mock-up.

1.4.4 Formative evaluation of intervention designs for AtGentNet

AtGentNet is to be based upon the ICDT platform, a collaborative learning and group discussion and coordination platform. Its strengths and versatility allowed the AtGentive partners to use this platform as the primary conduit for communication within the project. This provided a ready group of users with which to conduct formative evaluations on the software. The methodology employed was for the partners to analyse their own platform use, with respect to the Conceptual Framework, and to propose and validate attention-related interventions, again using the platform itself as a conduit.

2. Formative Evaluation – Conceptual Framework

2.1 Evolution of the Conceptual Framework

The previously stated (see Section 1.4.1), creation of scenarios for the Conceptual Framework was greatly influenced by the partners' scenario presentations at an AtGentive workshop held early in the project. Throughout the project, continuing discussions between the partners, primarily using the ICDT platform, has formed an essential element in clarifying and extending the Conceptual Framework document in an ongoing, iterative process.

As part of this clarification of the Conceptual Framework, a visualisation of the Framework was created. This PowerPoint-format presentation (see Appendix C for an example) elaborates on the Framework document, facilitating the discussion of the document between the AtGentive partners, as well as the presentation of the Framework externally.

Below are detailed specific ideas which were added to the Conceptual Framework as a result of the partner's discussions. Note that many more subtle changes which were influenced by the discussions are not listed here, nor are discussions that did not directly change the Conceptual Framework.

2.1.1 Work groups

A form of user profiling was discussed, and later added to the Conceptual Framework. It involves the automatic inference of Workgroups by noting previous collaborations. This could, for example, allow interruptions to be better prioritised according to how closely the interrupter and interruptee are working together. The user's colleague-network is often difficult to define. A traditional way of defining these workgroups is to maintain formal lists, i.e. workgroup name / users that belong to the group. An alternative would be to infer the existence of groups based on users' activities. For example, it is logical to assume that members of a workgroup will be both working on / with the same documents and communicating with each other. Applications may track these activities. Workgroups could be identified from a combination of trackable events, such as:

- Co-authorship of documents (i.e. the user who creates a document plus all users who subsequently edit that document
- Respondents (i.e. all users who contribute to a threaded conversation (add "sub-documents"))
- Mutual interest (i.e. all users who read the same document)
- Meetings (i.e. attendees of on-line meetings)
- Text chats (i.e. users that 'chat together' (at the same time))

2.1.2 Temporal Context

The ICDT platform is designed specifically to display one document at any one time—the focus is on creating a community and collaborating / competing within that community (Angehrn, 2004). "Documents" are therefore intended to be discussed, and are thus smaller and more focussed than one may otherwise create.

The Conceptual Framework initially discussed the storing and restoring of context, in the form of concurrently open documents. However, since the ICDT Platform does not permit concurrent documents, this form of denoting context is impractical as only one document, or "knowledge asset", may be seen at any one time. To circumvent this, the idea of *temporal* context was created. Temporal context refers to information in use in the same place but at a nearby *time*. Thus, temporal context may be created by reading two documents in succession – each provides context for the other. This approach was added to later versions of the Conceptual Framework.

2.1.3 Recommendations

An approach based on Recommender systems (e.g. Chalmers, 2000) was discussed and adopted. This follows on from the concept of Temporal Context by considering the sequence of events a user chooses and comparing those events against a large database of recorded events from all users. By finding other users who had previously chosen the same sequence of events as a current user, the found sequence may be used to extrapolate suggestions for the current user.

For example, User A is reading an ICDT posting on eye-tracking, having just read a posting on mouse use. The system would look at every access by every user of the mouse use posting followed by the eye-tracking posting, and see which posting was most frequently selected as the next posting to read. Suppose, then, that several users have, in the past, selected the sequence of postings: mouse use / eye-tracking / cognitive load. In this case, the system would recommend to User A the posting on cognitive load as something that may be helpful.

2.1.4 Implicit user feedback

A method still under discussion is to allow the user to give feedback as to his / her preferences by allowing questions to be answered as "Yes / Not now / Never" rather than "Yes / No". This would allow preferences to be gathered without specifically interrupting the user for that purpose.

2.2 Relationship with Design

The partners addressed the relationship between the Conceptual Framework and the design. Specifically, to what extent the design should be constrained by the Conceptual Framework and to what extent the Framework should be expanded to incorporate any new areas of attentive intervention identified during the design phase. Overall, it was concluded that the Conceptual Framework should act as a guide, but the design should not be constrained by the Framework. This requires an increased flexibility for the evaluation methodology, since the design may change at a later stage than the Conceptual Framework, possibly requiring later formative evaluation.

2.3 Experimental evaluation of Conceptual Framework Scenarios

An experimental evaluation of selected scenarios was created (Rudman & Zajicek, 2006 - to appear)—see Appendix D. The investigation at OBU placed participants in situations where interventions occurred as described in each of scenario one (Support to task resumption, restoring task context) and scenario two (Support to limited time resources

allocation) (see Appendix A for summary of scenarios, and Clauzel, Roda, & Stojanov, 2006 and deliverable D1.3—"Attention Framework Report"—for details). Afterwards, participants filled in a questionnaire and took part in a short interview, to elicit their feelings and opinions about the interventions. The purpose of the investigation was to look for potential user-related problems with these specific interventions, so that such problems can be circumvented or minimised as far as possible in any future agent implementation.

The investigation was conducted using paper-based low-fidelity prototyping tools. The intention was to ensure that any problems found were not created by the software interface used by the investigation, rather than the task situation itself. The chosen domain comprised general information on herbs for medicinal use, as it comprises a large amount of well documented and inter-related information. Participants were enlisted from within Oxford Brookes University, comprising administrative staff and Computer Science PhD students. Each was given a set of cards, showing details of one herb per card. They were then given questions to answer which required the use of multiple cards. For example, "You are going on a long journey. Briefly describe the flowers of a plant that could help."

After an initial familiarisation question, two main questions are given, one each for the evaluation of scenarios one and two. During the first question (relating to scenario one—Support to task resumption, restoring task context) participants were interrupted by the experimenter and a note taken of the cards in use. Later (after the second question) they are asked to continue, with or without being told which cards were being used previously. During the second question (relating to scenario two—Support to limited time resources allocation) participants were given a choice of two questions. After a short time, the experimenter informed them that "the other question" (regardless of which they have chosen to answer) was quicker to complete.

At the end of the experiment, participants were asked to fill in a purpose-designed questionnaire to ascertain their views on the experience. In particular, they were asked about their feelings when interrupted with "useful" information (scenario two) and about the usefulness of the information offered when restoring the context (scenario one). Finally, a short interview took place to allow the participants' views to be given in their own words.

2.3.1 Results

Intervention A - Restore task context

The intention of this intervention was to help the participant restore the task's context by reminding him or her of contextual information previously-used in this context—in this case herb names.

Participants generally appreciated being given assistance in restarting this question, finding it helpful and being pleased that someone was trying to help. However, four participants found this contextual information unhelpful or annoying, with one person only using the contextual information because (s)he felt (s)he ought to.

It seems that there was a separate group with an identifiable reason for not wanting the "assistance". These participants, on seeing the question a second time (or possibly in the

intervening time) decided to take a different approach to their answer. Thus, when the contextual information on the coloured paper was made available to them it was no longer relevant. How they dealt with this varied, from disinterest to annoyance. It is possible that the offer of contextual information was taken as a suggestion that their new direction was somehow being called into question.

By offering now-outdated contextual information, the agent may be seen as criticising the person's decision to change their approach to answering the question. It seems that any agent needs to consider that information may be out of date, since offering it in these circumstances appears likely to generate a negative emotional response.

Intervention B - Propose alternative task

The intention of this intervention was to help the participant to be more productive overall by suggesting they change from a recently-begun task to one that will be more appropriate given the time available.

Telling participants that the alternative question was (supposedly) quicker than the one they were attempting was generally seen as annoying and/or unhelpful. The reason given was consistent: once they had decided upon a task they were committed to that course of action. They did not want to change and restart in a different direction. Even when they believed that the advice was correct, and it would indeed be quicker at that point to change questions as suggested, they did not necessarily take the advice and go back on their original decision.

The investigation was deliberately set up to offer the suggestion very shortly after this decision. Clearly, stating on the question sheet "Please note that question n is quicker to answer" would be different. The investigation therefore attempted to offer the suggestion at the point where a person's actions reveal their decision.

In one case this strategy failed, with the suggestion being given inadvertently before the participant had decided which question to answer. As expected, the person involved was not unhappy with the suggestion (but stated that (s)he would have felt confused if the suggestion had come after making a decision).

As one would expect, all participants who changed their question considered the information (that the "other" question would be quicker) to be accurate (i.e. they believed the researcher). What is interesting is that most (six out of eight) people who did not change their question said that they did not believe that the other question would be quicker. There was no evidence in the materials provided that either question would be quicker to complete and no obvious reason for this being the case. Indeed, participants were told the "other" question would be quicker whichever question they attempted. (The need for this deception was explained to them after the trial.) Any decision to ignore the suggestion then was a subjective decision. It seems, in fact, to be a matter of belief—or otherwise—in the information given. Thus, it seems that believability and related issues (trust and likeability for example) are important attributes for any agent in this situation.

In addition, giving details of why the suggestion has been made may help make the intervention less annoying, as it would assist the person to justify any change of decision, something that seemed important in this study.

As regards the timing of this intervention, there seems to be a critical point at which the effect of the suggestion changes; it is the point at which the participant makes a decision as to which question to answer. This is problematic to any agent that intends to offer advice based on the user's decision to start a task. If, for example, an agent waits until a document has opened and the first line has been read (which is likely as the agent software may take a moment to react) the user will feel that they have "begun" the task, and may not appreciate being told, for example, that reading an important email now would be better than starting this long document. Strategies that put the user more in control, such as offering to assist but without saying immediately what the assistance is, may be worth employing in this type of situation.

2.3.2 Conclusion

In both interventions trialled in this investigation, the source of negative feelings was similar. Where a participant's viewpoint was in some way called into question, without there appearing to be sufficient reason, the result was negative feelings, such as annoyance and frustration.

With intervention A (restore task context) the intervention was acceptable provided the person had not changed the manner in which they intended to approach the task. After this point, offering assistance that implied the original strategy was "correct" may generate a negative emotional response, possibly based on the person not wanting to have the new approach undermined, rather than the contextual information simply being unhelpful.

Results from intervention B (propose alternative task) show that suggesting to a person that they switch tasks, having just begun a task of their choosing, is difficult to achieve without a negative emotional response. An alternative strategy would be to assist the person in justifying (to themselves) the suggested change by offering supporting evidence. Maximising the believability of the agent making the suggestion would also seem to reduce the negative aspect of this form of suggestion (related factors, such as trust and likeability may also be relevant).

Overall, this investigation suggests that in human-agent interactions the agent needs to take account of the human's likely feelings towards any intervention. Simply giving information that "should" be helpful is not sufficient. Results suggest that a software agent in these situations needs to take into account its position in the "social hierarchy": the agent will not usually be in a position to tell the human what to do, and must avoid actions that appear to do this (even when that is its intention). This general conclusion may relate to other scenarios and needs to be borne in mind with the application of AtGentive scenarios generally.

3. Formative Evaluation – Agent embodiment

Formative evaluations of the avatars were run in parallel and separately from formative evaluations of the interventions. This was to ensure that emotional effects are fully investigated as a separate factor in their own right. It was agreed that any avatar-based pilot interventions carried out before completion of avatar evaluation should use a 'neutral' type avatar, such as a stick man, or plain text without an avatar. For formative

evaluation purposes, Cantoche provided a number of avatars. These were representative of the main avatar types available.

3.1 Formative evaluation of the emotional effect of Avatars

In spite of the potential benefits of embodied agents, there is little empirical evidence to help in designing characters and social cues that are effective in guiding attention, whether directly or through other processes (e.g. emotional responses). The purpose of UTA's two-part study was to investigate the effects of different agent characteristics on the subjective ratings of emotion, attention, usability, and the perceived role of the agent. Thus, the first, more extensive, questionnaire study focuses on the effects of agents' appearances whereas the second study was a laboratory experiment investigating the specific effects of size and emotional facial expressions.

3.1.1 Part One - Questionnaire study

A questionnaire compared users' views of, and responses to, five of the avatars provided by Cantoche (see Figure 2). These avatars were chosen to represent available categories of avatar:

A="Dino" – Imaginary character, no gender
B="William" – Lifelike human, male
C="Julie" - Lifelike human, female
D="Philippe" – Cartoon human, male (cartoon female not represented)
E="Onty" – Existing avatar for the Ontdeknet system

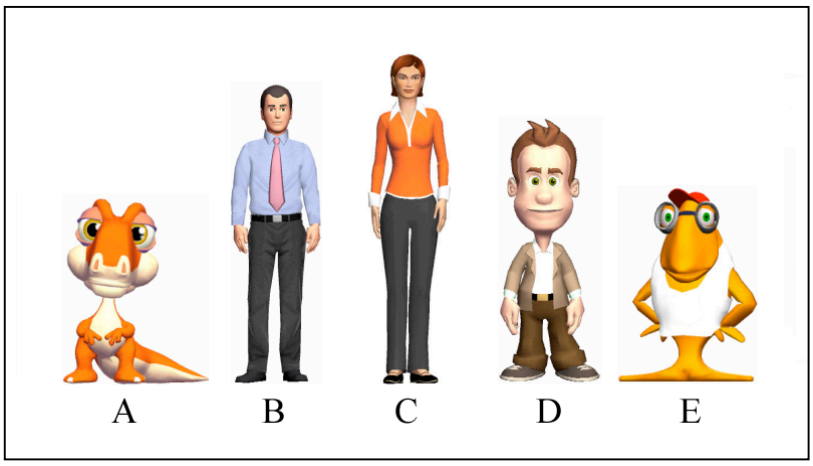


Figure 2 - Five avatars used for comparison

The questionnaire was designed specifically for the AtGentive project and covers four different areas, each using three scales: emotion (valence, arousal and dominance), usability (effectiveness, efficiency and satisfaction), attention and the social role of the agent. In addition, three facial expressions were evaluated for each character (neutral, happy and sad)—examples are shown in Figure 3.

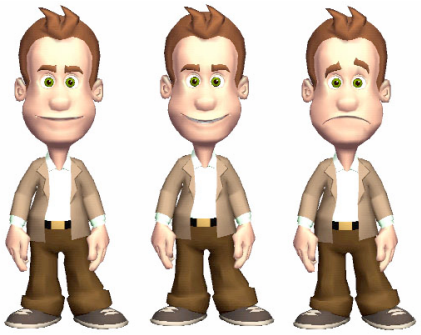


Figure 3 - Examples of the three facial expressions used (Neutral / Happy / Sad)

Each question was presented on a separate web page, which showed a static image of the agent character against a white background, the question, and the scale from which the answer was chosen. After the subject had chosen his or her answer, the questionnaire automatically moved to the next question. The static images of the characters were resized so that they all had approximately the same area.

A total of 75 volunteers participated in the study. 63 % of the participants were male and 37 % were female. Participants' age ranged from 18 to 56 years old with a median age of 26

Results

The responses were categorised according to the following two-dimensional indicators:

- Affect Space: Valence vs. Arousal
- Affect Space: Valence vs. Dominance
- Usability Space: Efficiency vs. Satisfaction
- Conspicuousness vs. Interestingness
- Role Space: Pertinence vs. Believability

All characters were chosen at least once both in the most favourite and the least favourite categories. Dino and Julie won the most favourite character category with equal number of votes. Julie was the most popular in the happy expression condition while Dino was preferred in the conditions of angry and neutral expression. Onty was the least favourite character in all the conditions for both female and male participants. For detailed results see Appendix E.

Discussion

Results showed that both the character itself and its expression had a significant effect on most of the ratings. For example, the non-human characters Dino and Onty were rated as the most conspicuous and interesting, suggesting that they are efficient in influencing attention. However, they were also rated the least pertinent and believable, which suggests that these characters are less suited for tutoring. The most pertinent and believable characters were Julie and William which are realistic human avatars. On the other hand, they were quite uninteresting and inconspicuous as compared to Dino and Onty. Thus, our results suggest that non-human characters are better suited for

attracting attention, while human characters have qualities that are potentially effective when instructing people.

However, it must be noted that the evaluated characters were clearly different from each other with respect to body proportions, colours, features, clothing etc. For this reason, we cannot be exactly sure which features of the agents caused the reported differences in ratings. Further analysis is planned of the open-ended questions concerning the reasons for liking and disliking certain agents.

The facial expressions of agents influenced significantly participants' ratings of emotional valence, efficiency, and satisfaction. However, when looking at the mean ratings for different expressions, it is evident that the differences are much more modest than between different characters. Most mean ratings are close to neutral, varying between four and six. This may be due to the method used to present the emotional cues, since the facial expressions are quite minor and subtle (see Figure 1). Further, as emotional cues of the character were varied between subjects, the effect of emotional cues was produced without reference to the other emotional expressions. Thus, considering these issues, the significant effects of facial expressions are particularly impressive.

Interestingly, angry expressions were rated as significantly more unpleasant, more arousing, less efficient, and less satisfying than other expressions, but there were only a few differences between the ratings for happy and neutral expressions. This might suggest that negative (angry) expressions of embodied agents are potentially more effective for influencing emotional and cognitive processes than positive and neutral expressions. On the other hand, happy expressions were rated as the most interesting. Thus, these may be more useful for maintaining attention and acting as psychological conditioners, for example, when rewarding students.

Further, there were also several interaction effects between facial expressions and gender. For example, male subjects were least aroused and most satisfied by a happy expression and female subjects by a neutral expression, while a character of the opposite gender was seen as the most satisfying and least distracting. This suggests that it might be useful (or maybe even necessary) to use different emotional cues (i.e. facial expressions) in human-avatar interaction depending on the person's gender. On the other hand, there was little difference in the preference of characters between male and female participants. For example, most men and most women chose either Dino or Julie as their favourite character.

Onty was the least favourite character for both male and female participants and each expression. This result is in accordance with the evaluation scales, as Dino and Julie were seen as the most pleasant and Onty generally received the poorest evaluation with respect to almost all of the categories. However, Onty was designed for, and has been used in, a learning environment for children, while the participants in the present study were adults. It is therefore possible that children may prefer Onty.

Further, agents were presented as static images without any dynamic movement or interaction. Dynamic behaviours (e.g., gestures, body movements) are arguably a large part of a character's personality and can greatly affect how a character is experienced. However, in the present work we aimed to study the effects of the character's appearance. Investigating all possible character behaviours, or even a subset of them,

would increase the number of variations significantly and complicate the interpretation of the results. In other words, if characters not only looked but also behaved differently, it would be impossible to tell which factors contributed to the respondents' ratings and to what extent.

3.1.2 Part Two - Laboratory experiment

A laboratory experiment was performed to investigate the specific effects of size and facial expressions using physiological measures of cognitive and emotional processes. Based on the results of the first study, two agent characters were chosen for this laboratory experiment. Participants viewed agents varying in their size and facial emotional expression. Size and facial expression were chosen as variables for the study based on previous results suggesting that they are effective social cues, not only in human-human interaction, but in human-agent interaction as well (Partala, Surakka, & Lahti, 2004). Also, these two characteristics can be easily and dynamically varied independently of other characteristics (e.g., appearance) of the agent. During the experiment, participants reported their experiences using the emotional and attentional scales of the developed questionnaire. Further, in order to provide converging and validating data, we collected several physiological measures of cognitive and emotional processes. These measures included both heart and facial muscle activity measures.

Another aim of the laboratory study was to investigate and validate potential physiological measures of attentional processes. Psychophysiological measures have the potential to act as continuous estimates of the cognitive and emotional state of a person, and are used extensively in emotion research (e.g. Larsen, Norris, & Cacioppo, 2003)). It is difficult to acquire as accurate information on a competitively fine time scale and in real time with other measures (Öhman, Hamm, & Hugdahl, 2000). For example, the exact time of a reaction to a certain surprising event is more easily identified as a change in physiological parameters than using, for example, a post-study questionnaire. Thus, if attention is to be continuously monitored, psychophysiological measures potentially offer a viable alternative that is less invasive and does not require intervention during acquisition. Identifying suitable, robust measures can help in later stages of the AtGentive project where attention is to be continuously monitored in less strictly controlled setups and environments, for example, while freely exploring a Web-based learning environment with embodied agents in a classroom.

Results

Smiling agents were experienced as most pleasant and frowning agents as most unpleasant. There were significant differences between small-sized Julie portraying anger and happiness, large-sized Julie portraying anger and with a neutral expression, small-sized William portraying anger and happiness, medium-sized William portraying anger and happiness, medium-sized William portraying anger and with a neutral expression, and large-sized William portraying anger and with a neutral expression. Other comparisons were not significant.

There was a significant interaction effect of the three factors agent, size, and expression in dominance ratings. Conspicuousness ratings showed a significant main effect of agent's size; small agents were rated as least conspicuous, while large agents were rated as most conspicuous. Interestingness ratings showed a significant interaction between

agent's size and facial expression; small- and medium-sized agents with neutral expressions were rated less interesting than angry and happy ones. However, this difference diminishes when the agent is large. Results suggest that anger was experienced as the most distracting facial expression, while neutral and happy expressions had little effect on concentration.

As regards physiological responses, heart rate decelerated during all stimuli compared to baseline, but deceleration was smaller when viewing larger stimuli. There was somewhat less corrugator supercilii (facial muscle) activity when viewing positive expressions compared to neutral and negative expressions.

Discussion

Preliminary data suggest that both size and facial expressions of a virtual character have effects on attentional processes. According to subjective ratings from participants, larger agents are more conspicuous and, in some cases, more interesting. Similarly, mean heart rate deceleration was largest when viewing small stimuli and smallest when viewing large stimuli. Previously, heart rate has been linked with both emotional and cognitive processes, including attention (Bradley, 2000) (Öhman et al., 2000).

Results also suggest that emotional expressions of virtual agents influenced attentional processes. Preliminary analysis of subjective ratings suggests that, generally, agents with emotional expressions were more interesting than neutral agents. However, there is a significant interaction between the effects of size and emotional expression. This is understandable, as the accuracy in detecting and recognizing facial expressions clearly depends on the size of the stimulus. Nonetheless, anger especially seems to evoke responses more easily. Firstly, in the present study only anger evoked significantly different emotional valence experiences when compared to other expressions. Secondly, these responses were significantly different between responses to facial expressions portrayed by the smallest agents. Thirdly, previous studies suggest that, in general, emotionally negative stimuli have a larger capacity for evoking and maintaining attention (e.g. Hansen & Hansen, 1988, Carretié, 2003 #154). It is interesting and useful for the present project to note that this effect holds for the emotional expressions of artificial agents as well. Emotionally negative stimulation should, however, be used in moderation.

Finally, these investigations do not show how specific emoting—smiling, frowning, etc.—may be best employed in relation to given situations. For example, stating that an email has arrived from the user's manager does not immediately suggest that the avatar should smile, frown, etc. These questions will be considered during the more detailed design phase.

3.1.3 Formative evaluation of the emotional effect of Avatars - General discussion

The results from both parts of the study together show that all three features, that is appearance, size, and facial expression, were significantly associated with the three-dimensional spaces of emotion, cognition, and usability. Overall, results suggest that both positive and negative emotional expressions can be used to influence attention. According to the results of the first study, negative stimulation seems especially efficient

at influencing emotional and cognitive processes, but positive stimulation also attracts interest. Further, the preliminary results of the second study suggested that, generally speaking, agents with emotional expressions can be more interesting than neutral agents. However, in both studies there were several interaction effects involving emotional expression, gender, and size. Thus, the results also suggest that we should consider using different emotional cues for male and female computer users as well as depending on an agent's other properties.

The ratings of the first study also suggest that when pertinence and believability are important, realistic characters should be used. On the other hand, non-human characters were more efficient in influencing attention (i.e. were rated more conspicuous and interesting). Further, the non-human character called Dino was generally seen as very pleasant and satisfying by both male and female participants. Thus, different types of characters are suited to different roles.

Subjective ratings from the second study showed that larger agents were experienced as more conspicuous. As size can be dynamically varied during the interaction regardless of the agent character, it is potentially a very useful and practical attentional and social cue. Further, preliminary data suggest that different agent sizes also induce different types of physiological responses. However, the results of the second study are still rather speculative, as data has been collected from very few participants. Nonetheless, the results seem promising and suggest that robust physiological measures for attentional and emotional processes can be found, once additional data have been collected.

In summary, it was shown that all three features of agents, that is, size, facial expression, and appearance, affected subjective experiences and physiological responses related to cognitive and emotional processes. Further, a number of characteristics influencing and moderating these effects were identified, for instance gender. Results so far are mostly based on the first study, as data for the second study are still being collected. Nonetheless, results already provide a useful insight for designing human-avatar interaction, for example, considering the advantages and disadvantages of using human and non-human characters for influencing attention.

3.2 Formative Evaluation of a character for AtGentNet

The formative evaluation of the emotional affect of Avatars on adults (described above), have allowed Cantoche to begin designing a character for AtGentNet That character will be serious is style, i.e. not too "cartoonish". For example, the style may be based around the image projected by a good business school or a strategic consulting company—serious, but not too much so, and with a little bit of humour and detachment. The character needs to be multi-cultural.

Overall, the AtGentNet character is most likely to be Realistic / Adult / Female (such as "Julie", as used for the formative evaluation). The actual design process will follow Cantoche's standard, proprietary, methodology. Formative evaluation of the character will continue once design and creation of the character has been completed.

3.3 Formative Evaluation of a character for AtGentSchool

AtGentSchool is to be based upon the Ontdeknet collaborative guided learning environment, designed for 11 to 13 years old. Designing an avatar for this age group requires a greater input from the users than designing for adults, since many assumptions that adult designers may make about adults generally cannot be made with younger people. A questionnaire study was carried out to guide the design of an avatar for AtGentSchool. 64 students were recruited by Ontdeknet, all within this age range and experienced with using the Ontdeknet system. Currently, this system includes the avatar "Onty" (see Figure 4). Therefore, evaluation began with a brief discussion of this character.

Figure 4 - "Onty" - Ontdeknet's existing avatar

The discussion found that students considered Onty rather stupid. This judgement seems to be more affected by his functioning then by his appearance. The foremost reason was that you cannot turn him off. He was considered really helpful when the students first started to work with Ontdeknet, but once they knew how Ontdeknet functioned then Onty should not intervene so frequently.

3.3.1 The questionnaire

The questionnaire was designed to get a perception of the world view of these children; the world view of 11 to 13 years old is difficult to determine as this age group is in-between being a child and a teenager. Thus, more general questions were asked in order to ensure that the appropriate direction was taken in designing an avatar for this age range (rather than to ask about specific avatars having assumed a direction). The questionnaire consisted mostly of open questions. This decision was made to prevent students from just crossing a box—selecting from answers the experimenters deemed important. Questions were created in four areas:

- the world of 11 to 13 year olds
- the type of character
- the appearance and colours of the character
- the voice of the character

3.3.2 Results

Open questions were given for each of the four areas. For example, "What cartoon figure do you relate to?". The answers were collated and those answers freely given the

most were taken as most representative of the participants' views. Below are a list of the questions and the two most popular answers for each question. The number of participants that independently chose each answer is given in square parentheses after each answer. For detailed results see Appendix F.

It is important to note that of the 64 respondents, 47 were boys and 17 girls. In order to control for the imbalance between boys and girls, the difference between the answers of the boys and girls was examined. In general, no big differences were found; where there is an important difference, it is mentioned specifically.

The world of 11 to 13 year olds

"What cartoon figure do you relate to?"

- Donald Duck [10] (see Figure 5)
- Sponge Bob [9] (see Figure 5)

"What person do you relate to?"

- Cristiano Ronaldo [7] (football player - see Figure 6)
- Lil jon romeo [3] (singer - see Figure 6)
- Ronaldinho [3] (football player - see Figure 6)
- Riguelme [3] (football player)

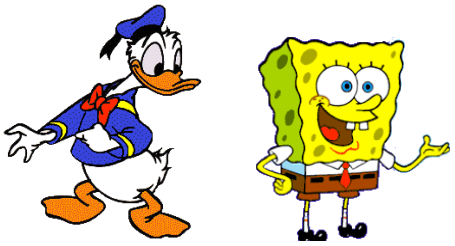


Figure 5 – Most-chosen cartoon characters – "Donald Duck" and "Sponge Bob"

Figure 6 - Most-chosen people – Cristiano Ronaldo, Lil jon romeo and Ronaldinho

The type of character

What type of character would you prefer?

- A human [30]
- A cartoon [29]

What gender should the character have?

- Male [25]
- Female [22]

(See discussion for interpretation of these answers)

Do you have ideas for the character?

- Talk less [12]
- Funny [8]

What features would you appreciate in the character?

- Funny [9]
- Erasable [5]

What features would you NOT appreciate in the character?

- When it talks a lot [21]
- Onty [10]

The appearance and colours of the character

What colours would you like to see in the character?

- Blue [9]
- White [9]

If the character wears cloth, what kind of clothing would it wear?

- Hip hop cloth [6]
- Normal cloth [5]
- Football cloth [5]

The voice of the character

What kind of voice should the character have?

- Heavy [10]
- Sweet and soft [9]

How would the character talk?

- As a human [18]
- Tough [5]

3.3.3 Discussion

The world of 11 to 13 year olds

The cartoons these students mention are cartoons famous within the participants' culture: currently Sponge Bob and Donald Duck. All figures that are mentioned have a strong human or animal appearance. It is, unlike Onty, clear what they are, e.g. a sponge or a duck. Students relate to famous people: football players, artists, movie stars and singers. In particular, they mention far more football players than any other category. This age group does not seem to relate to fantasy figures.

The type of character

The preference of this age group is for either a human or an animal cartoon figure. It may be of either gender. When we analysed this question according to gender of the respondent we concluded that most girls prefer a male character and most boys are indifferent about the gender.

The character type of the character could be human, a handsome man, a football player or a sporty type. Priority number one is that the agent should talk less (than Onty), be funny, cheerful, friendly and nice, beautiful or handsome, brave and last but not least realistic, both in appearance and behaviour. It almost sounds as if these students would look for their perfect friend in the agent. They would like a character that they can look up to and that takes them as users seriously. It should not be stupid, or ugly or irritating and most importantly it should only be there when they need it.

The appearance and colours of the character

The colours need to be realistic and comply with the design of the character. The clothing should be normal and fitting with the character according to the students. Hip hop clothing is named often and football cloth, sexy and tough.

The voice of the character

The clearest conclusion is that the character should sound like a normal human. It should talk with normal human language. Students would prefer a voice that you instantly like and that is clear and easy to understand.

3.3.4 Conclusion

Based on the results of this questionnaire study, it was concluded that students in the age group 11-13 are relating to well-known cartoons and famous people in football, music and the movies.

They have a strong preference for a normal, real, funny, friendly, beautiful and brave character. It could be either human of animal. The gender can be male or female. The personality type should be human, sportive or a cartoon. The voice, colours and clothing should be in keeping with the character.

It almost seems as if the students are looking for their role model or perfect friend in the character. The character should respect the user and keep his or her wishes in mind. It should not talk too much and should not be on the screen all the time. This age group

has a very strong desire to be taken seriously and the appearance of the character should support that desire.

4. Formative Evaluation – AtGentSchool

4.1 Introduction

The two pilot systems AtGentNet and AtGentSchool have certain properties that lend themselves to differing formative evaluation methods. AtGentSchool is a guided learning environment, in which the range of user options is limited. This allowed a comprehensive evaluation of the intended interventions. The approach was to perform a "Wizard of Oz" evaluation (Gould, Conti, & Hovanyecz, 1983) of the applicable scenarios.

The Wizard of Oz method presents to the user an apparently working system. They are unaware that a critical part of the system—in this case the decision making process that matches the user's actions with agent interventions—is missing, and being performed by a human (the "Wizard of Oz" or "WOZ"—a confederate of the experimenter who observes the situation and manually initiates each of the agent's actions).

The aim of this formative evaluation was to examine the potential for applying the scenarios developed in the conceptual model to the existing processes within the Ontdeknet system. Mary Zajicek worked closely with Ontdeknet on the development of the Wizard of Oz system and it was successfully demonstrated at the First AtGentive Project Meeting in Oxford, 22 / 23 May 2006, and subsequently trialled with experts and user representatives by Ontdeknet in The Netherlands.

Figure 7 shows how the Wizard of Oz formative evaluation fits into the overall Ontdeknet formative evaluations. The process is essentially iterative whereby each of the scenarios developed within the Conceptual Framework is considered for relevance to AtGentSchool. If considered relevant then it is implemented using WOZ and evaluated. This process is used to inform design decisions concerning the proposed AtGentSchool pilot system.

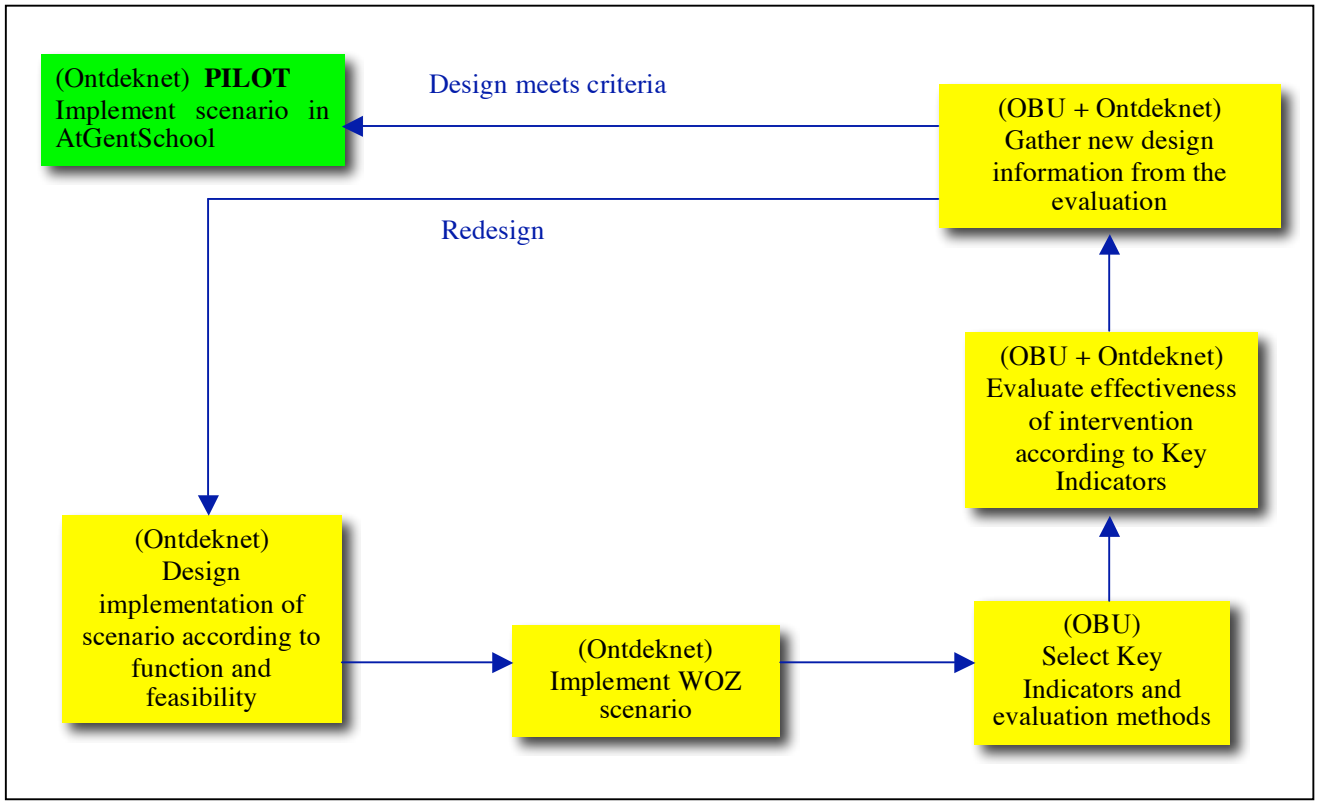


Figure 7 - Ontdeknet Wizard of Oz Plan

4.2 Wizard of Oz evaluation

The "Wizard of Oz" method (Gould et al., 1983) is commonly used when developing systems in order to achieve rapid feedback on design decisions. The attention intervention algorithms to be used in the pilots depend upon the availability of system information and the properties of the user. We can learn a great deal about the implications of attention interventions by using the Wizard of Oz method for both heuristic evaluation and user test evaluation. With this method, all intervention decisions normally taken by the AtGentive module are taken by a human (the experimenter) who sits (logically) between the participant and the available application responses. This also allows the "Intervention decision model" to be created on paper, so that it may be adapted quickly to incorporate the outcomes of previous Wizard of Oz experiments (see Figure 8).

The Wizard of Oz system enables:

- Evaluation of the attentive interventions in terms of users' responses
- Evaluation of the emotional quality of the agent's responses (currently "Onty")
- Evaluation of the efficiency of moving through the interface

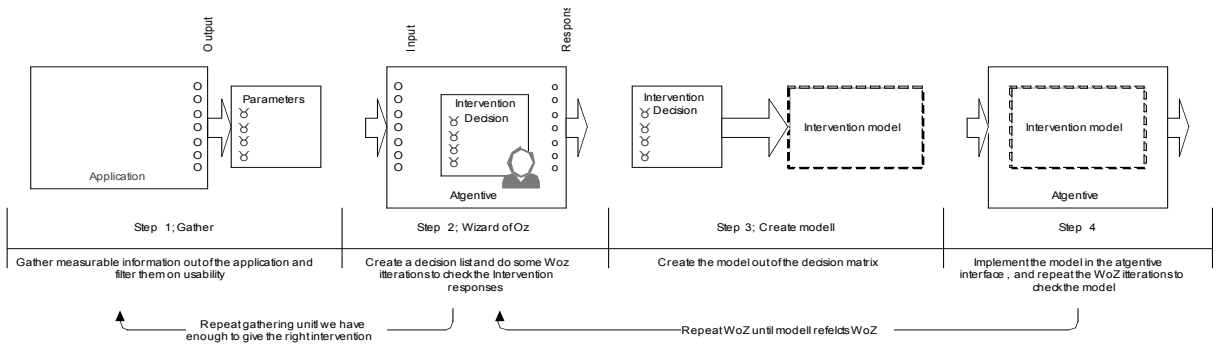


Figure 8 - Ontdeknet Intervention model

4.2.1 The Event Model

The events output from the application are the inputs for the WOZ (i.e. the experimenter / confederate) to decide which intervention to send to the user.

The first step towards the event model was a list, including all parameters, which can be derived from the Ontdeknet platform. The second step was filtering that list on the "not feasible to reason" and "out of scope" parameters to get a short list. This short list was included in the initial WOZ model. The list was modified based upon the results of the WOZ. For detailed results see Appendix H.

4.2.2 The Intervention Model

The Intervention Model in the application is the set of interventions from which the WOZ can select the most appropriate intervention. The Intervention Model was determined based on a research study, the current functioning of the agent "Onty" and with reference to the Conceptual Framework. Problems that students experience working on the Ontdeknet environment were analysed and intervention solutions were created. These cases were used to complete the Intervention Model. (For a full description of the development of the Intervention Model see Appendix I.) The three main intervention categories are Regulation, Cognition and Metacognition. The function of each category is described below:

- Regulative interventions support the user in working effectively with the environment. These interventions are directed at support of the user to move effectively between tasks
- Cognitive interventions support the user's learning process during the execution a task within Ontdeknet. It is directed at activating cognitive learning behaviour in users
- Metacognitive interventions support the user to retain an overview of his or her learning process. This learning process is supported by the

learning sequences of Ontdeknet. The support is directed at activating adequate metacognitive behaviour in users

4.2.3 The Intervention Decisions

During the WOZ process, one person takes the role of the WOZ. That person monitors events as they occur. It is for the WOZ to decide – manually, using their skill and judgement – which of the interventions from the Intervention Model to initiate, and when.

Thus, responses to user actions are manually selected from those in the Intervention Model, based on the WOZ's interpretation of events from the Event Model which the system has identified as taking place.

4.2.4 Stage one - Results

Stage one comprised three tests with adult users familiar with the Ontdeknet system. This was to compensate for the unfamiliarity with their role of the person acting as the WOZ.

Session one

Interventions were mostly on the regulative level. The average rating for fit was 3. The interventions were scored as helpful, average 4. The information value was neutral scored 3. The user was mostly happy scored 5, occasionally confused and irritated. Confusion and irritation were caused by:

- Misinterpretation of events by the WOZ
- Regulative navigational support without an introduction of the activity the agent was navigating the user to

Session two

Interventions were mostly regulative and metacognitive. Sometimes the WOZ gave an intervention to the user (for example navigation support) and at the same time, the user corrected his or her (navigation) error. It might be useful to be able to retract / abort the intervention. High fit was given for rightfully provided introductions. Low fit were also quoted by the WOZ. Interventions that were misfits caused confusion and irritation from the user. Misinterpretation of events by the WOZ occurred for the following instances:

- Regulative navigational support that does not make sense to the user
- Interventions missing in the intervention model

Session three

The person playing the part of the WOZ was not overly familiar with the interventions. This significantly impacted on their ability to select appropriate interventions in the time available, and thus on the validity of this particular Session.

The fit was average with a good correspondence between the WOZ and the users. Interventions were quoted as confusing when the timing is wrong. An extensive navigation process quoted as irritating, but this has little to do with the interventions. Misinterpretation of events by the WOZ occurred for the following instances:

- Regulative navigational support came too late

Fit

The fit in these sessions transformed from reasonably low to reasonably high. There was good correspondence of the determination of the fit between the users and the WOZ. When WOZ is doubtful, the fit is low and also when the WOZ is sure the fit is high. Misfits were due to missing interventions and in cases when the intervention was given simultaneously with an action of the user. In AtGentive, the actions of the users will need to overwrite interventions to prevent these events. Misfits due to misinterpretation of the events occurred quite often in the first session. After adjustments in sessions two and three, this occurred only occasionally.

User reactions

The user's states written down in the intervention log were mostly happy. Users found the interventions informative and helpful. Irritation and confusion occurred in the following situations:

- when the wrong intervention was given to the user
- when navigational interventions are not properly introduced
- when interventions are wrongly timed.

After the session, users were asked to fill in a questionnaire to evaluate the interventions. The user reactions on indicators performance, satisfaction and attention were calculated based on the answers in the questionnaire and are shown in Table 1. (Learning was included in the questionnaire, but the sessions were not long enough to permit significant learning by the users, so this has been excluded from the results shown here). The maximum score for each indicator is stated in the table header. Note that there was an increase on all indicators after the adjustments to the WOZ setting following session one.

User	Performance(20)	Satisfaction(20)	Attention(15)
Session 1	11	10	10
Session 2	16	15	14
Session 3	16	15	14

Table 1 – Stage one - User scores of Performance, Satisfaction and Attention, per session

4.2.5 Results – Stage two

Stage one comprised three sessions with adult users unfamiliar with the Ontdeknet system. This was to increase the feedback that the users were able to give, but still ensure that the session proceeded smoothly.

A number of extra events and parameters were included in the WOZ interface after stage one. This improved the fit and the user reactions.

Fit

The fit in these sessions was determined by the user state. Users were mostly stateless or happy. Occasionally the state confused was given. This was related to the following instances:

- The user was lost in the navigation toward an activity. The WOZ only has navigation interventions following the logical steps toward an activity. In some cases the users did not follow these paths. Then the WOZ could not intervene effectively. In these cases it was incidentally necessary to communicate between the WOZ and the Users. This problem is a known issue and there will be a special module to guide navigation in the AtGentSchool application.
- In the test application there were only a few experts included (as available within the Ontdeknet system). In their search for an expert, the users needed to be cued on the right search term to use.

User reactions

As with stage one, users were asked to fill in a questionnaire after the session, to evaluate the interventions. The user reactions on indicators performance, satisfaction and attention were calculated based on the answers in the questionnaire and are shown in Table 2. (Learning was included in the questionnaire, but again the sessions were not long enough to permit significant learning by the users, so this has been excluded from the results). The maximum score for each indicator is stated in the table header.

User	Performance(20)	Satisfaction(20)	Attention(15)
Session 1	18	17	12
Session 2	17	17	10
Session 3	16	17	13

Table 2 - Stage two - User scores of Performance, Satisfaction and Attention, per session

In these sessions, the emphasis was on regulative support, which was between 50% and 79% of all the interventions (see Table 3). The fact that all three users were new users explains the need for regulative support. The metacognitive interventions were between 30% and 40% of the interventions. The cognitive support was relatively little in the first two sessions—between 15% to 20%—except for the last session were it contained 50%. The first two users finished their tasks rather quickly, so there was little space or need for cognitive support.

Category	Regulative	Metacognitive	Cognitive	Total
Session 1	19	10	5	24
Session 2	25	15	8	48
Session 3	14	11	15	30

Table 3 - Quantitative breakdown of support types per session

4.2.6 Discussion

The main question to be answer was: Is the WOZ able to determine a useful intervention from the Intervention Model for the user based on the information from the Event Model? This experiment answers that question in the affirmative—a useful intervention may be selected in this manner; there is good reason to assume that the Event Model and the Intervention Model are sufficient for the first version of AtGentSchool.

Fit

The fit in these sessions transformed from reasonably low to reasonably high. There was a strong correspondence between the WOZ and the Users on the judgment of fit. This would indicate that the event model (consisting of events and parameters) implemented in the WOZ could provide enough information to determine a useful intervention.

There is one restriction on this conclusion, namely that the WOZ was not only interpreting the events and parameters, but also applying his or her knowledge of the application and learning situation. An analysis of the log files may later provide insights into the reasoning of the WOZ.

User reactions

The user's states were mostly happy, suggesting that users generally appreciated the interventions. Users indicated states of irritation or confusion when the wrong intervention was given, when navigational interventions were not properly introduced or when interventions are wrongly timed.

The user judgment on the indicators performance, satisfaction and attention were high enough to establish that the interventions have a positive effect on their functioning in Ontdeknet.

Adjustment to the models

Event Model - the adjustments made to the Event Model improved the ability of the WOZ to determine a useful intervention. The last version of the Event Model allows for a good selection of agent interventions and is expected to be suitable for use with the first version of AtGentSchool.

Intervention Model - a number of missing interventions were added and adjustments were made to the intervention model based on feedback from the users and the WOZ. The division between interventions provided was around 50% regulative interventions, 30% metacognitive interventions and 20% cognitive interventions. It is expected that these percentages will change with the experience of the user. We can conclude that all

intervention categories are necessary to support the user. The General Intervention Model derived can function as a first version of the intervention framework for AtGentSchool.

A study of the log files will yield information on the interventions used by the WOZs under which circumstances and will assist in formalising the intervention model for use with AtGentNet.

There was an important point raised about the form of the interventions; these were all directive of tone. It was suggested that a more open tone for the interventions could contribute to different user reactions. This is a significant point that will need further attention.

4.2.7 Future evaluations

Information gained using this WOZ system has already greatly assisted the understanding of applying the Conceptual Framework to Ontdeknet system. However, the setting required some adjustment from the users. The interventions were provided in text only without a spoken voice. Users needed to be focussed on the agent to follow the interventions. Participants needed to read very fast. For these reasons it was found that the current WOZ system is not suitable for testing with children. This situation is currently being reviewed.

5. Formative Evaluation – AtGentNet

5.1 Introduction

AtGentNet is to be based upon the ICDT platform—a collaborative learning, group discussion and coordination platform. As such, it differs from AtGentSchool significantly in the form in which it will be employed. AtGentSchool is primarily a guided learning system. Since this requires the students to complete tasks in a relatively predefined sequence, it was possible to define and test a large proportion of circumstances in which interventions (based upon the Conceptual Framework) could be made. With AtGentNet, the circumstance of use differs greatly. Individual users have complete discretion as to their use of the platform. Formative evaluation must therefore be much more flexible than the more formal WOZ approach used for AtGentSchool.

In addition, the user freedom allowed within AtGentNet allows for additional intervention approaches. Perception is one of the four main levels of support identified by the Conceptual Framework. This "bottom-up" process has the ability to "capture" a user's attention, diverting him or her from their current task. Such a diversion may be beneficial or disruptive, according to circumstance. The likelihood of this capture effect depends on many factors of the screen layout and display. AtGentNet has the opportunity to explore this area. The independence of the user allows him / her to select their own route around the platform. It also allows the presentation of information to influence the user's choices. Therefore, the influence of perception, and its effect on the user's attentional choices, is an important aspect within AtGentNet and has been incorporated into the formative evaluation in addition to direct agent-based interventions (as used also for AtGentSchool).

5.2 Agent-based interventions

The AtGentive partners themselves have been major users of the ICDT Platform—it has been the primary medium for collaboration and communication within the project. While the AtGentive partners are not a learning community as such, it was decided that the collaborative and communicative aspects of the platform are the primary areas to which attention intervention will be directed, while learning aspects will be indirectly supported. Thus their use of the platform has given the partners the necessary experience and understanding to enable them to act as participants for the ICDT platform's formative evaluation.

5.2.1 Questionnaire to evaluate scenarios

With the completion of a first draft of the Conceptual Framework deliverable, a more complete set of usage scenarios became available (see Appendix A). These scenarios exemplified the forms of intervention indicated by the Conceptual Framework. The scenarios were modified (see Appendix B) to concretise their descriptions of the concepts involved, making them specifically relevant to the ICDT platform. For example, Scenario one became:

> "You ask to respond to a posting. You spend 10 minutes creating the response. The agent tells you that most other users took just a few seconds and only wrote a short response."

A questionnaire was created to evaluate likely responses to the implementation of these modified scenarios. Respondents were asked to rate their response to each scenario on the following seven point Likert scale (the numbers were not shown to respondents but were used to analyse the data):

1. Really Annoyed
2. Annoyed
3. Prefer not to have this
4. Not particularly bothered
5. Slightly Pleased
6. Happy
7. Very Happy

As users of the platform, the AtGentive partners were in a good position to comment on these scenarios. OBU distributed this questionnaire to the partners to obtain formative feedback on the desirability of implementing these scenarios.

Results

Early quantitative results from 12 respondents are described here. The scenarios under investigation are listed below (see Appendices A and B for further details):

- Scenario 1: Support to task resumption, restoring task context (I)
- Scenario 2: Support to limited time resources allocation
- Scenario 3: Notification of external events
- Scenario 5: User requests notification
- Scenario 4: Learning guidance
- Scenario 6: I don't want to do this … bug me no more!

- Scenario 7: Re-attracting an idle-user attention
- Scenario 8: Re-attracting distracted user's attention
- Scenario 9: Support to task resumption, restoring task context (II)
- Scenario 10: Restore historical context
- Scenario 11: Propose task continuation
- Scenario 12: Suggest community relevant resources
- Scenario 13: Suggest community relevant tasks
- Scenario 14: Task sequencing
- Scenario 15: Encourage slow user

The mean scores obtained are shown in Figure 9. Note that a score of four is the mid-point ("Not particularly bothered", between "Prefer not to have this" and "Slightly Pleased"). Figure 9 shows deviations from this value.

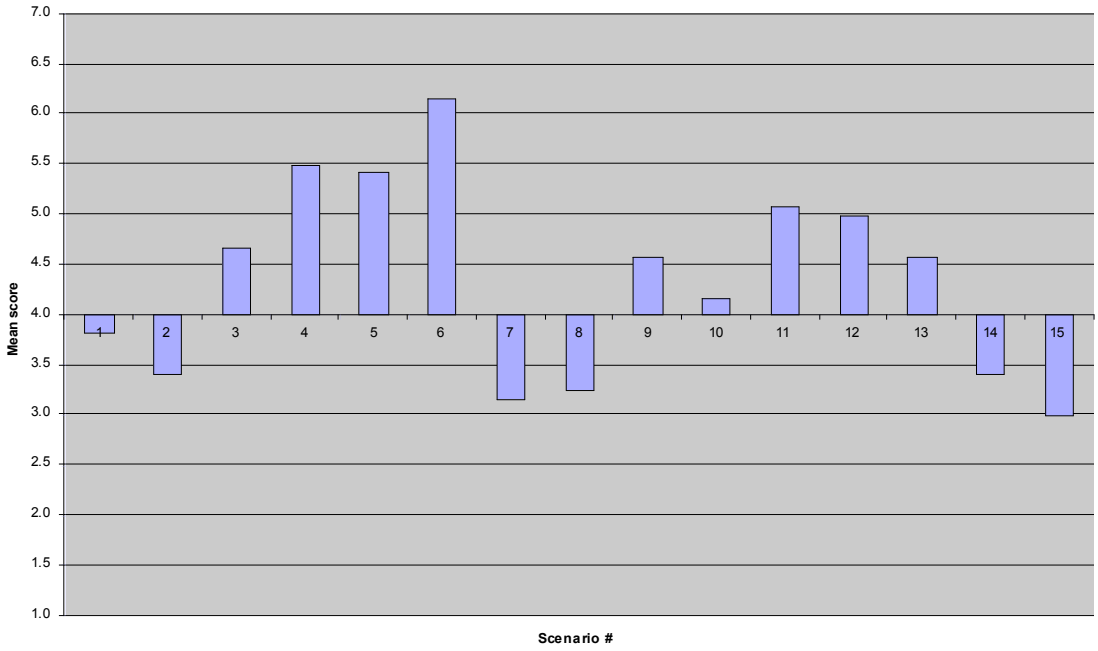


Figure 9 - Mean scores for respondents per scenario

Discussion

Figure 9 shows a range of values, both above and below the mid-value. For example, Scenario 8 (Re-attracting distracted user's attention) was considered the most useful (in the context of the ICDT platform, and thus AtGentNet), while Scenario 15 (Encourage slow user) was seen as the least useful scenario to implement.

At the time of writing, further analysis is ongoing. Discussions are also under way to prepare and distribute a related questionnaire to STC's users of the ICDT platform.

5.3 Perception

As discussed in the introduction to this section, Perception is one of the four main levels of support identified by the Conceptual Framework. The independence of the user when using AtGentNet means that the presentation of information may significantly influence the user's choices. Therefore, the influence of perception, and its effect on the user's attentional choices, represents an important aspect of AtGentNet, and is reflected in the formative evaluation.

5.3.1 Graphical community usage presentation

The formative evaluation process has generated a number of ideas for graphing facilities that could allow users to perceive, and thence reflect upon and change, their collaborative use of the platform. These are:

1. Number of postings made to the platform per day. (Clicking on a day lists the postings made that day)
2. Total number of postings on the platform, ever, broken down by user. (Clicking on a user lists their postings)
3. Number of log-on events per day
4. Total number of times each user has ever logged on
5. A combination of (1.) and (3.), shown as a multiple-line graph
6. The top-20 read postings over the last 24 hours. (Clicking on a posting opens it) (not yet implemented)

Graphs (1.) to (5.) have been incorporated into the ICDT platform to allow more detailed evaluation, which is currently ongoing.

5.3.2 Questionnaire to evaluate graphical community use

A questionnaire was created to evaluate actual and likely responses to the implementation of these graphs. Respondents were asked to rate their response to each graph on the following seven point Likert scale (the numbers were not shown to respondents but were used to analyse the data):

1. No Use At All
2. Unlikely To Use
3. Just Possibly Useful Occasionally
4. Useful Sometimes
5. Quite Useful
6. Useful
7. Very Useful

As with the questionnaire described in Section 5.2.1, the AtGentive partners were in a good position to comment on these graphs. A questionnaire was again used to obtain formative feedback on the desirability of implementing these graphs in AtGentNet.

Results

Early quantitative results are described here. The mean scores obtained are shown in Figure 10. While only a score of one for a graph would indicate that all respondents

considered that graph to be of no practical use, a low score is not favourable to a graph's ultimate implementation in AtGentNet. There is no absolute definition of the number required for a graph to be worth implementing—the quantitative data will be used in tandem with qualitative data collected in order to make that decision. These data are being analysed at the time of writing.

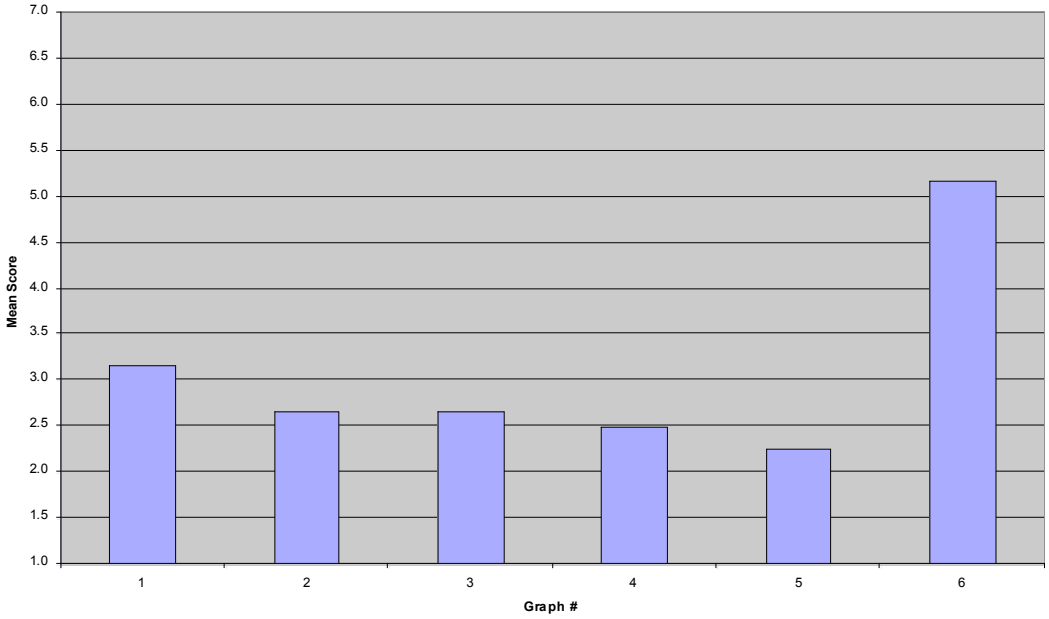


Figure 10 - Mean scores for respondents per graph

Discussion

It may be seen from Figure 10 that only graph number six has a mean score from these respondents that suggests it may be of real use (5.2, or just above "Quite Useful"). Further analysis is ongoing at the time of writing, and will be made available in due course.

5.3.3 Personalised home page

A personalised home page would be able to focus the user's perception on salient data relating to the learner at that time, and the learning community in general. More concretely, the purpose of this home page is to provide to the user with a summary of the most important elements happening on the platform. It is proposed that this "Home Page" will help the user to get an overview very quickly of all the elements that are most relevant to him or her. This should reduce the information overload and cognitive effort otherwise required to collect such information. This idea is currently undergoing development.

6. Further formative evaluation

This document has described the formative evaluation carried out within AtGentive to date, in terms of its four main functions: to validate and enhance the Conceptual

Framework itself (in particular the scenarios of use), to create a profile for embodied agents suitable for AtGentSchool and AtGentNet, and to propose, validate and enhance design elements of AtGentive interventions for each of the specific platforms AtGentSchool and AtGentNet.

The formative evaluation described in this document will continue to act as a reference, informing and guiding the design process and maximising the effectiveness of AtGentive overall. In addition, smaller amounts of continuing formative evaluation may prove beneficial throughout the remainder of the design phase (WP2). Areas that may benefit from such evaluations include the development of avatars, design changes generated by existing formative evaluation, changes to the design that may be necessitated by implementation constraints, and evaluation of early partial implementations prior to the final release. Such additional formative evaluations will be based upon, and add to, the formative evaluation described in this document.

7. References

Angehrn, A. A. (2004). *Designing Effective Virtual Communities Environments: The ICDT Platform* (CALT Report 10-2004): INSEAD CALT.

Bradley, M. M. (2000). Emotion and motivation. In J. T.Cacioppo & L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (2nd ed ed., pp. 602-642).

Chalmers, M. (2000). When Cookies Aren't Enough: Tracking and Enriching Web Activity with Recer In R. Rogers (Ed.), *Published as Preferred Placement: Knowledge Politics on the Web* (pp. 99-102). Maastricht: Jan van Eyck Academie Editions.

Clauzel, D., Roda, C., & Stojanov, G. (2006). Tracking task context to support resumption, *Workshop on Computer Assisted Recording, Pre-Processing, and Analysis of User Interaction Data*. London, UK.

Gould, J. D., Conti, J., & Hovanyecz, T. (1983). Composing letters with a simulated listening typewriter. *Commun. ACM, 26*(4), 295-308.

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: an anger superiority effect. *Journal of Personality and Social Psychology, 54*, 917-924.

Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over the zygomaticus major and corrugator supercilii. *Psychophysiology, 40*, 776-785.

Öhman, A., Hamm, A., & Hugdahl, K. (2000). Cognition and the Autonomic Nervous System: Orienting, Anticipation, and Conditioning. In J. T. Cacioppo & L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (2nd ed ed., pp. 533–575).

Partala, T., Surakka, V., & Lahti, J. (2004). *Affective effects of agent proximity in conversational systems.* Paper presented at the NordiCHI, Tampere, Finland.

Rudman, P., & Zajicek, M. (2006 - to appear). *Autonomous agent as helper - Helpful or Annoying?* Paper to be presented at the IAT 2006 - IEEE/WIC/ACM International Conference on Intelligent Agent Technology, Hong Kong.

Appendix A - Scenarios from the Conceptual Framework

These scenarios are taken from the draft Conceptual Framework as at 02/08/2006 – the version current during the preparation of this report.

Scenario 1: Support to task resumption, restoring task context (I)

The student is working at an assignment. In order to perform this activity he/she has opened the Web page of the course containing the text of the assignment (window 1), a word processor where he/she is typing some text (window 2), as well as a PDF document containing some notes from the professor (window 3). Before completing the assignment, the student switches to another task. Later the student returns to the assignment task; as soon as the student resumes the interrupted task the system proposes to restore the context of the assignment task, as it was left at interruption time, by reopening (or bring to front) the three windows 1, 2, and 3.

Scenario 1: Support to task resumption, restoring task context (I) (Applied to: AtGentSchool)

The student is building the mind map (current focus) using the expert's introduction diary and personal information, as well as a PDF document opened in an Acrobat window. The student switches to a questionnaire. The AtGentSchool application reports a start event for the new activity (questionnaire). The agent saves, possibly with the help of the user, the context of the previous focus (which includes the mind-map window, the diary and information of the expert, and the PDF document window). Later the student returns to the mind-map building activity; the application sends a resume event; the agent proposes to restore the saved context.

Scenario 2: Support to limited time resources allocation

The student starts reading the text for a new lecture. The system recognises that a relevant exercise task was previously interrupted (or that the exercise was previously suggested by the application). The agent also evaluates that the exercise task could be completed within the time available to the student whilst reading the text for the new lecture requires longer than the time available to the student. The system suggests working at the exercise.

Scenario 2: Support to limited time resources allocation (Applied to: AtGentSchool)

The student starts working at the mind-map (start event). The agents recognise that a relevant exercise task was previously interrupted (or that the exercise was previously suggested by the application). The agent also evaluates that the exercise task could be completed within the time available to the student whilst the mind-map task requires longer than the time available to the student. The agent suggests working at the exercise.

Scenario 3: Notification of external events

The user is performing a task. An email addressed to the user (or other notification event), is received. The system recognises that the message is of average importance (e.g. the sender is listed in the user social network, and the subject is relevant to one of the interrupted tasks) however the system also recognises that the current task is urgent and it requires a heavy workload. The system decides to delay notifying the user about

the message until the occurrence of a breakpoint in the task execution (e.g. the user completes the current activity, or starts a new activity).

Scenario 3: Notification of external events

The user is performing a task (e.g. user is working at an assignment in the AtGentSchool application / the user is browsing a space in the AtGentNet application). An email addressed to the user (or other notification event), is received by the application. The application originates a new information available event. The agents recognise that the message is of average importance (e.g. the sender is listed in the user social network, and the subject is relevant to one of the interrupted tasks) however the agent also recognises that the current task is urgent and it requires a heavy workload. The agents decide to delay notifying the user about the message until the occurrence of a breakpoint in the task execution (marked by a new user-application, or user event).

Scenario 4: Learning guidance

The user is reading some information and the application evaluates that the user should also read another document that he / she has not yet explored. The system evaluates the best manner to propose the new focus (on the basis of the user's current and past activity) and makes the suggestion to the user. The user disregards this suggestion (without dismissing it). The system saves the proposed focus to be able to propose it later.

Scenario 4: Learning guidance (Applied to: AtGentNet)

The user is visiting one of the platform's knowledge area and the application evaluates that the user should also visit another knowledge area, which he / she has not explored. The application generates a propose focus event. The agent evaluates the best manner to propose the new focus (on the basis of the current and proposed foci characteristics) and makes the suggestion to the user. The user disregards this suggestion (without dismissing it). The agents save the proposed focus to be able to propose it later.

Scenario 5: User requests notification

The student requests to be notified immediately and with confirmation, about any message coming from a given sender. Upon reception of the email message the system recognises that the conditions for notification are verified, consequently it notifies the user immediately (as requested). Since the user indicated that the notification is with confirmation, the notification is repeated at successive breakpoints until the user acknowledges it.

Scenario 5: User requests notification

The user requests to be notified immediately and with confirmation, about any message coming from a given sender (notify-me). The application, upon reception of email messages, notifies the agents (new-focus). The agents recognise that the user wants to be notified about the email. The agents notify the user immediately (as indicated by the notify-me event). Because the user indicated that the notification is with confirmation, the notification is repeated at successive breakpoints until the user acknowledges it.

Scenario 6: I don't want to do this … bug me no more!

The system proposes to perform a certain task; the user dismisses the proposal. The system will not propose the task again unless the application requires it one more time, in which case the task will be proposed the intervention with further motivation. May ask for reasons for dismissal to the user (e.g. obsolete, too busy, etc.)

Scenario 6: I don't want to do this … bug me no more! (Applied to: AtGentSchool)

A child has logged in the AtGentSchool application and is expected to complete the introduction activity. The child is new to the activity (he / she has never completed the introduction before), has been rated by the teacher as a weak student, has been inactive for a few minutes, and has not reached the introduction screen yet. The agents propose some navigational help explaining how to reach the introduction screen (e.g. "By clicking on the top left button you will reach the introduction screen"). The child dismisses the suggestion. Because the intervention has been dismissed, the Agents will not propose this type of intervention again unless the application requires it, in which case the task will be proposed with further motivation (e.g. "Before you start working at the mind map you must introduce yourself; it looks like you are having troubles reaching the correct screen. By clicking on the top left button you will reach the introduction screen").

Scenario 7: Re-attracting an idle-user attention

The student has started an activity requiring that he / she supplies some input. The student does not provide input for longer than the maximum input inactivity time for the task. The system evaluates whether the task being performed is still the best-suited one for the user; it verifies whether the learner is busy with offline activities. Following these evaluations the system may propose to the user: (1) to continue the task, possibly by providing motivation for the task; (2) to receive help on the task; (3) to switch to another relevant task (if available).

Scenario 7: Re-attracting an idle-user attention (Applied to: AtGentSchool)

The student has started browsing the expert's information (start event). The student does not provide input (idle input) for longer than the time indicated as the maximum input inactivity for the task. The agents evaluate if the task being performed is still the best-suited one for the user. The agents consult the user's agenda to verify whether he / she is busy with offline activities. The agents propose to the user to either: (1) to continue the task, possibly by providing motivation for the task; (2) to receive help on the task; (3) to switch to another relevant task (if available).

Scenario 7a: Re-attracting an idle-user attention (a)

The student initiates a task that he / she has never performed before. The student does not provide input for longer than the time indicated as the maximum input inactivity time for the task. The system proposes to the student to focus on a support task (e.g. explanation, help) for the task just initiated by the user.

Scenario 7a: Re-attracting an idle-user attention (a) (Applied to: AtGentSchool)

The student works at the introduction (start event). He / she has never performed an introduction before. The student does not provide input (idle input) for longer than the time indicated as the maximum input inactivity for the task. The agents propose that the

application should provide support for the introduction task. This support may depend, amongst others, on the input already supplied by the student.

Scenario 8: Re-attracting distracted user's attention

The user is active in an application that is not AtGentive enabled as a consequence AtGentive cannot assess whether the user's current focus is more "important" than any of the foci associated to AtGentive enabled applications and doesn't interrupt the user. However, being able to capture window activities such as copy and paste between windows, or frequent windows switches between an AtGentive-application and an "unknown" application, may allow the system to infer which "unknown" windows are part of the context for the current task and therefore make more informed decisions about the user activity.

Scenario 8: Re-attracting distracted user's attention (Applied to: AtGentNet)

The user is working at a high priority task on the platform: writing a posting that is due in a few hours. The tracking devices recognise that the user is frequently switching between the platform's window for the post-writing and the window of a document D in a word processor (not AtGentive enabled). The agents tentatively associate the word processing window to the context of the post-writing task. Another tracking device reports an idle input event on the post-writing focus. Although this event would normally give rise to an agents' intervention to re-attract the user's attention to the post-writing task, the agents recognise that the user is active in the word processor window for document D. Since this window is associated to the context of the post-writing task, the agents assume that the user is working at the task in another application window and do not intervene.

Scenario 9: Support to task resumption, restoring task context (II)

While browsing a document A, the learner has opened several windows; he/she accesses a new document B; the system proposes to the user to select the windows associated to the interrupted browsing activity on document A, in order to save the context of this activity. Later the user re-accesses document A, the system verifies whether all the windows in the context are already open. If not, it proposes to restore (one of) the saved environment(s) associated to the task of reading document A. The intervention modality will depend, amongst others, on how long the task has been idle.

Scenario 9: Support to task resumption, restoring task context (II) (Applied to: AtGentNet)

While browsing a knowledge area A, the learner has opened several windows; the user enters a new knowledge area B (start event); the agent proposes to the user to select the windows associated to the interrupted browsing activity on A, in order to save the context of that activity. Later the user re-enters the knowledge area A (start or resume event), the agent verifies whether all the windows in the context are already open. If not, it proposes to restore (one of) the saved environment(s) associated to the task of browsing the knowledge area A. The intervention modality will depend, amongst others, on how long the task has been idle.

Scenario 10: Restore historical context

After replying to an email, and reading a document, the user is interrupted while writing a further email. When resuming this last task the system reminds the user that the last actions performed before the interruption consisted in replying to the email and reading the document.

Scenario 10: Restore historical context (Applied to: AtGentNet)

The system will keep track of the sequence in which the user opens Knowledge Assets KAs. For every KA, a 'list' will be held of the KAs that were selected immediately both before and afterwards (referred to as a "contextual Knowledge Asset"—cKA).

When a user selects a KA the system will look at the last time they opened the same KA and offer the user the n (number to be determined) cKAs which (s)he had previously selected immediately before and after the original KA.

Once a KA has been selected n times without accepting the contextual KAs the agent will stop offering cKAs for that particular KA (but the user may ask for contextual KAs at any time).

Scenario 11: Propose task continuation

After N observations the user has executed a certain task X after – or interleaved to – a task Y. The user is now focusing again on task Y. Once the task is completed the system proposes to continue with task X.

Scenario 11: Propose task continuation (Applied to: AtGentNet)

After 10 observations the user has looked at the platform's action-log immediately after reading all new messages on the platform 8 times out of 10. The user is now focusing again on the new messages, once this task is completed the agents proposes to continue the activity by looking at the platform's action-log.

Scenario 12: Suggest community relevant resources

As the learner accesses an online resource, say R1, the system offers a set of "related resources". These related resources correspond to those most frequently selected, by all users, immediately both before and after R1. While the user may select one of the proffered related resources, no action need be taken by the user if they so choose.

When a resource is reopened, (i.e. after the first time for that user) the user will be offered the related resources, as described above, AND any related resource accepted previously.

Scenario 12: Suggest community relevant resources (Applied to: AtGentNet)

The system keeps track of the sequence in which all users open Knowledge Assets (KAs) in the platform. For every KA, a 'league table' is maintained of the KAs most frequently selected immediately both before and after the main KA (we will refer to each of these as a "related Knowledge Asset" - rKA).

When a user selects a KA he / she will be offered the n (number to be determined) rKAs most likely to be of relevance in understanding the KA they chose (i.e. most temporally related).

To reduce the cost of interruption, the user will be offered the additional documents (rKAs) immediately upon selection of a KA. While the user may select one of the proffered rKAs (which will each open in an additional new window), no action need be taken by the user if he / she so choose.

When a KA is reopened, (i.e. after the first time for that user) the user will be offered the most frequently selected rKAs, as described above, AND any rKAs they accepted previous times for the current KA (if they do not now appear as the top n entries in the 'league table').

Once a KA has been selected n times without accepting the related KAs the agent will stop offering rKAs for that particular KA (but the user may ask for related KAs for that KA).

Scenario 13: Suggest community relevant tasks

If a sequence of N events E1 … En generated by this user matches (the event is the same and the task is the same) the beginning of a sequence of M (M>N) events of other users B1, …, Bn, Bn+1, …, Bm, then the task contained in the N+1 event of the sequence (Bn+1) is proposed to this user.

Scenario 13: Suggest community relevant tasks (Applied to: AtGentNet)

The sequence of foci <"read D1 on the platform", "read D2 on the platform"> performed by the current users matches the beginning of the sequence <"read D1 on the platform", "read D2 on the platform", "reply to posting D3"> of 5 out of 6 other members of the community. The agents proposed to this user to continue his/her activity by performing "reply to posting D3".

Scenario 14: Task sequencing

The learner has completed a task T1 that must be followed by task T2. Upon completion of T1, the learner is informed that the next task to be completed is T2. Similarly, other constraints may be defined on tasks sequences, for example, that a task T1 must be completed before initiating task T2.

Scenario 14: Task sequencing (Applied to: AtGentSchool)

The application has informed AtGentive that the task login must be followed by the task introduction. Once the learner has completed a task login he / she is informed that the next task to be completed is the introduction. Similarly, other constraints may be defined on tasks sequences, for example, the introduction must be completed before contacting the expert.

Scenario 15: Encourage slow user

The student initiates a task that he / she has never performed before. The student provides input with a frequency lower than the minimum input frequency for the task.

The system supplies some encouragement and perhaps some simple explanations. When the learner's input frequency increases, the system gives a positive feedback.

Scenario 15: Encourage slow user (Applied to: AtGentSchool)

The student starts with the introduction task that he / she has never performed before. In the start event the application has indicated a minimum input frequency for the task. The student provides input with a frequency lower than the minimum input frequency (low input frequency event). The system supplies some encouragement and perhaps some simple explanations relative to the introduction task. When the learner's input frequency increases, the system gives a positive feedback.

Scenario 16: Tools for various levels of interruption conspiquity

The learner must be notified about new documents available for his/her course. This information is defined as having a low urgency and a high content level. The system will pass on this information as an email.

Later, the learner must be notified about a real time chat meeting with the teacher that will take place in 5 minutes. This information is defined as having a high urgency and a low content level, and an action tracking on the "user connecting in the chat meeting". The learner is notified about the chat event by an instant message.

Later yet, if the user has not connected in the chat event, he is notified, with a further instant message, about the number of participants already in the chat meeting.

Scenario 16: Tools for various levels of interruption conspiquity (Applied to: AtGentNet)

For each entry on the platform, the AtGentNet application generates a new information available event indicating that this is a "new platform entry", that the urgency is low, and that the content level is high. The user has indicated, with a set interruption frequency event, that the maximum interruption frequency for the "new platform entry" information is weekly, and that the interruption modality should be by email. The agents collect all "new platform entry" information and inform the user with a weekly email summarizing the activities of the last period (such as the number of messages that have been posted, the title of the messages, and some indicators of the activity of the community).

Later, the AtGentNet application generates a new information available event indicating that:

- this is a "new chat meeting"
- the urgency is high (the meeting will take place in five minutes)
- the content level is low
- the application requires notification if the user does not connect to the chat within 5 minutes.

The agents notify the user about the chat event with an instant message. Since the user does not login in the chat within 5 minutes, the agents notify the application. The application generates a further new information available event that results in the user receiving a further instant message, about the number of participants already in the chat meeting.

Scenario 17: Task delegation

In a virtual learning community, the community organiser creates a message to be sent to the community, he / she can also indicate presentation style and media, the time of delivery, as well as the operations that should take place after delivery (for instance the message may be archived after it has been read by all recipients, or a reminder may be sent to recipients who did not reply). The system will take charge of completing after delivery actions.

Appendix B – Scenarios Concretised for the Questionnaire

These scenarios are taken from those shown in Appendix one, modified to give a concrete situation in order to consider their use.

Scenario 1: Support to task resumption, restoring task context (I)

You select a posting that you have read before. The posting has an attachment. The agent asks if you would like to see the attachment. You say yes, and the attachment opens at the part you had previously looked at the most.

Scenario 2: Support to limited time resources allocation

You select a posting from the list of unread entries. The posting has a document attached. You click on the document to open it. The agent tells you that the document has 26 pages and asks if you would prefer to save it on your desktop for later and read some of the other unread entries now.

Scenario 3: Notification of external events

Two users that you have had chat conversations with in the last week start using the chat window. The agent tells you that these users are currently using the chat facility.

Scenario 5: User requests notification

You ask that you be notified if a chosen user uses the chat window. Later that day, the agent tells you that the chosen user has placed text in the chat window.

Scenario 4: Learning guidance

You select a recent unread entry from the list. The agent tells you that there is an older unread entry for the same user (that may be needed to understand the current post

Scenario 6: I don't want to do this … bug me no more!

With the previous scenario, when the agent tells you that there is an older unread entry for the same user, you select that the agent stops telling you this sort of thing. You don't get this sort of information again.

Also, how long should this change last?

Scenario 7: Re-attracting an idle-user attention

You have not used the keyboard or mouse for 15 minutes. The agent asks if you would like to be reminded of the most recent postings that you looked at.

Scenario 8: Re-attracting distracted user's attention

You have been "surfing" the web (looking at many pages for a short time) outside of the platform for 10 minutes. The agent asks if you would like to be reminded of the last postings that you looked at.

Scenario 9: Support to task resumption, restoring task context (II)

You select a posting that you have read before. The agent asks if you would like to look at a chat that you participated in last time you had that posting on the screen.

Scenario 10: Restore historical context (Applied to: AtGentNet)

You find an old email with a link to a posting that you read before. You click on that link and the platform opens to show the posting again. The agent tells you that you've read this posting before and asks if you would like to see the other postings that you read around the time that you last looked at this one

Scenario 11: Propose task continuation

You select a posting. The agent knows that you have more than once visited a certain other posting either before or after this one. The agent asks if you would like to look at that posting next. You say "Yes" and the agent adds it to your "unread entries" list.

Scenario 12: Suggest community relevant resources

You select a posting. The agent knows that many other users have read a certain other posting either before or after this one. The agent asks if you would like to look at that posting next. You say "Yes" and the agent adds it to your "unread entries" list.

Scenario 13: Suggest community relevant tasks (Applied to: AtGentNet)

You select two different postings in succession. The agent knows that many other users have read these two postings (in either sequence) and then replied to a third posting. The agent suggests that you reply to this third posting as your next task.

Scenario 14: Task sequencing

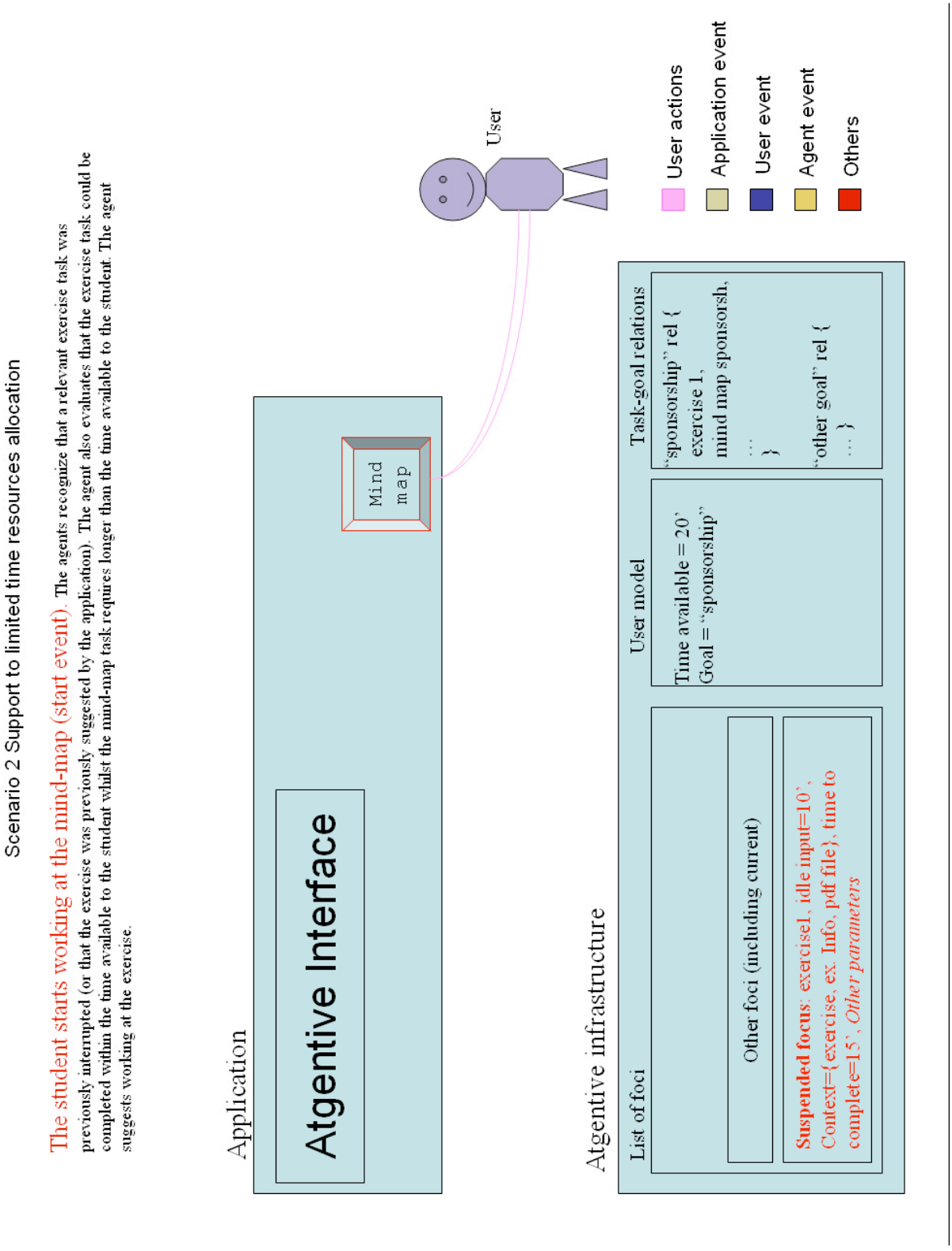
You select a posting from the "unread entries" list. Instead of the posting appearing, the agent tells you that the system administrator has defined that another posting must be read before the one you asked for. The agent asks if you would like to see the other posting.

Scenario 15: Encourage slow user

You ask to respond to a posting. You spend 10 minutes creating the response. The agent tells you that most other users took just a few seconds and only wrote a short response.

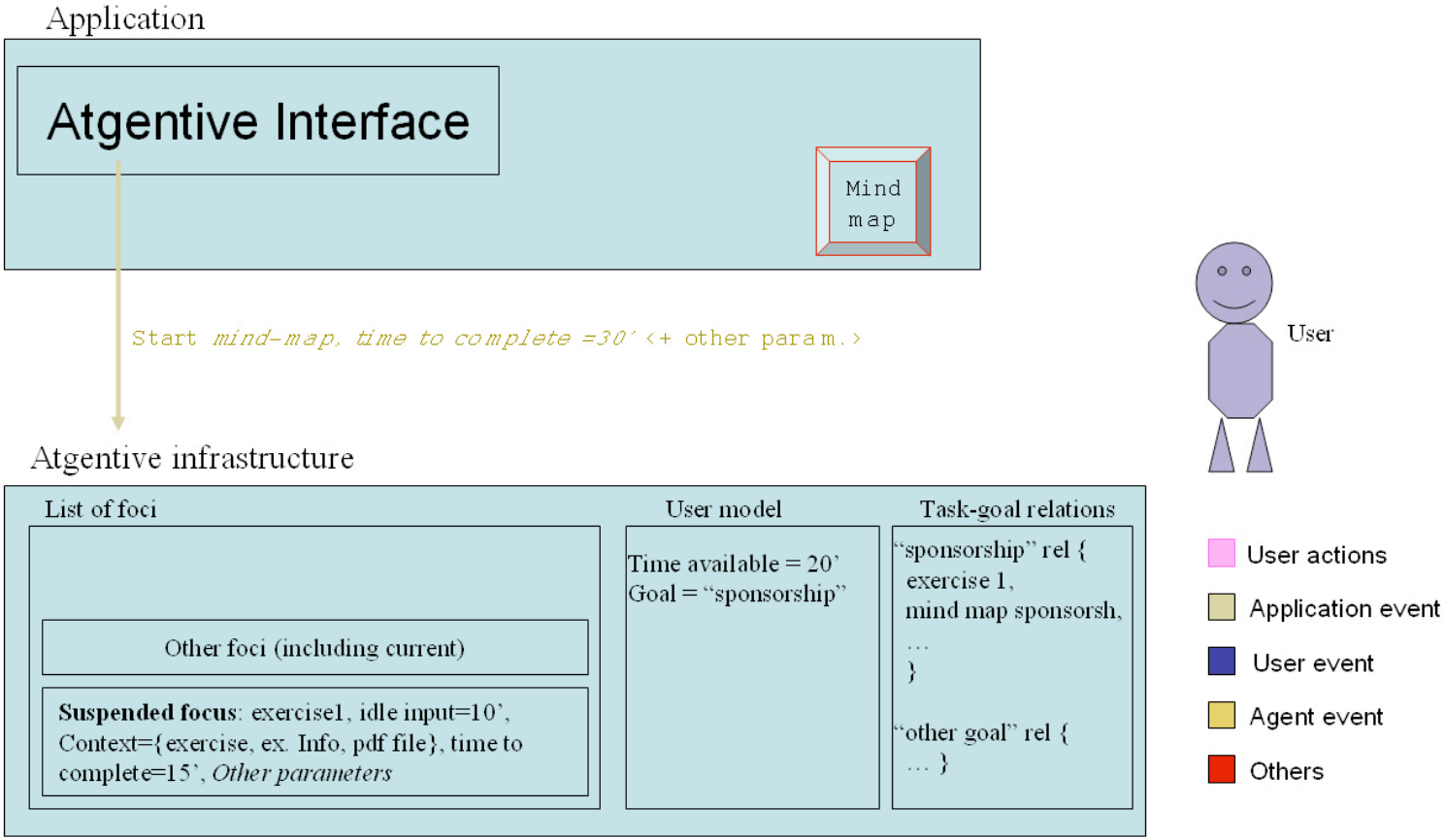
Appendix C – Visual demonstration of the scenarios

As an example of the visual demonstration, the slides for Scenario two are shown below:



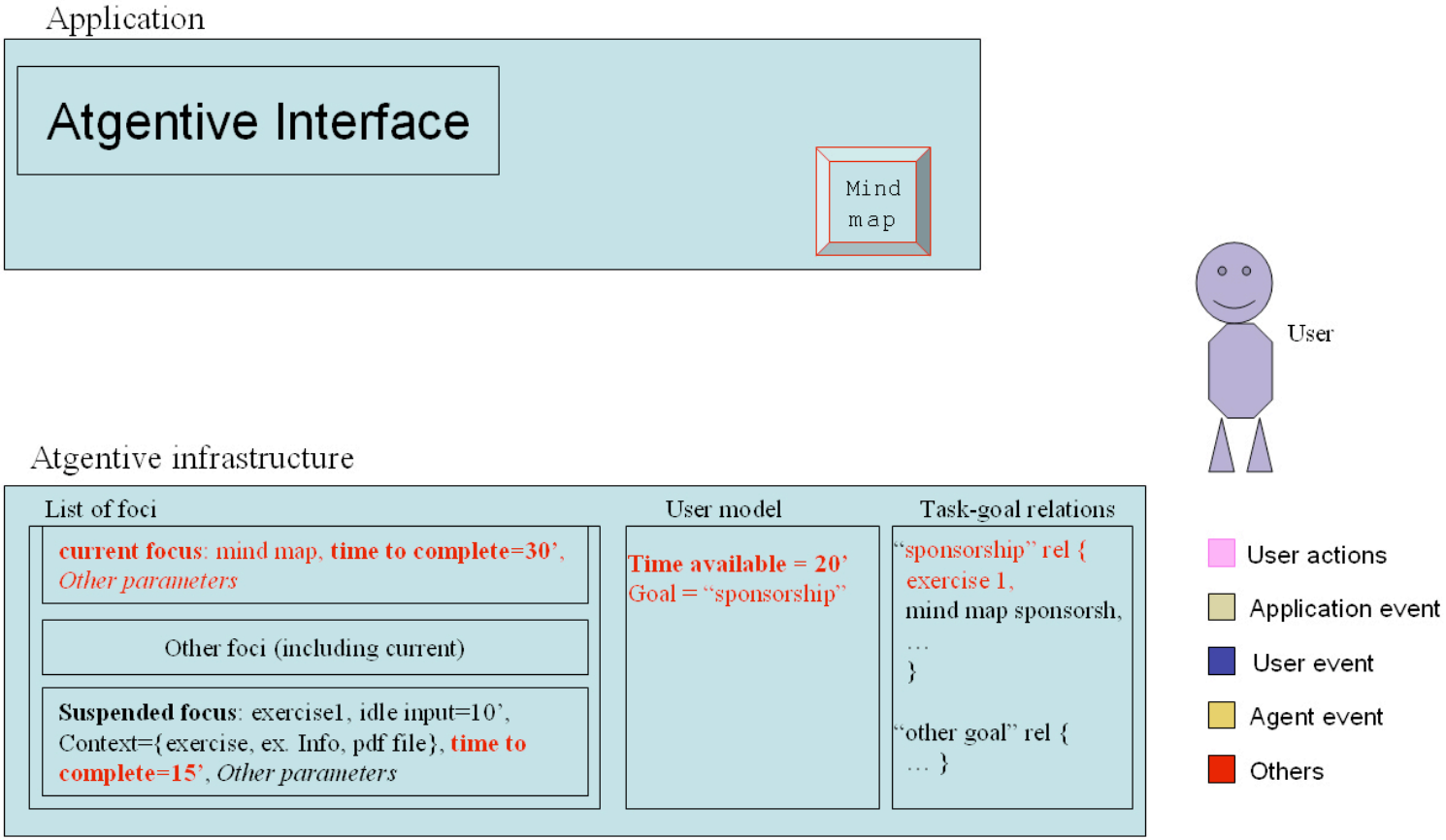
Scenario 2 Support to limited time resources allocation

The student starts working at the mind-map (start event). The agents recognize that a relevant exercise task was previously interrupted (or that the exercise was previously suggested by the application). The agent also evaluates that the exercise task could be completed within the time available to the student whilst the mind-map task requires longer than the time available to the student. The agent suggests working at the exercise.



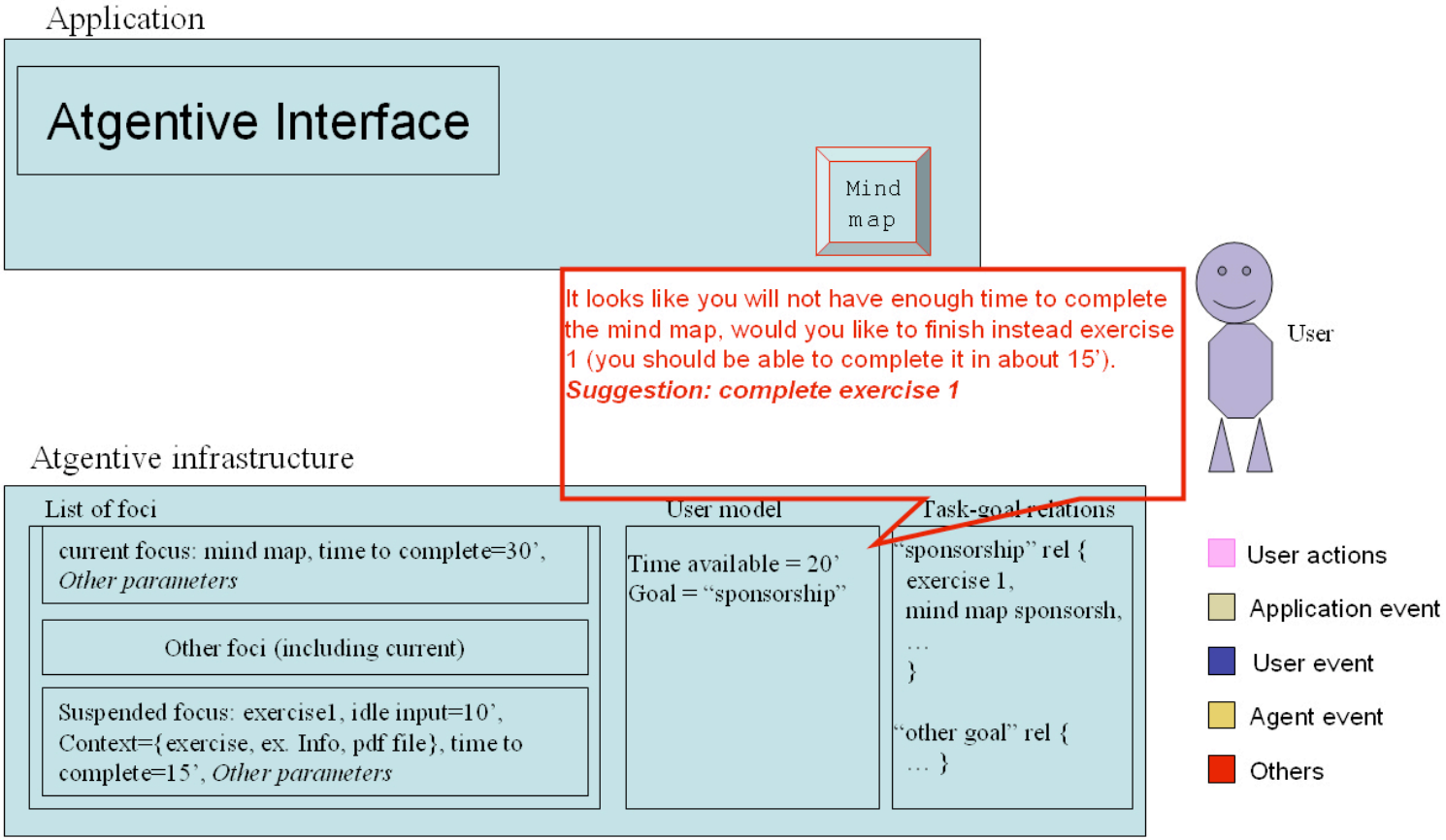
Scenario 2 Support to limited time resources allocation

The student starts working at the mind-map (start event). The agents recognize that a relevant exercise task was previously interrupted (or that the exercise was previously suggested by the application). The agent also evaluates that the exercise task could be completed within the time available to the student whilst the mind-map task requires longer than the time available to the student. The agent suggests working at the exercise.



Scenario 2 Support to limited time resources allocation

The student starts working at the mind-map (start event). The agents recognize that a relevant exercise task was previously interrupted (or that the exercise was previously suggested by the application). The agent also evaluates that the exercise task could be completed within the time available to the student whilst the mind-map task requires longer than the time available to the student. The agent suggests working at the exercise.



Appendix D – Autonomous agent as helper – Helpful or Annoying?

The paper "Autonomous agent as helper – Helpful or Annoying?" (Rudman & Zajicek, 2006 - to appear) has been accepted to the peer-reviewed International Conference on Intelligent Agent Technology. This necessitated the transfer of the paper's copyright to the Institute of Electrical and Electronics Engineers. The paper therefore does not appear in this publicly-accessible deliverable, but is available on request from the authors, or from the conference proceedings.

Appendix E – Formative evaluation of agent characters

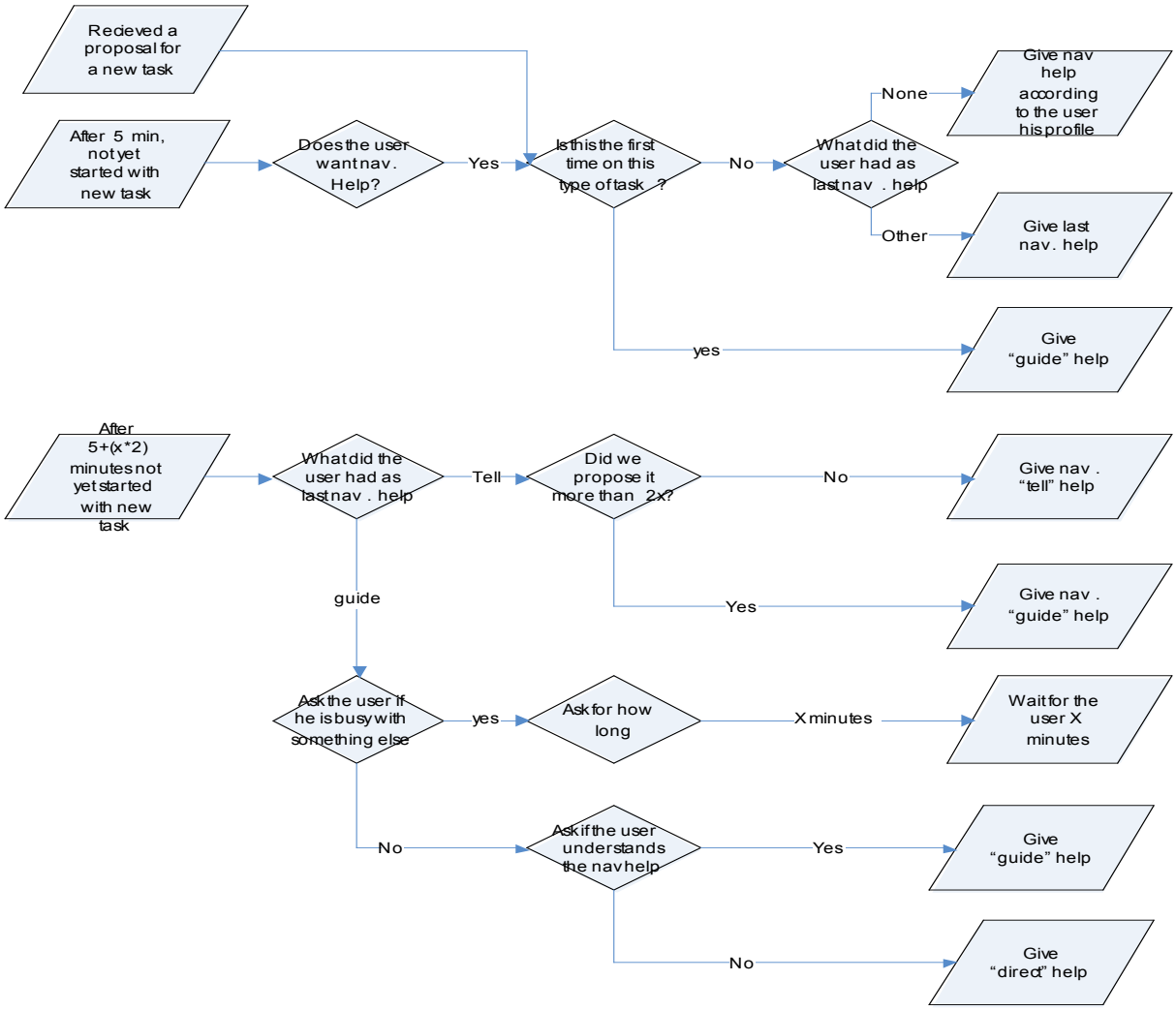
To facilitate a future original publication of the document "Formative evaluation of agent characters", it has not been included in this public document. It is available from the authors on request.

Appendix F – A character for AtGentSchool

To facilitate a future original publication of the document "A character for AtGentSchool", it has not been included in this public document. It is available from the authors on request.

Appendix G – Wizard of Oz Example Decision Tree

This is an example of navigational help. It would occur when the application has proposed a new (required) task but the user has not yet started it. Note that this is an example of how the WOZ may intervene. While the possible user events and potential WOZ interventions are predefined, it is for the WOZ to decide on individual interventions during each trial.



We think that you can have 3 levels of help:

Tell the user; Say where (s)he must go to:

- The user knows the navigation

Guide the user; Show the buttons he must press:

- It's the first time on this task
- He didn't get the "tell" navigation

Direct the user:

- The user doesn't understand the navigation
- There is no time to navigate

The schematic includes some internal parameters (like: how many times did we propose something, is it the first time on task?) The input data has some combined conditions. The first reacts to a "new task proposal" and looks to see if it's the user's first time on this type of task. In the second we received a "propose task" event from the application, but did not receive a "start task" event from the user and it has been 5 minutes since the proposal.

Appendix H – Wizard of Oz Event model

To facilitate a future original publication of the document "Wizard of Oz Event model", it has not been included in this public document. It is available from the authors on request.

Appendix I – Wizard of Oz Intervention Model

To facilitate a future original publication of the document "Wizard of Oz Intervention model", it has not been included in this public document. It is available from the authors on request.