Evaluating the Effects of Displaying Uncertainty in Context-Aware Applications

Stavros Antifakos[1], Adrian Schwaninger[2,3], and Bernt Schiele[1,4]

[1] Perceptual Computing and Computer Vision, ETH Zurich
antifakos@inf.ethz.ch
[2] Department of Psychology, University of Zurich
[3] Max Planck Institute for Biological Cybernetics, Tübingen
adrian.schwaninger@tuebingen.mpg.de
[4] Multimodal Interactive Systems, Darmstadt University of Technology
schiele@informatik.tu-darmstadt.de

Abstract. Many context aware systems assume that the context information they use is highly accurate. In reality, however, perfect and reliable context information is hard if not impossible to obtain. Several researchers have therefore argued that proper feedback such as monitor and control mechanisms have to be employed in order to make context aware systems applicable and useable in scenarios of realistic complexity. As of today, those feedback mechanisms are difficult to compare since they are too rarely evaluated. In this paper we propose and evaluate a simple but effective feedback mechanism for context aware systems. The idea is to explicitly display the uncertainty inherent in the context information and to leverage from the human ability to deal well with uncertain information. In order to evaluate the effectiveness of this feedback mechanism the paper describes two user studies which mimic a ubiquitous memory aid. By changing the quality, respectively the uncertainty of context recognition, the experiments show that human performance in a memory task is increased by explicitly displaying uncertainty information. Finally, we discuss implications of these experiments for today's context-aware systems.

1 Introduction

Context awareness is often seen to be a key ingredient for ubiquitous computing. In the literature, several frameworks and architectures for context awareness have been proposed such as the Context Toolkit [1], Context Fabric [2], or the Location Stack [3]. Experience with context-aware systems however shows that often context is not as simple to deal with as it may seem at first glance. This is mainly due to the inherent uncertainty and ambiguity in context information. Greenberg [4] for example argues that external things, such as objects, the environment, and people, might be relatively simple to capture but that internal things such as people's current interests, objectives, and the state of the activity people are pursuing, is extremely difficult to capture. Bellotti and Edwards [5] even argue that there are human aspects of context that cannot be sensed or

N. Davies et al. (Eds.): UbiComp 2004, LNCS 3205, pp. 54–69, 2004.

even inferred by technological means. Such effects have been reported by others in several application domains [6, 7]. So it is important to take into account that context information might be faulty and uncertain because of missing information, unpredictable behavior, ambiguous situations, and differing interpretations.

Even though many of today's context aware systems do not deal with uncertainty of context information they could be extended to do so. Obviously, systems exist which explicitly model and use uncertainty during inference and decision making. Maybe the most advanced systems like the Lumiere [8] project, the Lookout project [9] or the Activity Compass [10] are based on techniques such as Bayesian modelling and inference, utility, and decision theory.

In the context of ubiquitous computing it has been suggested, however, that modelling uncertainties and advanced inference mechanisms might not be enough. Starting from the observation that there are human aspects of context that cannot be sensed or inferred by technological means, Bellotti and Edwards [5] conclude that context systems cannot be designed simply to act on our behalf. Rather they propose that those systems will have to defer to users in an efficient and non-obtrusive fashion. They also present design principles which support intelligibility of system behavior and accountability of human users. Greenberg [4] also states that actions automatically taken by the system should be clearly linked to the respective context through feedback. Chalmers [11] even argues for "seamful rather than seamless design" to reveal the physical nature of the Ubicomp systems in, for example, the uncertainty in sensing and ambiguity in representations. Mankoff et al. [12] developed a toolkit that supports resolution of input ambiguity through mediation by building on various methods of error correction in user interfaces. More recently Newberger and Dey [13] have extended the Context Toolkit by a so-called enactor component that encapsulates application state and manipulation to allow users to monitor and control context-aware applications. Horvitz and Barry [14] extend their framework to also estimate the expected value of revealed information to enhance computer displays to monitor applications for a time-critical application at NASA.

All of the above-mentioned approaches offer solutions to deal with the inherent uncertainty problem of context information. What is common to all of them is to propose the use of different feedback mechanisms and to involve the user in various degrees and forms. While those approaches are well motivated in their respective application context, it is currently difficult to compare and evaluate those approaches and to judge which of those methods are effective and to which degree.

So, the goal of this paper is to propose and explore a particular way of user feedback and involvement in order to deal with uncertain context information. The proposal is based on the fact that users are actually used to and highly successful in dealing with uncertain information throughout their daily lives. So, rather than using uncertainty of context information to try to infer the most sensible action on behalf of the user with mechanisms such as Bayesian inference, we propose to display this uncertainty explicitly and leverage from the user's ability to choose the appropriate action.

In order to explore the display of uncertainty of context information, we use the running example of a context-aware memory aid. As has been noted by Lamming [15] at the Conference on Ubiquitous Computing 2003: "Forgetting is a truly pervasive problem". Humans tend to forget all sorts of things, ranging from objects and appointments to promises made to friends. While everybody is prone to forget something from time to time, this can have more serious consequences for certain professions, such as for airplane pilots, construction workers, or doctors.

Through two experiments we would like to inform the Ubicomp community if displaying uncertainty is indeed useful as a feedback mechanisms in the sense that it improves human performance in a measurable way. The first experiment is a pure desktop-based study in which we analyze the effects of displaying uncertainty in detail. The second experiment replicates the main findings of the first one in a realistic setting using wireless sensor nodes.

In the following Section we give a brief overview over research on memory aids in the ubiquitous and wearable computing community. In Section 3 we introduce the two experiments, which we use to examine specific aspects of displaying uncertainty information. In Sections 4 and 5 we present the details of the experiments. Finally, in Section 6, we set the results into context and give an outlook on the implications of this work.

2 Memory Aids

Our studies on the effects of displaying uncertainty were motivated by the possibility for ubiquitous computing devices to provide context-aware memory aids. As in other context-aware applications it is difficult to reliably extract context information in scenarios of realistic complexity. Even so, in the last decade quite some effort has been put into developing such memory aids. Lamming and Flynn's "Forget-me-not"-project [16] is one such approach. They build upon the idea that humans can remember things better if they know in what context the events occurred. For example, people are better at remembering where and from whom they received a document, than at remembering the document name. By associating such context information over time with file names, the user has the possibility to remember past events by context. Here, wrongly inferred context information would render such associations useless.

CyberMinder, described in [17], is a tool to support users send and receive reminders. These reminders can be associated with context, making their delivery possible in appropriate situations. The Remembrance Agent [18] is a system that exploits the notes people make on a wearable computer. Whenever a word is entered, the system scans previous data for related notes. Here, the notion of context is limited to previously provided information. Context inference is then similar to an information retrieval task in a database system. Again, the relevance of the retrieved information has a direct effect on how useful the system is. Other research efforts towards building context-aware memory aids can be found in [19] and [20].

Besides such prototypes, approaches have been taken to help people remember objects they have in everyday use. Smart-Its friends [21] is such a technique. Users shake two enhanced objects together, thus allowing them to become "friends". These objects then notify their carrier as soon as they loose communication contact between each other. A similar principle has recently even been introduced as a product, see [22] for details. Such systems rely on the explicit action of the user to build associations between objects. If the system should decide automatically which objects are to be associated, then we need some form of context awareness. Lamming [15] describes such a system that consists of simple low-cost sensor nodes which can store information about proximity to other nodes over a whole day. A simple scripting language enables each node to notify its carrier when an object is out of proximity and thus missing (possibly forgotten). Again, the ultimate goal for such a system would be to infer which objects users want with them at which time. Inferring such information from simple sensor readings may work for certain scenarios, but will undoubtedly cause frustration with users, applied to the full complexity of everyday life. Even if the system takes additional schedules and upcoming events into account it will still be missing much personal information that the user may not even be willing to share.

Rather than implementing our own memory aid, we assume a system exists that can infer for what activity a person is packing and which items he would like to have with him. We further assume that this system would infer the correct activity, and thus the correct set of objects, with some known uncertainty.

3 Experiment Overview

In the following, we give a brief introduction into the experiments detailed in Sections 4 and 5. In both experiments we use numbers instead of different sets of real objects. By taking the semantics out of the experiments, we make the experiments repeatable and generalizable across several people. For example, it may be very unfortunate for some people if they forget their mobile phone, whereas others may not care about the fact. Further, associations between real objects can significantly influence the outcomes of memory tasks.

3.1 A Short-Term Memory Task with an Imperfect Memory Aid

The first experiment is a short-term memory task in which volunteers are asked to remember numbers out of a list. The task is designed to be hard enough so that volunteers can only remember approximately half or even less of the numbers. However, before the user is asked to enter the remembered numbers, the system provides a tip on what the numbers might have been. This tip is equivalent to the notification a context-aware memory aid would provide.

While varying the uncertainty of this tip and whether or not the uncertainty is displayed, we measured participants' performance. Often the users reliance on uncertain information is dependent on the stakes at hand. To be able to control this variable, two groups of participants were tested with opposite costs and gains for correctly remembered and wrongly taken objects, respectively.

The experiment was a four-factorial mixed design including the following independent variables:

– **task difficulty** – by varying the stimulus display duration
– **cost** – by varying the number of points gained and lost for hits and false alarms respectively
– **knowledge about uncertainty** – by displaying the uncertainty or not
– **level of uncertainty** – by varying the quality of the tip

3.2 A Short-Term Memory Task with Sensing and Inference Uncertainty

A large number of applications envisioned by the ubiquitous computing community rely on inference based on uncertain sensor values. For some recent examples see [23, 24, 10]. With this second experiment we hope to gain knowledge about the use of displaying uncertainty in such applications.

Experiment 2 uses wireless sensor nodes to simulate a simple packing scenario in which people have to pack objects. Again, participants have to remember as many numbers as possible from a display on a computer screen. Then they have to pack the respective sensor nodes (see Figure 1) into a cardboard box. The sensor nodes, in turn, use light sensors to detect whether they are being packed or not.

To make the task more realistic, we introduced sets of possible numbers that represent objects, which people may want to take with them at the same time. This concept is based on the vision of having a system that infers for what activity a person is packing. Depending on the inferred activity, a different set of objects is proposed for packing. In other words, if the user often goes swimming on Sunday afternoons and he starts packing his bathing suit, the memory aid will infer the *going swimming* activity. It could then notify the user not to forget the shampoo and a bathing towel assuming he might want to take a shower after swimming.

In our experiment we infer which set of objects is being packed by matching the already packed objects with the possible sets. Uncertainty is introduced at the sensing level by artificially discarding objects that have been sensed as packed and accepting objects that were not sensed packed. As the scenario is tested in a laboratory setting, a high reliability in sensing can be achieved. This makes it possible to produce equivalent sensing uncertainty for all participants of the study. By introducing inference and artificially manipulated sensing uncertainty, we hope to come as close as possible to a real-world scenario without making a controlled experiment unfeasible.

(a) (b)

Fig. 1. Figure (a) shows a participant during a trial run of Experiment 2 packing the sensor nodes. Figure (b) shows the sensor nodes given to the participants.

4 Experiment 1: A Short-Term Memory Task with an Imperfect Memory Aid

Experiment 1 consists of a short-term memory task aided by an imperfect memory aid. Subjects were asked to remember as many numbers as possible from a list of 10 numbers (chosen from 1–20) that is only displayed a very short time. We call this the *subject's task*. After seeing the numbers the participants can enter what they believed to have seen in an array of checkboxes. To aid the user, the program displays a tip by marking some of the numbers in red. This tip is generated by choosing each object from the subjects task with probability p and the other objects with probability $1 - p$. For an example see Figure 2.

4.1 Method

Subjects: 24 students from either the Computer Science department of ETH Zurich or the Psychology department of the University of Zurich participated in this study. Nine were female and fifteen were male. The median age of the participants was 25. All participants reported normal or corrected-to-normal vision.

Design: This experiment was a mixed design study with four independent variables. The participants were randomly distributed between two equally sized groups. The cost variable was tested between groups. This means that both groups completed the same set of experiments with the only difference being the cost function. The *low-cost* group received two points for each correct answer (hit) and minus one point for each wrongly checked answer (false alarm). The *high-cost* group oppositely received only one point for a hit and minus two points for a false alarm.

Four blocks were carried out with each participant in each group. In all blocks the task display duration was randomized between the three values 200, 800, and

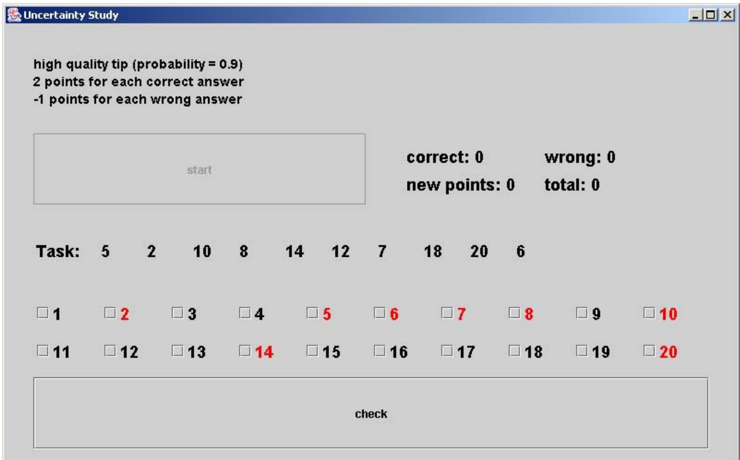

Fig. 2. Screenshot of Experiment 1 with the task field displaying the numbers to be remembered. The task line will disappear as soon as the display duration is elapsed. Information about the tip uncertainty is displayed in the upper left corner. The tip given by the computer consists of the red numbers.

3200 milliseconds. This was approximately perceived as being able to see hardly any objects (two to three), about five, and all of the objects, respectively. Even with the long display time it is hard to remember all ten objects due to the limitations of human memory.

One block did not display any information about uncertainty. Within this block the uncertainty level was randomized between the tip probabilities of 0.6, 0.75, and 0.9. The other three blocks had a fixed tip probability level marked with low ($p=0.6$), medium ($p = 0.75$), and high ($p = 0.9$). It was explained to the subjects that on average, the low quality tip ($p = 0.6$) would render 6 correct objects, the medium 7.5 and the high would render 9 objects. The order of these blocks was counterbalanced using a Latin Square design.

For each time and probability level, 10 trials were completed, resulting in 90 trials for the blocks with displayed uncertainty and 90 trials for the block with no uncertainty displayed. In total, each participant completed 180 trials.

Equipment: The experiment was conducted using a personal computer running Windows 2000 with the screen resolution set at 1280x1024 on a TFT screen. A program was written to display the memory task and to accept the users answers (see Figure 2).

Procedure: First, the experimental settings were explained to each participant. Next, the graphical user interface elements were described. Each participant was told to try to make as many points as possible in accordance with the actual

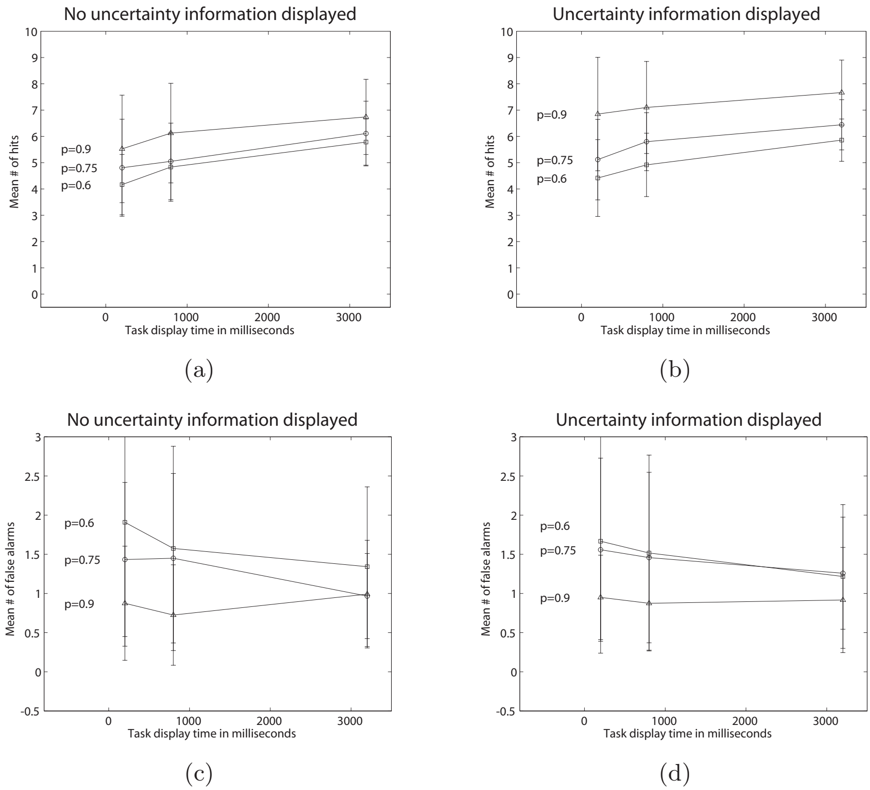

Fig. 3. Results from the low-cost group. The Figures suggest the increase in hit rates when uncertainty information is displayed (compare Figures (a) and (b)). False alarm rates remain similar for both conditions.

cost situation. Prior to the experiment, 20 practice trials were completed using a random order. Each of the four different block settings was represented by 5 trials.

4.2 Results

Figure 3 displays hit and false alarm rates for the low cost condition and Figure 4 for the high cost condition. The plots suggest that displaying uncertainty information results in higher hit rates, especially when tips of high probabilities are shown. This effect seems to be more pronounced in the most difficult condition (short display times). Both effects on hit rates seem to be more pronounced in the high cost condition. The effect of displaying uncertainty is less clear when false alarm rates are concerned, but false alarm rates are substantially reduced in the high cost condition.

The conventional cut-off of $p < .05$ was used for all tests of statistical significance in this study. The performance measures (hit and false alarm rates)

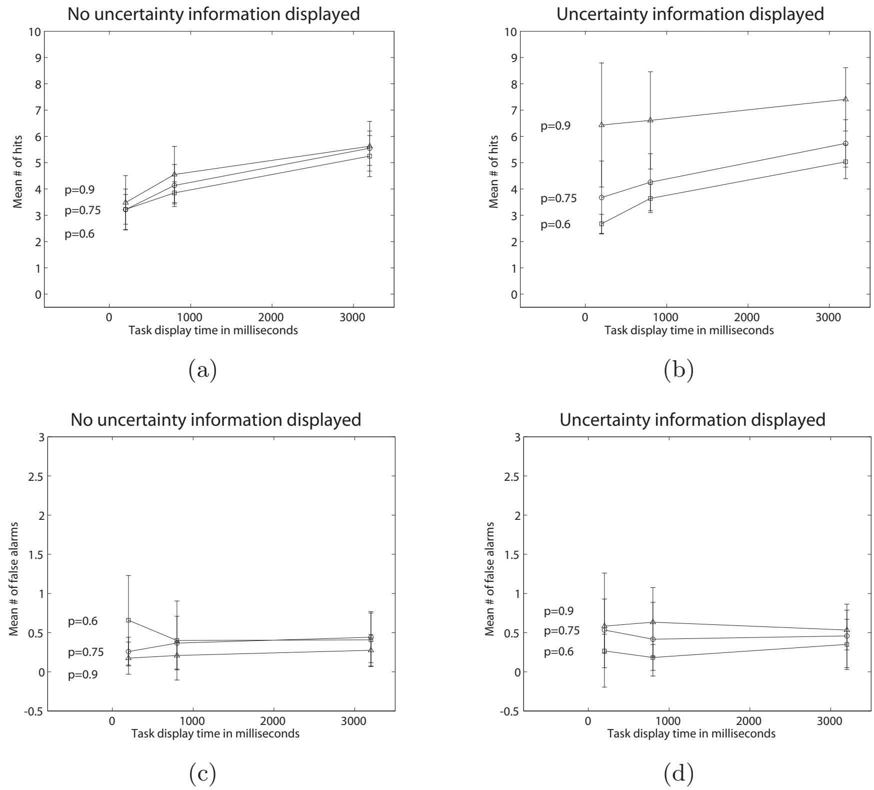

Fig. 4. Results from the high-cost group. The Figures again suggest a large increase in hit rates when uncertainty information is displayed (compare Figures (a) and (b)). False alarm rates remain similar independent of the display of uncertainty information. However, false alarm rates are generally lower than in the low-cost group.

were subjected to a multivariate analysis of variance (MANOVA) with cost (low vs. high) as between-subjects factor and the following within-subject factors: Task difficulty (display times of 200, 800, 3200 milliseconds), knowledge about uncertainty (displayed uncertainty or not), and level of uncertainty (tip probabilities of 0.6, 0.75, 0.9). All main effects were significant. Providing the knowledge about uncertainty affected performance, $F(2, 21) = 8.32$, $p < .01$, as well as costs, $F(2, 21) = 6.27$, $p < .01$, level of uncertainty, $F(4, 19) = 17.50$, $p < .001$, and task difficulty, $F(4, 19) = 65.64$, $p < .001$. There was an interaction between knowledge about uncertainty and the level of uncertainty, $F(4, 19) = 6.52$, $p < .01$. There was also a three-way interaction between task difficulty, knowledge and level of uncertainty, $F(8, 15) = 3.19$, $p < .05$. No other interactions reached statistical significance.

Since the effects of providing the knowledge of uncertainty were of main interest in this study, selective univariate analyses were carried out on hit and false alarm rates regarding main effects and interactions of this factor with the other factors. Providing knowledge about uncertainty affected hit rates, $F(1, 22) = 15.32$, $p < .001$, but not false alarm rates. There was an interaction with the level of uncertainty, both for hit rates, $F(2, 44) = 18.08$, $p < .001$ and for false alarm rates, $F(2, 44) = 7.06$, $p < .01$. The interaction between providing the knowledge of uncertainty and task difficulty was significant only for hit rates, $F(2, 44) = 3.49$, $p < .05$. There was also a three-way interaction between costs, knowledge and level of uncertainty for hit rates only, $F(2, 44) = 4.26$, $p < .05$. No other interactions involving knowledge about uncertainty were significant.

4.3 Discussion

Experiment 1 clearly showed that displaying the degree of uncertainty affected performance. Showing uncertainty information had a clear effect on hit rates. They increased substantially when uncertainty information was displayed, especially when tips of high quality were shown and when the task was difficult. This effect was more pronounced in the high-cost condition. The effect of displaying uncertainty is less clear when false alarm rates are concerned, but they were substantially reduced in the high-cost condition.

5 Experiment 2: A Short-Term Memory Task with Sensing and Inference Uncertainty

As mentioned above, Experiment 1 was designed to test for main effects and interactions between knowledge and level of uncertainty, costs and task difficulty. The aim of Experiment 2 was to replicate the main results of Experiment 1 in a more realistic setting using a less complex experimental design. To this end, a two-factorial design was used in which knowledge and level of uncertainty was manipulated.

The main difference to Experiment 1 is that we introduce real sensing with wireless sensor nodes and inference based upon this uncertain sensing. Smart-Its were used as sensor nodes; for details see [25, 26]. In the Experiment, participants have to remember as many numbers as possible from a list of 7 numbers between 0 and 9. Then they have to physically pack the Smart-Its that represent the numbers into a cardboard box. These detect whether they have been packed or not using a light sensor. Packing objects into a closed cardboard box makes sensing a simple task. To guarantee perfect recognition during all the experiments, an operator constantly checked whether the correct objects were sensed.

To vary the uncertainty in a controllable manner we introduced artificial sensing uncertainty. This was done by only propagating sensing information from packed objects with a certain probability. Similarly, objects that were sensed as not being packed can be regarded as packed by the system. Upon this artificially uncertain packing data, we add inference to determine which objects the user should pack next. This is done by introducing five groups of seven randomly chosen numbers. One of these groups is the actual list of objects the user is trying to pack. The group to be packed by the user is inferred by calculating the matching probability between what has supposedly been packed and all the possible groups.

Figure 5 displays a scenario in which it is most probable to pack the objects 0 and 1 based on an artificial sensing probability of 0.9. Figure 1 displays a person packing the nodes and the wireless sensor nodes in detail.

5.1 Method

Subjects: 10 students from either ETH Zurich or the University of Zurich participated in this study. Two were female and eight were male. The median age of the participants was 28. Two people had participated in the first study. All participants reported normal or corrected-to-normal vision.

Design: A two-way within-subjects design was used. The first independent variable tested the benefit of displaying uncertainty information. The second variable was level of uncertainty. Each participant completed three blocks. In the first block no uncertainty information was displayed. The uncertainty level however, was varied randomly between 0.7 and 0.9. In the second and third block the uncertainty information was displayed. Once it was set to 0.7 and once to 0.9. Block order was counterbalanced using a Latin Square design. For all experiments the task display time was held constant at 400 milliseconds, which let the participants remember 4 numbers on average. Costs were held constant at 1 point for each correct answer and minus 2 points for each wrong answer. Each participant completed 10 trials for the blocks with uncertainty displayed (0.7 and 0.9). In the block with randomized uncertainty and no uncertainty information displayed, 20 trials were completed. This resulted in 40 trials per subject.

Equipment: The experiment was conducted using a personal computer running Windows 2000 with the screen resolution set at 1280x1024 on a TFT screen. A program was written to display the memory task and to display the inferred results (see Figure 5). The program communicated with a sensor node via the serial port of the personal computer. This node acted as a receiver for the data from the other 10 sensor nodes. Finally, we gave the participants 10 Smart-Its sensor nodes and a cardboard box for packing them (see Figure 1). More technical details on the sensor nodes used can be found in [26].

Procedure: The participants were introduced to the experiment by letting them imagine they had a system at home that could help them during packing for different occasions. It was explained that the system would infer which occasion one is packing for and would give hints based on this inference. After the introductory example, the user interface and the handling of the sensor nodes was explained. Each user completed at least two trial runs with each of the three experimental conditions. The order of the three experimental blocks was counterbalanced using a Latin Square design.

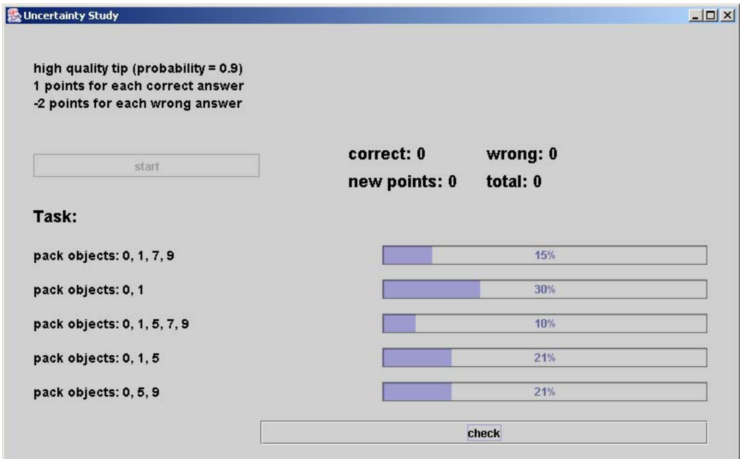

Fig. 5. Screenshot of Experiment 2 after the numbers to be remembered have been displayed. Information about the tip probability is displayed in the upper left corner. The lower part of the screen displays the five possible groups of objects to be packed. Based on the objects that were supposedly packed it is most probable to continue packing with objects 0 and 1. (second line of "pack objects:")

5.2 Results

Hit and false alarm rates were subjected to univariate analyses of variance (ANOVA) with knowledge about uncertainty (displayed uncertainty or not), and level of uncertainty (tip probabilities of 0.7 and 0.9) as within-subjects factors.

As in Experiment 1, there was a main effect of knowledge about uncertainty for hit rates, $F(1, 9) = 6.11$, $p < .05$, but not for false alarm rates. The level of uncertainty also affected hit rates, $F(1, 9) = 9.63$, $p < .05$, while there was no main effect on false alarm rates. In contrast to the results of Experiment 1, the interaction between knowledge and level of uncertainty did not reach statistical significance in Experiment 2, neither for hit rates nor for false alarm rates. [5]

[5] It must be noted however, that the smaller sample size and/or the different uncertainty levels used in Experiment 2 could have prevented revealing these interactions.

5.3 Discussion

Experiment 2 provides converging evidence for the view that displaying uncertainty information increases performance in terms of hit rates, whereas false-alarm rates are much less – if at all – affected. Thus the main finding of Experiment 1 was replicated in the more realistic setting used in Experiment 2.

6 General Discussion and Conclusions

For context-aware systems, we often cannot rely on the assumption that context information is highly accurate. Several proposals have been made to deal with those ambiguities and uncertainties through various feedback, monitor, and control mechanisms. However, their respective strength is hardly known since they are rarely evaluated. In this paper, we propose a simple but effective feedback mechanism by displaying the uncertainty of context information. The effectiveness of the feedback mechanism is shown and replicated in two different user studies in the context of a ubiquitous memory aid.

In the first experiment, we analyzed the effects of four factors and their interactions. Displaying uncertainty information resulted in a substantial increase in hit rates when tips of high quality were shown. This benefit was more pronounced for high task difficulty in high-cost situations. False-alarm rates were less affected by displaying uncertainty, whereas a substantial reduction was observed in high-cost situations.

While the first experiment was desktop-based only, experiment 2 was designed in a way as to make the setting as realistic as possible for a Ubicomp scenario. Therefore, we introduced physical objects with sensing, communication and processing capabilities. In order to avoid, however, that humans add too much semantic meaning to the individual objects by having for example objects like keys, towels, pens, or coats, we still used numbered objects. This 'semantic-free' setting allows to compare the results across people by reducing the semantic bias of each individual person. In this more realistic setting, the main results of Experiment 1 were replicated. Both the display of uncertainty and the level of uncertainty showed significant effects on hit rates, whereas the false-alarm rate remained constant.

One issue to be considered in future work is the tradeoff between the cognitive load, which displaying uncertainty information causes, and the added value that it provides. First design guidelines can be gained from the field of signal detection theory in cognitive science. Results presented in [27] show that people perform best in a signal detection task when uncertainty information is encoded as luminance of a display element. This means it is effective to display more-certain information in a brighter mode than less-certain information. However, as feedback presented on a computer screen is only one of many possible modalities in a ubiquitous computing scenario, it remains to be shown how such results can be transferred.

Experiments with similar objectives have also been carried out in domains with very high costs, such as air traffic control and military pilot training [28, 29, 30]. Here the subjects are highly-trained individuals that have practiced dealing with uncertainty information. In our experiments we show that equivalent results can be achieved with untrained individuals.

Another effect mentioned by several participants is that when uncertainty is displayed, it is easier to understand what the system is doing and how well it is doing it. This postulates that displaying uncertainty information as feedback may be a possibility to build intelligible context-aware systems, as desired by Bellotti & Edwards [5].

Last but not least, we'd like to argue that the procedure we adopted in this paper using two user studies has several interesting properties and might be more widely applicable in the context of Ubicomp. In the first experiment, we used a rather idealistic desktop-setting which allowed us to employ a 4-factorial analysis. Looking at four factors simultaneously would be quite hard and time-consuming in a realistic Ubicomp setting. This first experiment then allowed to measure the most significant effects involved and to test those in a different second experiment. We designed the second experiment then to be more realistic in the Ubicomp sense and used a 2-factorial analysis using the two most important factors from the first experiment. In our case, this way of proceeding has four interesting properties. The first is that the experimental design of the second experiment is informed from the first experiment. The second advantage is that the experiment involves only 2 and not 4 factors as in the first experiment and therefore makes it more feasible as a Ubicomp experiment. The third advantage is that the second experiment itself is more realistic. Finally, the fourth advantage is that we were able to replicate the most important findings in two different experiments.

References

[1] Salber, D., Dey, A.K., Abowd, G.D.: The context toolkit: Aiding the development of context-enabled applications. Technical report, Georgia Institute of Technology (2000)

[2] Hong, J.L.: The context fabric: an infrastructure for context-aware computing. In: Conference on Human Factors in Computing Systems (CHI). (2002)

[3] Graumann, D., Lara, W., Hightower, J., Borriello, G.: Real-world implementation of the location stack: The universal location framework. In: Proceedings of the 5th IEEE Workshop on Mobile Computing Systems & Applications (WMCSA 2003), IEEE Computer Society Press (2003) 122–128

[4] Greenberg, S.: Context as a dynamic construct. Human-Computer Interaction **Volume 16** (2001) pp.257–268

[5] Bellotti, V., Edwards, K.: Intelligibility and accountability: Human considerations in context-aware systems. Human-Computer Interaction **Volume 16** (2001) pp.193–212

[6] Cheverst, K., Davies, N., Mitchell, K., Efstratiou, C.: Using context as a crystal ball: Rewards and pitfalls. Personal Technologies Journal **Volume 3** (2001)

[7] Edwards, W.K., Grinter, R.E.: At home with ubiquitous computing: Seven challenges. In: Conference on Ubiquitous and Handheld Computing (UbiComp). (2001)

[8] Horvitz, E., Breese, J., Heckerman, D., Hovel, D., Rommelse, K.: The lumiere project: Bayesian user modeling for inferring the goals and needs of software users. In: Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence. (1998)
[9] Horvitz, E.: Principles of mixed-initiative user interfaces. In: Conference on Human Factors in Computing (CHI). (1999) 159–166
[10] Patterson, D.J., Liao, L., Fox, D., Kautz, H.: Inferring high-level behavior from low-level sensors. In: Proceedings of the Internation Conference on Ubiquitous Computing (UbiComp). (2003)
[11] Chalmers, M., MacColl, I.: Saemful and seamless design in ubiquitous computing. In: Workshop At the Crossroads: The Interaction of HCI and Systems Issues in UbiComp. (2003)
[12] Mankoff, J., Hudson, S., Abowd, G.: Interaction techniques for ambiguity resolution in recognition-based interfaces. In: UIST 2000. (2000)
[13] Newberger, A., Dey, A.: System support for context monitoring and control. In: Workshop At the Crossroads: The Interaction of HCI and Systems Issues in UbiComp. (2003)
[14] Horvitz, E., Barry, M.: Display of information for time-critical decision making. In: Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence. (1995)
[15] Lamming, M.: SPECs: Another approach to human context and activity sensing research, using tiny peer-to-peer wireless computers. In: International Conference on Ubiquitous Computing (UbiComp). (2003)
[16] Lamming, M., Flynn, M.: Forget-me-not: Intimate computing in support of human memory. In: FRIEND21, International Symposium on Next Generation Human Interface. (1994)
[17] Dey, A.K., Abowd, G.D.: CybreMinder: A Context-Aware System for Supporting Reminders. In: Handheld and Ubiquitous Computing (HUC'00), Bristol, Uk (2000)
[18] Rhodes, B.: The wearable remembrance agent: A system for augmented memory. In: Proceedings of the International Symposium on Wearable Computing. (1997)
[19] Kern, N., Schiele, B., Junker, H., Lukowicz, P.: Wearable sensing to annotate meeting recordings. In: Proceedings of International Symposium on Wearable Computers (ISWC), Seattle, USA (2002)
[20] Antifakos, S., Holmquist, L.E., Schiele, B.: Grouping mechanisms for smart objects based on implicit interaction and context proximity. In: Adjunct Proceedings of International Conference on Ubiquitous Computing (Ubicomp), Seattle, USA (2003)
[21] Holmquist, L.E., Mattern, F., Schiele, B., Alahuhta, P., Beigl, M., Gellersen, H.W.: Smart-its friends: A technique for users to easily establish connections between smart artefacts. In: Proceeding of the International Conference on Ubiquitous Computing (UbiComp). (2001)
[22] Dipo: Memory Aid. http://www.dipo.fr/ (2004)
[23] Neustaedter, C., Greenberg, S.: The design of a context-aware home media space for balancing privacy and awareness. In: Proceedings of the Internation Conference on Ubiquitous Computing (UbiComp). (2003)
[24] Koile, K., Tollmar, K., Demirdjian, D., Shrobe, H., Darrell, T.: Activity zones for context-aware computing. In: Proceedings of the Internation Conference on Ubiquitous Computing (UbiComp). (2003)
[25] Smart-Its: Project web-page. http://www.smart-its.org/ (2004)

[26] Beigl, M., Zimmer, T., Krohn, A., Decker, C., Robinson, P.: Smart-its - communication and sensing technology for ubicomp environments. Technical Report ISSN 1432-7864 2003/2, TecO Karlsruhe (2003)

[27] Montgomery, D.A., Sorkin, R.D.: Observer sensitivity to element reliability in a multi-element visual display. Human Factors **Volume 38** (1996) 484–494

[28] Nicholls, D., Barsotti, V., Battino, P., Marti, P., Pozzi, S.: Presenting uncertainty to pilots and controllers. In: 5th International seminar on ATM, FAA and Eurocontrol, Budapest (2003)

[29] Gempler, K.S., Wickens, C.D.: Display of predictor reliability on a cockpit display of traffic information. Technical Report ARL-98-6/Rockwell-98-1, Aviation Research Lab, Institute of Aviation, University of Illinois (1998)

[30] Banbury, S., Selcon, S., Endsley, M., Gorton, T., K.Tatlock: Being certain about uncertainty: How the representation of system reliability affects pilot decision making. In: Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting. (1998)