Towards the right assistance at the right time for using complex interfaces

Blandine GINON
Université de Lyon, CNRS, UMR5205, F-69622, France
blandine.ginon@liris.cnrs.fr

Simone STUMPF
City University London, Centre for HCI Design, School of Informatics
London EC1V 0HB, United Kingdom
Simone.Stumpf.1@city.ac.uk

Stéphanie JEAN-DAUBIAS
Université de Lyon, CNRS, UMR5205, F-69622, France
stephanie.jean-daubias@liris.cnrs.fr

ABSTRACT
Many users struggle when they have to use complex interfaces to complete everyday computing tasks. Offering intelligent, proactive assistance is becoming commonplace yet determining the right time to provide help is still difficult. We conducted an empirical study that aimed to uncover what user factors influenced following advice. Our results describe a user's background and expectations that appear to play a role in heeding assistance. Our work is a step towards understanding how to provide the right assistance at the right time and build proactive assistance systems that are personalized for individual users.

CCS Concepts

• **Human-centered computing → Human computer interaction (HCI)**; user model, user studies, graphical user interfaces.

Keywords
User assistance; predictive model; proactive assistance; activity monitoring; trace-based systems.

1. INTRODUCTION
While user interfaces should be as intuitive and easy to learn as possible for a wide audience, many users need help and guidance to complete their tasks. Traditional help systems typically are reactive: they require users to recognize that they need help, enter a help mode, and search/browse for the required topic. More recently, research has been directed at designing and implementing *proactive assistance* in a variety of task-based systems [1][8][9]. However, identifying when best to intervene with advice is still challenging 0. Too soon and the assistance will interrupt the user unnecessarily, but helping too late means that the user is left struggling.

It has been recognized that every user is different and some progress has been made to personalize assistance by developing predictive models that take user characteristics into account [1][3], for example, knowing a user's self-efficacy [5] might allow the system to be tailored to provide assistance at the right time.

In this work, we investigated the impact of user characteristics that influence the reaction to proactive help messages. We conducted an empirical study using an assistance system that monitored users' interactions with a photo-editing application, and if they deviated from expected tasks, provided assistance for carrying out the next action. We collected user characteristics, interaction logs and preference data during the study to build a predictive model of when to provide assistance. Our research contributes to a better understanding of the impact of proactive assistance on user satisfaction, and provides first steps toward predicting the right time to provide help to users based on their individual background.

2. RELATED WORK
Proactive assistance systems interact with users by embedding some intelligence in the user interface. All task-based assistance systems rely on being able to monitor and trace the user's activity at a fine-grained level, e.g. [6][9][11][16]. There is prior research into how to identify the task a user is carrying out, when users switch from one task to another and optimal times to interrupt them (e.g. [10][14]) but our approach is trying to help users *continue* with their task. Most research in proactively assisting users has focused on determining the task that users are doing that requires assistance, and possibly automating it, while there is little research to detect the right time to assist the user.

Determining the task structure from traces of user activities takes lots of data. Hence, most assistance systems, especially in an educational setting, are designed with an *a priori* structuring of the task, by scheduling subtasks that are required to complete the overall task. This can be done either by the user or by a teacher [7]. Our study uses such a system in which the task structure is predefined.

Some Intelligent Tutoring Systems intervene based on a user's characteristics, such as their current emotion and motivation (e.g. [2][4]). For example, the assistance system can try to detect when a student is disengaging from a pedagogical activity, using pupillary response and other sensor information. Because this involves a complex technical setup, they have not found their way into everyday use. However, more static user characteristics, such as their willingness to seek help and their self-efficacy, could be a fruitful avenue to explore for determining the best time to intervene. In this study, we aimed to investigate such user characteristics that could lend themselves to personalize assistance intervention.

3. STUDY SETUP
We conducted an empirical study with a prototype assistance system for using photo-editing software to complete a task, in which we manipulated how long the system waited before intervening with a help message to complete the next task step.

We recruited 144 students and staff from a French university from a range of disciplines and backgrounds with an average age of 27.8 years. Each participant was randomly assigned to a group of 12; each group completed the same task but we varied the time the system waited for the next correct step in the task. We varied timing between groups in 3-second steps, for example, group A's assistance "fired" immediately whereas the system gave participants 21 seconds in group H to get back on task.

For their task, we asked the participants to create a holiday card from a photo they were given using PhotoScape (http://www.photoscape.org/ps/main/index.php). The task participants were asked to do consisted of: opening a given photo in edit mode; cropping the photo; adding a speech balloon; adding a frame; saving it. For each of these steps, the participant had to carry out several actions with PhotoScape. For instance, for the step “add a speech balloon”, the participants had to open the "Object" tab, draw a shape, open the shape properties, enter a text, set the font “Verdana” with size 24 points and color blue, set the balloon shape and then save the properties.

The assistance system we used, SEPIA [8][9], can monitor an application and "trace" all user interactions with this application, e.g. clicking on a button or opening a menu, without any need to access or modify the application source code. SEPIA can be leveraged to provide contextualized assistance, and user interface enhancements and automated actions can be injected into the application. To do so, we defined a "trace" which comprised a set of low-level actions that led to successfully completing the task. Then, using this trace, we specified *how* to assist for any low-level actions in the trace. In our study, assistance included simple help messages explaining to the user what to do next, coupled with a UI enhancement to an object on which the user should act on e.g. an arrow pointing to the button on which to click next (Figure 1). The assistance was non-modal; the user can close the pop-up or ignore it. We also specified *when* to assist by determining the maximum amount of time that the user could spend completing a step before assistance is triggered, meaning that the system can delay triggering the assistance once it has determined deviation from a trace. During completing the task using PhotoScape, a user's interaction with the application was monitored by SEPIA and provided contextualized assistance based on the triggers specified in the assistance specification.

Each participant started the session by filling in a background questionnaire, including details about their demographics, personality and help preferences that might be useful as variables in predicting the right timing of assistance, based on previous research into help-seeking and intelligent tutoring systems. We developed a set of questions that asked for participants' gender, age and previous experience in photo-editing, self-efficacy in completing a computer photo-editing task [5], and their self-esteem [13]. We also developed a set of questions that probed their help-seeking behavior, based on factors identified by [1][12], such as locus of control, need for achievement, authoritarianism, mastery and patience. We also asked participants to rate their perseverance when faced with a difficulty in the use of software and wished assistance frequency.

They then completed the task using PhotoScape; no tutorial was given how to use this application to succeed. During the use of PhotoScape, all the participants' actions were traced, as well as the assistance actions. Thus, we were able to determine what the participants did in PhotoScape, when and how often the assistance system provided help and for which subtask(s). We also measured how long participants waited after the system provided assistance before carrying out the next step in the task, i.e. following the advice.

Finally, we administered an exit questionnaire capturing the participants' feedback regarding the assistance provided. The aim of this questionnaire was to measure the participant’s satisfaction with the assistance provided by the system. We measured perceived timeliness (1– far too slow, 5 – far too quick) as a measure of getting the timing of the advice right.

4. RESULTS

To investigate how to provide the right advice at the right time, we analyzed participants' traces and questionnaire data. We excluded 3 participants from our analysis because they never received any assistance.

4.1 Did participants follow the assistance and how quickly?

We first analysed the trace logs to establish whether participants followed assistance given, no matter when we intervened. On average, all participants followed the assistance provided in less than 2 low-level actions and within 20 seconds. Overall, 64% of all instances when assistance was provided were followed *immediately*, i.e. the user did 0 low-level actions before carrying out the suggested action (Figure 2). For participants that had very large number of low-level actions before following the assistance, we noted from their traces that it was because they deviated from the task instructions we had given them, for example, a participant added a black and white effect to the photo.

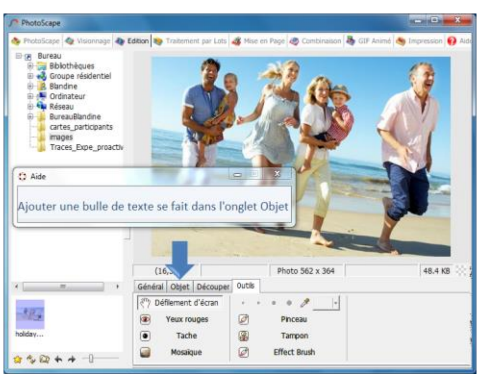

Figure 1. Example of assistance actions: help message and an arrow pointing to the next user action required.

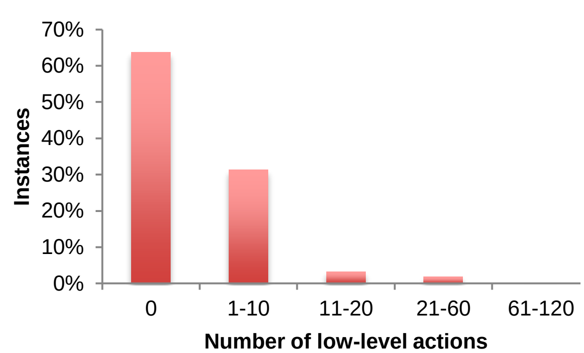

Figure 2. Frequency of number of low-level actions before advice is followed.

It might be assumed that the longer the system delayed the assistance, the less useful our advice was. We indeed found that participants in group A with an assistance timing of 0 seconds tended to follow the advice more quickly (mean=1.09 low-level actions), whereas the number of low-level actions participants carried out after advice was received increased as assistance was delayed (max mean=3.59 in 24 seconds timing interval). However, there was no difference in the number of low-level actions before advice was followed whatever the timing interval that participants experienced (F=1.145, p=0.335). Instead, it seems that there are individual user differences at play that affect how quickly participants react to assistance.

4.2 Towards Predicting The Right Time to Assist

We asked participants how they felt about the timing of the advice through the perceived timeliness ratings in the exit questionnaire, based on a 5-point Likert scale (recall that 1 means "far too slow", and 5 means "far too quick"). Using an ANOVA, we found that there was no significant difference between ratings based on the timing of when interventions were made (F=1.106, *p*=0.363). However, perceived timeliness decreased as actual timing increased (r=-0.205, *p*=0.015). This correlation is statistically significant but only weak, indicating that there also appeared to be other factors that played a role in participants' perceived timeliness.

It has been surmised that background factors play an important role in whether help is sought [1] and we therefore investigated the impact of participants' background factors on perceived timeliness. Using a multiple regression, we found that the regression model significantly predicted perceived timeliness (r=0.454, *p*=0.0003) and that there were three important factors that mattered in the prediction (Table 1). First, *assistance timing* had a negative impact on the perceived timeliness rating (B=-0.027) i.e. the slower advice was given after the participant deviated from the task, the slower they also perceived it. However, the coefficient shows the contribution of timing is quite low. Second, their *previous experience* with carrying out the task mattered and also had a negative impact on perceived timeliness ratings (B=-0.164). Hence, the higher their self-assessed expertise rating, the lower the perceived timeliness. This implies that the more they knew about photo-editing previously, the slower they perceived the assistance to be given. Third, the *amount of help they wanted* appeared to matter, again in a negative relationship (B=-0.386): the higher their rating on required assistance the lower the rating on perceived timeliness. This means that the more they wanted help, the slower they perceived help to arrive. Last, it should be noted that there is a very important "anchor" from which participants seemed to judge the timeliness of advice. This baseline is represented by the constant (B=4.524), sitting very close to the extreme end of the 5-point Likert scale, meaning that participants started out as perceiving advice given as "too quick" in most cases, and then decreased their ratings based on other factors. Indeed, 53 out of 141 (38%) participants rated the timing of the advice as "too quick", whereas 54 rated it as "right" (38%), and only 34 (24%) as "too slow".

	Coefficients (B)	*p*
Constant	4.524	0.000
Assistance timing (sec)	-0.027	0.003
Age	-0.002	0.800
Gender	0.078	0.688
Expertise in photo editing	-0.164	0.045
Self efficacy	0.008	0.199
Self esteem	0.003	0.672
Help-seeking	-0.012	0.187
Perseverance	0.142	0.106
Wished assistance frequency	-0.386	0.001

Table 1. Factors in timeliness regression model (shaded shows significant)

Attribute	Cluster#				
	0	1	2	3	4
N	15	38	31	23	34
Assistance timing (sec)	9.6	25.1842	3.0968	11.8696	18.0882
Expertise in photo editing	3.4	2.2368	1.5161	1.913	1.2941
Wished assistance frequency	2.0667	3.0526	3.2258	2.6522	2.5
Timeliness	4	2.4211	3.2581	2.4348	4.2647

Table 2. Cluster centroid information

A common approach in personalizing interfaces based on user characteristics is to "stereotype" users and then to determine the behavior of the systems by how well a new user fits this stereotype. One way this could be done is by dividing all user data into clusters (i.e. the "stereotypes"), assign a specific user to a cluster based on a distance metric and then change the assistance timing in some way that makes sense for the stereotype. We decided to use the three factors described in the regression model in a cluster analysis to investigate what separates different groups of users.

We produced five clusters over the data set containing 141 participants, giving us reasonably distributed and separated data (Table 2). We can identify three different approaches to timing advice: increase the time after which advice is given, speed up giving advice, or instances when the timing was about right. Cluster 0 (N=15) and cluster 4 (N=34) both contained participants who rated the advice as being given "too quickly" (4 or 4.2647, respectively, with an assistance timing of 9.6 seconds and 11.87 seconds, respectively). What these two groups of users share is a pre-established desire for less frequent interventions and hence it might be an idea to increase the timing for users similar to these participants. Cluster 1 (N=38) and cluster 3 (N=23) show participants who considered timeliness too slow and therefore we could decrease the assistance timing. Cluster 1 contained participants who requested a moderate amount of assistance, were not greatly experienced with photo editing and had a long timing interval of more than 25 seconds. In contrast, cluster 3's participants were equally not very experienced with photo editing but wanted less assistance and had a shorter timing interval of advice of about 11 seconds. In contrast, cluster 2 (N=31) seems to contain most of the participants who judged the advice as coming at the right time. These individuals were not very experienced in photo editing but wanted more help. In this case they experienced assistance timing of about 3 seconds.

5. DISCUSSION

Even though predicting the best timing of advice is difficult, our results are a first step towards doing so. Most participants appreciated the advice our system gave: 65% of the participants said that the advice was effective, 77% of them found it relevant or very relevant and 81% stated that it helped them to achieve the task more quickly. We also found that 87% of participants appreciated the way we enhanced the UI with the advice.

Our study can also point the way to the next steps in research in this area. First, we did not focus on task difficulty in our study design and some of the steps in our task instruction were very complex to do. Further analysis could illuminate whether the given assistance was more useful for these steps. Second, we only captured an overall, aggregated measure of the timeliness of assistance from participants instead of feedback about each intervention. For example, participant A1 saw 14 interventions but possibly some of the advice given was well-timed whereas others might have been offered too quickly or not quickly enough. Furthermore, our data was not ideally balanced; only 38% of participants rated the timing of the advice "right", whereas the majority thought the advice did not arrive at the right time. This means that there is a substantial amount of "noise" in our data which makes prediction and modeling difficult. Last, identification of the user's task is currently quite basic, based on a deviation from an expected task path which has to be demonstrated by the assistance developer. However, if the current task along with an expected sequence could be predicted from users' actions, assistance could be made more accurate. Of course, gathering enough examples for task prediction might be challenging in this context but possibly previous work in detecting frequent procedures could be useful in these circumstances [14].

6. CONCLUSION

We conducted an empirical study to investigate how best to time assistance to users, capturing feedback by participants through logged interactions and their subjective ratings. We showed that a user's perceived expertise in a task, their wished assistance frequency and the timing of advice are important factors to consider in personalizing assistance systems to an individual user. The right time to intervene is difficult to predict accurately, however, our work has shown some early indications how to adjust assistance timing based on a user's characteristics and preferences, in order to provide the right assistance at the right time.

REFERENCES

[1] Aleven, V., Stahl, E., Schworm, S., Fischer, F. and Wallace, R. (2003). Help Seeking and Help Design in Interactive Learning Environments. Review of Educational Research, 73(3), 277-320.

[2] Beck, J. (2005). Engagement Tracing: Using Response Times to Model Student Disengagement. Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology, Looi C. et al., eds., IOS Press, 88-95.

[3] Bull, S. and Kay, J. (2010). Open Learner Models. Advances in Intelligent Tutoring Systems, R. Nkambou, J. Bourdeau, R. Mizoguchi (Eds.), Springer Berlin Heidelberg, 301-322.

[4] Cocea, M. and Weibelzahl, S. (2011). Disengagement detection in online learning: validation studies and perspectives. Learning Technologies, 4(2), 114–124.

[5] Compeau, D. R. and Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. MIS Quarterly, 189–211.

[6] Dragunov, A. N., Dietterich, T. G., Johnsrude, K., McLaughlin, M., Li, L. and Herlocker, J. L. (2005, January). TaskTracer: a desktop environment to support multi-tasking knowledge workers. Intelligent User Interfaces, ACM, 75-82.

[7] Dufresne, A. and Paquette, G. (2000). ExploraGraph: A Flexible and Adaptive Interface to Support Distance Learning. Ed-Media, Victoria, Canada, 304–309.

[8] Ginon, B. Thai, L. V., Jean-Daubias, S., Lefevre, M. and Champin, P.-A. (2014). Adding epiphytic assistance systems in learning applications using the SEPIA system. Ec-Tel, Graz, Austria, 138-151.

[9] Ginon, B., Champin, P. A. and Jean-Daubias, S. (2013). An approach for collecting fine-grained use traces in any application without modifying it. Workshop EXPPORT of ICCBR, New-York, USA.

[10] Iqbal, S. T. and Horvitz, E. (2010). Notifications and awareness: a field study of alert usage and preferences. Computer supported cooperative work, 27-30.

[11] Mille, A., Champin, P. A., Cordier, A., Georgeon, O. and Lefevre, M. (2013). Trace-based reasoning-modeling interaction traces for reasoning on experiences. FLAIRS Conference, 1-15.

[12] Nadler, A. (1997). Personality and Help Seeking. Sourcebook of Social Support and Personality, G. R. Pierce, B. Lakey, I. G. Sarason and B. R. Sarason (Eds.), Springer US, 379-407.

[13] Rosenberg, M. (1965). Society and the adolescent self-image. Princeton, NJ: Princeton

[14] Shen, J., Fitzhenry, E., Dietterich, T. (2009). Discovering Frequent Work Procedures From Resource Connections. Intelligent User Interfaces, ACM.

[15] Swartz L. (2003). Why People Hate the Paperclip: Labels, Appearance, Behavior, and Social Responses to User Interface Agents, Master's report, Stanford University, Stanford, USA.

[16] Zarka, R., Champin, P-A., Cordier, A., Egyed-Zsigmond, E., Lamontagne, L. and Mille, A. (2012). TStore: A Web-Based System for Managing, Transforming and Reusing Traces. ICCBR 2012 TRUE and Story Cases Workshop, Luc Lamontagne, Juan A. Recio-Garc ed. Lyon, France, 173-182.