

The Diary Study: A Workplace-Oriented Research Tool to Guide Laboratory Efforts

John Rieman

Department of Computer Science and Institute of Cognitive Science
University of Colorado
Boulder, CO 80309-0430
(303) 786-0447
rieman@cs.colorado.edu

ABSTRACT

Methods for studying user behavior in HCI can be informally divided into two approaches: experimental psychology in the laboratory and observations in the workplace. The first approach has been faulted for providing results that have little effect on system usability, while the second can often be accused of yielding primarily anecdotal data that do not support general conclusions. This paper describes two similar approaches in another field, the study of animal behavior, and shows how they produce complementary results. To support similar complementary interactions between research approaches in the HCI field, the paper describes the diary study technique, a tool for research in the workplace that achieves a relatively high standard of objectivity. A diary study is reported that focuses on exploratory learning.

KEYWORDS: Diary studies, methodologies, participatory design, situated cognition, exploratory learning

METHODS FOR STUDYING USER BEHAVIOR

In the field of human-computer interaction (HCI), methods for studying user behavior can be informally divided into two approaches: experimental psychology in the laboratory and observations of users in the workplace.

The laboratory-based approach of experimental psychology grows out of a long history of work in the human factors field. In this style of research, investigators bring users into a laboratory situation and make quantitative observations of their behavior. The users are chosen from a single population, and the laboratory situation is carefully designed to vary only a few independent variables. The situation may have little in common with real computer usage. In the conservative version of this approach, a large number of subjects are tested, yielding data that can be interpreted with statistical methods and projected to a larger population [10]. A related method uses protocols taken from a few subjects and analyzed in depth, as in thinking-aloud studies or individual user modelling efforts [11,13].

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

© 1993 ACM 0-89791-575-5/93/0004/0321...\$1.50

The second dominant approach is to collect data in the workplace, with the researcher often interacting closely with users in an informal way. The most radical version of this approach is participatory design, often identified with the European and particularly the Scandinavian HCI community [5]. In participatory design, users and designers work together to identify system goals and to iteratively produce systems that meet the users' evolving needs. Some other workplace-oriented methodologies include user surveys, monitoring the use of prototypes, and contextual design [18]. Although it is not always the case, investigations in the workplace are generally less controlled than laboratory efforts, involving fewer subjects and relying more heavily on subjective judgments and personal interactions. These efforts are most often described as design methodologies, not research tools, and the data they yield tend to be less predictive, more anecdotal, more limited to the individual situation under investigation.

Most HCI researchers are aware of this distinction, and many design projects use a mixture of the two approaches [6]. Nonetheless, the polarity is informative. On the one hand, the laboratory approach typically yields objective data that can be interpreted statistically to predict the behavior of a population, but the data concern the behavior of subjects in a contrived situation. Critics of this work, including proponents of the situated cognition view, argue that it has little impact on the usability of real systems [14,18]. By contrast, information gathered in the workplace as part of a design project reflects the real, day-to-day work of the participants, but extracting objective, general data from these reports may be difficult.

This paper offers some insights into the interaction between these two general approaches, based on the interaction of similar approaches in another research field, animal behavior. The paper then describes a method, the diary study, which is intended for use in the workplace, but which yields data that are more objective and more generalizable than many workplace-oriented methods. As an example, a diary study of exploratory learning is described.

LESSONS FROM THE STUDY OF ANIMAL BEHAVIOR

The study of animal behavior is much older than HCI, with the roots of its modern theories going back at least as far as Pavlov and Darwin in the nineteenth century. But as with HCI, there are fundamentally two research approaches. The

first of these is to study animals in the laboratory, an approach usually associated with experimental psychology. The second is to study animals in natural field situations, an approach taken by ethologists.

Like experimental psychology in HCI, experimental psychology with animals often involves contrived situations that are far from natural: rats are encouraged to visit the arms of radial mazes where cheese appears and disappears as if by magic; pigeons learn to dance from microswitch to microswitch in response to colored lights or photographs; cats are forced to escape from boxes with trick latching devices. Subject animals are carefully chosen from an artificially bred population, and groups for a given experimental condition may be surgically disabled, drugged, or raised in exceptional environments. But the data produced are clearly objective.

By contrast, ethological investigations are typically performed in the field, with randomly selected members of the naturally occurring population. Much of the work is simple visual observation of natural behavior. Other research relies on artificial devices, such as radio collars to track animal movement patterns, and some projects involve experimental manipulations, such as substitution of differently colored eggs in field sites. Ethologists strive to be objective, but the research is unavoidably influenced by behavioral coding decisions, and reported data are often based on small samples of a population. As with workplace HCI, some of this research may involve participation and highly subjective judgments: Konrad Lorenz with geese, Diane Fossey with gorillas. Interestingly, this approach, like workplace HCI, also has a history that is more closely associated with the European than the American research community.

Reconciling the Two Approaches

As with the two polar approaches to HCI research, there is clearly a tension between the two approaches to the study of animal behavior. However, several researchers have recently suggested that the two approaches can be complementary, in both a theoretical and a practical sense [1,4,12].

In a theoretical sense, the approaches can be seen to ask different, but complementary, questions. Ethological research in the field asks *functional* questions: how is the survivability of a species enhanced by a given behavioral characteristic, and why did that characteristic evolve? [7] To provide convincing answers to these questions necessarily requires that as much as possible of the organism's natural environment be considered in the analysis. On the other hand, experimental psychology in the laboratory asks *causal* (proximate) questions: what combination of cues does an organism perceive that causes it to behave in a certain way, and what cognitive algorithms does it use to define its behavior? The theoretical interplay is made especially interesting by the requirement that behavior which might be functionally optimal must be approximated by algorithms that an organism can process with its available memory and cognitive power [4].

It is this interaction between theoretical questions that also sets up the practical interaction between the two research approaches. The data and theories of ethologists suggest functionally optimal behavior; experimental psychologists can then design experiments to uncover mechanisms that implement or approximate this behavior.

Clark's Nutcracker: A Suite of Studies

As an example of the interaction between field and laboratory research in animal behavior, consider a series of studies investigating the behavior of Clark's Nutcracker, a bird found in the conifer forests in the western part of North America. Observations of naturalists have revealed that the seeds of pine trees form a major part of this species' diet. During the seed harvesting season, which lasts about one month, nutcrackers remove the seeds from pine cones and bury them in shallow holes, called caches, for retrieval during the rest of the year [15].

The relationship between the nutcracker and the pine trees is a close one. The beak of the nutcracker is specialized for removing seeds from cones and carrying them to cache sites or the nest, and the nutcracker's range closely parallels the range of the pine trees that provide its food. The range of the forest, in turn, is influenced by the nutcracker's activities: not all cached seeds are recovered, and unrecovered seeds may germinate and grow into new trees.

Field studies in the ethological style investigated this relationship between bird and forest, with particular attention to the nutcracker's effectiveness in caching and recovery [15,17]. Using a combination of energy calculations, field examinations of recovered caches, and direct observation, these researchers concluded that an individual bird must store on the order of 32,000 seeds each year, in roughly 8,000 caches. Although some of these caches are stolen by rodents, and others are unrecovered and left to germinate, an estimated 60 to 86 percent of the caches are located and recovered by nutcrackers.

These results of this investigation into functional issues raised causal questions that field research alone could not easily answer. Specifically, what were the mechanisms that nutcrackers used to achieve this high recovery rate? Was there some common pattern of storage and search to which all nutcrackers were privy, making the caches a communal resource? Or, did the birds actually possess the spatial memory needed to recall and recover their own caches? If so, did they rely on local, visual cues, such as the shape of a rocks and trees near the cache, or did they navigate using some more sophisticated system?

To answer these questions, researchers transferred the investigation from the field to the laboratory [16,9]. A large outdoor flight cage was constructed, with a dirt floor and various objects, such as logs and rocks, to simulate a forest setting. In this setting, a series of controlled experiments eliminated olfactory and other possible cues that could not be controlled in the field. The results confirmed that a nutcracker's primary means of cache relocation was, indeed, spatial memory of the locations in which it had stored its own seeds, cued by large objects near to each cache.

Applying Similar Principles to HCI Research

The time-course of the nutcracker investigations illustrates an effective research paradigm. First, informal observations of naturalists identified the basic relationship between the nutcracker and the pine forest. Second, formal fieldwork investigated the functional relationships, raising questions as to the causal mechanisms involved. Third, laboratory experiments, designed to mimic important features of the natural environment, probed those mechanisms.

A similar complementary research approach can be effective in HCI research. Anecdotal evidence and insights acquired during participatory activities with users can suggest questions for more objective field studies in the workplace. These studies, in turn, can point to questions for investigation in the laboratory. They can also be used to suggest laboratory situations that reflect the critical constraints of the user's real environment.

An essential component of this approach is a set of objective tools for workplace investigations during the second phase of the research, along with guidelines for interpreting the data they yield. While these tools should reflect participatory design's fundamental concern with the rights and interests of the user, they should be structured primarily for objective data collection. One such tool is the diary study.

THE DIARY STUDY TECHNIQUE

The diary study, as described here, is based on a technique used by Ericsson and his colleagues to investigate expert behavior of concert musicians [3]. In a diary study, participants are asked to record their daily activities on a preprinted log form. The log breaks the day into brief intervals, typically 15 or 30 minutes. The activities are initially recorded using the participant's own description, i.e., "had dinner with Kelly," or, "dentist — ouch!" Later, all activities are assigned to predefined categories, such as "meals" or "physical exercise." Diaries are typically kept for one or two weeks; any longer would be a burden for many participants. The diary record itself is usually supplemented with a post-study interview of the participant, addressing behavior over a longer term than the diary covers.

The activity categories in the daily log can be based on established categories for time-budgeting activities [8]. However, the log can be tailored to the questions of interest in the study by adding to or specializing these categories. The size of the time intervals and the hours of each day covered by the log can also be adjusted.

Data Provided by the Diary Study

The diary study provides several kinds of data. Most powerfully, the record of time spent in each activity category is quantitative data, amenable to statistical analysis. Ericsson and his colleagues used this data to correlate practice times with levels of expertise in concert pianists and violinists [3]. Because the diary participant is forced to record *all* activities for the period covered, the data reported in the diary are arguably more reliable than reports of time spent, if any, on one specific activity each day.

The second kind of data provided by the study are the statements collected in the interview. Although interview data standing alone are a relatively soft and unreliable, the participant's statements are "hardened" by relating them to the diary record. For example, if an interviewee says, "No, I hardly ever use my spreadsheet program," and a two-week diary confirms this, the evidence is much stronger than an interview statement without further support. The diary record can also suggest questions for the interview.

Finally, the diary study provides an opportunity for the researcher to interact with a number of users in their workplaces. The informal insights gained through this interaction are similar in kind to those achieved in less formal approaches, such as participatory design. This information is useful in suggesting appropriate tasks for controlled investigation in the laboratory.

EXAMPLE: A DIARY STUDY OF EXPLORATORY LEARNING

An ongoing project of our research group is the study of how users learn the functionality of computer systems during their daily work, outside a formal training situation. We refer to this activity as "exploratory learning" [2]. A subsidiary question is how users explore a system in the absence of a given task, simply out of curiosity. In preparation for laboratory studies of this behavior, we performed a diary study to find out how prevalent task-free exploration is and where it occurs.

The diary study involved four data-collection activities: (1) a daily log, (2) "Eureka Reports" of learning incidents, (3) daily debriefing interviews, and (4) a longer interview following the one-week diary period. These activities were all designed to focus on computer-related learning, while avoiding a bias in favor of any one learning technique.

The time during which the daily log was kept was limited to the participant's working hours. We had no interest in what a participant did during his or her time away from the job, and minimizing invasion of privacy helped the researcher achieve a good working relationship with the participants. As shown in Figure 1, the log was focussed on the area of interest by including categories for several computer-related activities, rather than a single one for "computing." The log did not include a category for "learning," an activity orthogonal to the categories listed. Learning episodes were identified during the daily debriefing and through the Eureka Reports (Figure 2). A Eureka Report was to be filled out whenever a participant realized he or she had learned something, or solved a problem, or attempted unsuccessfully to learn or problem-solve.

The daily debriefing interview was a critical part of the study. During the day, the participant filled in the activities column of the log, describing activities in half-hour intervals. At the end of each day, or early the next day, the researcher met with the participant and assigned the activities to categories. The discussion accompanying this interaction gave the researcher an opportunity to discover any learning episodes that had taken place and to ensure that they had been recorded on Eureka Reports.

Categories: Fill in at End of Day											
Day: <u>TUES-9/15</u>	Talk, In Person	Talk, Phone	Meetings	File, Organize	Fill in Forms	Copying	Paper Mail	E-Mail	Word Process	Spreadsheet	Other Compute
I.D.: <u>7</u>											Breaks/Personal
8-8:30	Got coffee Checked e-Mail										Reading
8:30-9	Phoned garage about car More e-Mail	*						*			
9-9:30	Met with student	*									
9:30-10	AI Class										*

Figure 1. The beginning of a diary log sheet for one day. The participant records activities on the left as the day proceeds. The researcher assigns categories during the end-of-day debriefing.

"Eureka" Report	
For Computers, Phones, Copiers, Fax Machines, Staplers, Clocks, Thermostats, Window Locks, Cameras, Recorders, Adjustable Chairs, and other Strange Devices.	
Name: <u>7</u>	Date & Time: <u>9/15 11am</u>
Describe the problem you solved, or the new feature you discovered, or what you figured out how to do:	
<u>Got copier to put staple in right corner!</u>	
How did you figure it out? (Check one or more, explain)	
<input type="checkbox"/> Read the paper manual <input type="checkbox"/> Used on-line "Help" or "Man" <input checked="" type="checkbox"/> Tried different things until it worked <input type="checkbox"/> Stumbled onto it by accident <input type="checkbox"/> Asked someone (in person or by phone) <input type="checkbox"/> Sent e-mail or posted news request for help <input type="checkbox"/> Noticed someone else doing it <input type="checkbox"/> Other	
Explain: <u>Can't read "international" copier symbols</u>	

Figure 2. A "Eureka Report," for noting episodes in which new information is sought or acquired.

The one-week diary study of each participant was concluded with a one-hour structured interview, covering the participant's experiences in learning about computers and other devices outside the time span of the diary. After a week of daily debriefings, the researcher and participant had established a rapport and common vocabulary that significantly enhanced the final interview.

The study provided converging evidence of varying quality. We did not simply ask the participants, "Tell us when and how you learn something." Rather, we asked for a complete report of their days, and we discussed each day with them to discover learning episodes. The Eureka Reports served to further focus on this issue, and interview questions allowed us to identify behavior that the subject considered uncharacteristic, as well as to gain an understanding of reasons for the activity patterns reported. This approach yielded evidence that was stronger than isolated anecdotal reports, while avoiding the biasing influence of researchers who actually participate in the users' work.

Results of the Study and Their Significance for Future Investigations

Ten participants completed five-day diaries (one work week) for the exploratory learning study. Participants were selected from a variety of backgrounds to get an overview of how and when computer learning activities occurred. Two of the subjects were secretarial personnel, two were undergraduates with part-time clerical assistant jobs, two were university faculty members (one in computer science), and four were full-time researchers in computer science. The last group was where we expected to find the most frequent occurrence of task-free exploratory learning.

As described, the sampling was too small and inhomogeneous to support strong projections to a larger population. However, the Eureka Reports, supported by the daily logs and the follow-up interviews, revealed several points that will guide the next phase of our research.

Based on the number of Eureka Reports generated, the participants could clearly be divided into two groups: explorers and non-explorers. The explorers included only two of the participants, one in the field of computer science and one outside that field. These users actively sought new information about their computer systems almost every day. Each of the explorers generated 18 Eurekas during the five-day diary study. The remaining eight participants, the non-explorers, generated an average of 4 Eurekas over the five days, with a minimum of 2 and a maximum of 8. Two of the four subjects with the lowest Eureka counts (2 or 3) were computer scientists.

The explorer who was not a computer scientist was the only participant to engage in a significant amount of learning that was not directly demanded by work-related tasks. As we had hypothesized, these activities were structured around tasks created primarily for the purpose of investigating the computer system.

The categories on Eureka Reports describe how participants made their discoveries. For the total of 69 Eurekas reported

by all participants, the three most common ways of learning were: trying things until something worked (42 times); reading the manual or using on-line help (41 times); and asking for help, either in person or occasionally by e-mail (18 times). There was often more than one method noted for a single Eureka, and a common combination was reading the manual along with trying things. Further insight is yielded by the distribution of these methods among the explorers and the non-explorers. Explorers were about twice as likely to use the manual (30 times in 36 Eurekas) as non-explorers (11 times in 33 Eurekas). They were also more likely to try things (26 times for explorers, 16 times for non-explorers), but slightly less likely to ask for help (6 times for explorers, 10 times for non-explorers).

These results provide guidance for our continuing research into exploratory learning. We can be alert to potentially large differences in behavior, perhaps bimodally distributed, and not tied to any obvious characteristics. And contrary to our original hypothesis, we have evidence that task-free exploration is not limited to computer scientists.

We can also design laboratory situations to investigate exploratory learning under real-world task conditions. Our diary participants showed a strong reliance on written documentation, often as the first step in investigating a system feature. Asking subjects to begin exploration in the laboratory without access to documentation would not reflect a natural situation. Further, the diaries and the follow-up interviews both revealed the frequency with which users ask for help. Again, the natural lab situation should somehow support this fall-back action, instead of forcing subjects into extended trial-and-error behavior.

Methodological Lessons From the Study

Experience in setting up and running this study has taught us several lessons about the diary technique. First, personal interactions are a key part of a successful study. Participants must be convinced to make a considerable effort over a week's time to record their activities. Several participants agreed to maintain a diary but initially failed to record sufficient detail. One participant filled in the log sheet for the first day with a single word: "programming." Another subject waited until the end of the day, then wrote down the entire day's activities, noting that this was the way he had kept log sheets in a previous job. After some discussion, the researcher convinced both participants to maintain their logs with greater detail, on an ongoing basis.

A second lesson was that participants inevitably tried to focus their logs, and even adjust their activities, to emphasize areas that they thought would interest the researcher. The participant who had written "programming" for the day's activities explained, "I didn't do anything you were interested in." More than one participant apologized to the researcher at the end of the day for not having learned anything new, and participants sometimes made comments such as, "I guess I should spend some time learning about this new program so I can fill out a Eureka Report."

Avoiding this bias is difficult. Our participants had several clues as to the study's purpose: the Eureka and log forms,

the legal release, the end-of-day debriefings, and sometimes prior knowledge. As noted, the current study broadened the focus by addressing learning in general, rather than highlighting task-free exploratory learning. In addition, the researcher made a conscious effort to impress upon participants the need to record their normal activities, not activities that they thought would be interesting.

A third lesson is that the process is quite effortful. Even though the debriefings only take about 15 minutes each day, those periods must be scheduled at the participant's convenience, and the researcher must arrive at the participant's workplace on time, prepared to wait until the participant finishes his or her current task. Logging the activities of more than two participants in a week is usually impossible for a single researcher.

Finally, because it takes a day or two for the participant to become comfortable with the process, it is a good idea to start participants on different days of the week. Otherwise, the diaries may all have questionable data for the weekday, typically Monday, on which all logs are started. Starting midweek also allows activities performed during the weekend to be captured as an integral part of the study.

SUMMARY

The two fundamental methods of HCI research, observations in the workplace and controlled laboratory studies, can complement each other in an effective research program. The diary study is a tool that can help bridge the gap between these methodologies, providing objective data about the workplace to guide further study in the laboratory.

ACKNOWLEDGEMENTS

Discussions with Anders Ericsson were instrumental in defining the diary study methodology and in clarifying the larger methodological issues. Thanks is extended to Evelyn Ferstl, Marita Franzke, Alcinda Lewis, Clayton Lewis, Peter Polson, and the CHI'92 Doctoral Consortium for their comments. Support for this research has been provided by the National Science Foundation, grants IRI 8722792 and 9116640, and by US West Advanced Technologies.

REFERENCES

1. Brain, P.F. Ethology and experimental psychology: from confrontation to partnership. In *Ethoexperimental Approaches to the Study of Behavior*, R.J. Blanchard, P.F. Brain, D.C. Blanchard, and S. Parmigiani, Eds., Kluwer, Dordrecht, The Netherlands, 1989, pp. 19-27.
2. Carroll, J.M., Mack, R.L., Lewis, C.H., Grischkowsky, N.L., and Robertson, S.P. Exploring a word processor. *Human Computer Interaction 1* (1985), 283-307.
3. Ericsson, K. Anders, Tesch-Römer, C., and Krampe, R.T. The role of practice and motivation in the acquisition of expert-level performance in real life: An empirical evaluation of a theoretical framework. In *Encouraging the Development of Exceptional Skills and Talents*, M.J.A. Howe, Ed., The British Psychological Society, Leicester, 1990, pp. 109-130.
4. Fantino, E., and Abarca, N. Choice, optimal foraging, and the delay-reduction hypothesis. *Behavioral and Brain Sciences 8* (1985), 315-330.
5. Floyd, C., Wolf-Michael, M, Reisin, F.-M., Schmidt, G., and Wolf, G. Out of Scandinavia: Alternative approaches to software design and system development. *Human-Computer Interaction 4* (1989), 253-350.
6. Gould, J.D., Boies, S., and Lewis, C. Making usable, useful, productivity-enhancing computer applications. *Communications of the ACM 34* (1991), 74-89.
7. Houston, A., Kacelnik, A., and McNamara, J. Some learning rules for acquiring information. In *Functional Ontogeny*, David McFarland, Ed., Pitman, London, 1982, pp. 140-190.
8. Juster, F.T., and Stafford, F.P. (Eds.) *Time, Goods and Well-being*. Institute for Social Research, The University of Michigan, Ann Arbor, Mich., 1985.
9. Kamil, A.C., and Balda, R.P. Cache recovery and spatial memory in Clark's Nutcrackers. *J. Exp. Psych., Animal Behavior Processes 11* (1985), 95-111.
10. Landauer, T.K., Research methods in human computer interaction. In *Handbook of Human-Computer Interaction*, M. Helander, Ed., Elsevier Science, Amsterdam, 1988, pp. 905-928.
11. Lewis, C. Using thinking aloud protocols to study the "cognitive interface." Research Report RC 9265, IBM Thomas J. Watson Research Center, Yorktown Heights, N.Y., 1982.
12. Miller, D. B. Methodological issues in the ecological study of learning. In *Issues in the Ecological Study of Learning*, T.D. Johnston and A.T. Pietrewicz, Eds., Lawrence Erlbaum, Hillsdale, N.J., 1985, pp. 73-95.
13. Shrager, J., and Klahr, D. Instructionless learning about a complex device: The paradigm and observations. *International Journal of Man-Machine Studies 25* (1986), 153-189.
14. Suchman, L. A. *Plans and Situated Actions*. Cambridge, England: Cambridge University Press, 1987.
15. Tombach, D.F. Foraging strategies of Clark's Nutcracker. *Living Bird 16* (1978), 123-161.
16. Vander Wall, S.B. An experimental analysis of cache recovery in Clark's Nutcracker. *Animal Behavior 30* (1982), 84-94.
17. Vander Wall, S.B., and Balda, R.P. Coadaptations of the Clark's Nutcracker and the piñon pine for efficient seed harvest and dispersal. *Ecological Monographs 47* (1977), 89-111.
18. Wixon, D., Holtzblatt, K., and Knox, S. Contextual design: An emergent view of system design. In *Proceedings of the Conference on Human Factors in Computing Systems* (Seattle, Wash., April 1-5, 1990). ACM, New York, 1990, pp. 331-336.