Are expert users always better searchers?

Interaction of expertise and semantic grouping in hypertext search tasks

Ladislao Salmerón

José J. Cañas

University of Granada, Spain

Inmaculada Fajardo

University of the Basque Country, Spain

MANUSCRIPT DRAFT- PLEASE CHECK THE ORIGINAL REFERENCE

Correspondence may be sent to:

Ladislao Salmerón

Department of Experimental Psychology
University of Granada
Campus de Cartuja
Granada, 18071, Spain
Tel. 34-958-240767
Fax 34-958-246239
e-mail: lalo@ugr.es

Abstract

The facilitative effect of expertise in hypertext information retrieval tasks (IR) has been widely reported in related literature. However, recent theories of human expertise question the robustness of this result, since previous works have not fully considered the interaction between user and system characteristics. In this study, the constraint attunement hypothesis (CAH) is considered in order to predict that the effect of expertise in IR would appear only when the user and system characteristics can be combined successfully. Results from an experiment revealed that expert users outperformed novice users in IR when the elements of a system interface are organized semantically, but not when organized randomly. Results are discussed in the framework of the CAH supporting the interactive nature of human behaviour in HCI.

Hypertext systems consist of a network of linked documents (texts, images, tables...) that users can access using different features like menus or embedded links. Due to its potential in organizing a great amount of information, one of the main tasks that users perform in this system is information retrieval (IR). Researchers have investigated several task and user factors that influence performance in this task (e.g. Chen and Rada 1996). Expertise on the information domain is one of the most important user factors to explain performance in IR: high knowledge users (i.e. experts) are found to be more efficient and accurate than low knowledge users (i.e. novices). This effect has been found repeatedly in the literature, in hypertexts with different content structure (Shin *et al.* 1994, McDonald and Stevenson 1998a, Patel *et al.* 1998), menu organization (Hollands and Merikle 1987), navigation tools (McDonald and Stevenson 1998b) or type of goal (Last *et al.* 2001). Researchers agree that this effect is due to the fact that experts hold an accurate mental representation of the contents hence allowing them to engage in a more efficient knowledge driven search (e.g. McDonald and Stevenson 1998b). One of the most important theories that have been proposed for explaining the mental representation that experts built while navigating through a hypertext is the cognitive flexibility theory (CFT) by Spiro and cols. (Spiro, Coulson, Feltovich and Anderson 1988; Spiro, Feltovitch, Jacobson and Coulson 1991). The theory affirms that a main feature of the expert mental representation is its flexibility. In this sense, the CFT proposes that the expert can adapt its representation of a domain in order to better understand an ill-structured medium like hypertexts. Applied to IR tasks, the CFT would affirm that experts would outperform novices even in the case in which the hypertext provide a poor defined structure of its content, because experts would adapt its flexible representation to match the organization of the system. In the present paper we question this theory of expertise and the robustness of the expertise effect in IR tasks. First we introduce an alternative account for the role of prior knowledge in IR that stresses the interactive (i.e. the human and the system) nature of this task. Second we test the hypothesis derived from this alternative approach in an experiment. Finally, we discuss the results in a framework supporting the interactive nature of human behaviour in human-computer interaction (HCI).

As stated before, a great amount of research supports the expertise effect in IR tasks. Furthermore, there is no direct evidence that contradicts this result. Why then questioning the robustness of this effect? A main reason for this challenge is that we have found that in other type of interactive tasks (complex problem solving) expert flexibility depends on constrains of the environment (Cañas, Quesada, Antolí and Fajardo 2003). However, in the literature of IR human (e.g. degree of expertise) and system characteristics (e.g. organization of the interface) have never been manipulated in conjunction. Consequently, we need to explore the interaction between expertise and environment constraints in IR in order to clarify the conditions under which the expertise effect appears. In order to do that, we start from considering Vicente and Wang (1998) constraint attunement hypothesis (CAH). This theory states that the effects of expertise on a given task do not always hold, because they depend on the level that the structure of the environment helps the person to use his expertise. In the words of Vicente and Wang (1998: 36): 'There can be expertise effects when there are goal-relevant constraints (i.e. relationships pertinent to the domain) that experts can exploit to structure the stimuli. The more constrain available, the greater the expertise advantage can be'. Following the CAH, we could expect that the effects of domain expertise on IR should not always be present, as previous works in the field suggest. They would depend, however, on the interaction between the mental representation of the user and the characteristics of the hypertext system relevant to the task (e.g. menus, interface or structure). Consider the case of a student that has had previous experience with class registration. As far as the registration system follows the task structure the person is used to, he should perform better than a freshman student that has not had previous experience in the domain. But in the case that the system violates some rule of the 'registration scheme' (e.g. by using a poor menu organization, like including the 'final exams schedule' under the label 'grading and enrolment information' instead of 'academic calendar') the effects of expertise on the domain could disappear. Some works assessing the effect of grouping system elements on IR partially support this hypothesis (Halgren and Cooke 1993, Niemela and Saariluoma 2003, Niemela *In press*). These works have shown that a semantic grouping of the elements of an interface improve performance (e.g. by using the Gestalt laws of similarity of closeness), but also, however, that this improvement disappears when labels are grouped randomly. Materials used in these experiments were relatively simple (e.g. semantic categories such as 'domestic animals' or 'sports'). Therefore, it can be assumed that most of the participants possessed a high knowledge of the information before interacting with the system. However, in these experiments the level of expertise of the user was not manipulated, so it can not be affirmed that these effects hold only for expert users, as the CAH would predict.

It is clear at this point that the CAH and the CFT predict different outcomes for expert interaction on IR tasks. On the one hand, the CAH proposes that experts would outperform novices only if the information of the interface is organized following an expert solution, but not otherwise. On the other hand, the CFT predicts that experts would outperform novices independently of the organization of the interface, because the flexibility of the expert mental representation will compensate for a random grouping of the information. For testing these predictions we run an experiment where participants with different domain expertise had to search for information in a hypertext system organized either according to an expert solution or randomly. However, it could be argued from the CFT that a randomly organized interface will be unexpected by an expert user, and thus he / she will need certain time interacting with the interface in order to reconfigure or adapt his / her flexible mental representation to that in which the information is presented. In order to control for that possibility, the search tasks were divided in three blocks, and in each block participants had to search for items that force them to visit all the main pages of the hypertext.

Experiment

Following the CAH, we predicted that users with some expertise on the topic of the hypertext will outperform novice users in terms of searching time, targets found and lostness, when the interface presented the information following an expert solution (we will refer to that as the semantic grouping), but not when presented in a random organization. In addition, we predicted that this effect will hold even when the users have gained some experience with the organization of the information on the hypertext (i.e search block 3).

Method

Participants

Thirty-six students from the University of Granada participated in the experiment. Twenty-two were undergraduates participating for class credit. Fourteen were graduate students participating voluntarily. None of them had had previous experience with the hypertext used in the experiment.

Materials

A web site of a scientific meeting was designed for this experiment. The site consisted of 11 pages hierarchically organized. The main page contained links to the rest of the pages. Each link was surrounded by a coloured rectangle (figure 1). Two menus were created: a semantically grouped menu and a randomly grouped menu. For each group of each condition the label colour was identical. Groups from the semantic menu were created according to an expert solution. Four experts in the domain with experience of scientific meetings performed a card sorting task with items from the main menu (Valero and Sanmartin 1999). The two grouping solution was chosen because it had the highest intra-expert agreement. Experts agreed in labelling the two groups as ‘Bureaucratic’ and ‘Content information’. The web site was written in HTML language, and the navigation behaviour along with the rest of the measures were controlled by a program written in Microsoft Visual Basic 6.0 running under PC computers.

-- Insert figure 1 about here --

Procedure

Participants performed a 30 items search task through the web site. For example, participants were prompted to search for 'Guidelines for contributions'. In order to find it, participants had to enter the menu 'Submissions' and click on the link 'Formatting suggestions'. The presentation of the items was divided into three blocks. In each block participants had to search for 10 items which were included in one of the 10 main pages. Therefore, in each block participants had to visit all the main pages. Participants had a limit of 60 seconds in order to find each item.

Design

We used a 2 x 2 x 3 Mixed Factorial design. Expertise (novice and expert users) and Menu (semantic and random) were the between-participants variables and Blocks (3 levels) was the within-participant variable. Participants' expertise was established by considering the number of times they had attended a scientific meeting. Participants bellow the median (2.5 times) were considered novices (M = 0.3, SD = 0.6), whereas participants above the median were included in the expert group (M = 8.2, SD = 8.9). Therefore, nine participants were assigned to each of four groups. Performance on the search task was measured by three dependent variables: the response time, lostness and number of targets found. Lostness was defined following the formula proposed by Smith (1996): $L = (N / S - 1)^2 + (R / N - 1)^2$, were N is the number of different nodes visited while performing the search task; S is the total number of nodes visited while performing the search task; and R is the number of nodes needed in order to accomplish the search task. Lostness index ranges between 0 and 1. The greater the value, the greater the lostness.

Results

Different performance data was submitted to an ANOVA with Expertise (novice and expert) and Menu (semantic and random) as between-participants variables and Blocks (3 levels) as within-participant variable. Analysis of the response time and lostness were conducted only on the trials where the participant had found the target. Results are summarized in table 1.

Response time. Results for response time showed a main effect of Block, $F(2, 64) = 12.43$, $MSe = 14.72$, $p < 0.001$. Participants were slower at block 1 ($M = 13.9$), than at block 2 ($M = 10.5$) and block 3 ($M = 9.7$). In addition, the main effect of Expertise was significant, $F(1, 32) = 5.38$, $MSe = 35.36$, $p < 0.05$. Experts were faster ($M = 12.75$) than novices ($M = 10.09$). However, the interpretation of this effect should wait until the significant interaction between Expertise and Menu is considered, $F(1, 32) = 4.37$, $MSe = 35.36$, $p < 0.05$. Supporting our hypothesis, simple effects analyses showed that experts were faster than novices in the semantic menu, $F(1, 32) = 9.72$, $p < 0.01$, but not in the random menu, $F < 1$. Therefore, the effect of expertise appeared only in the semantic menu. In the random menu expert performance dropped to the level of a novice user (see scores in table 1). To rule out the possibility that the expertise benefit will arise in both the semantic and random menus only after some practice has been gained with the system, a planned comparison was made considering only those scores from Block 3. Supporting predictions from the CAH theory, results revealed that experts continued being faster than novices in the semantic menu, $F(1, 32) = 4.89$, $MSe = 23.79$, $p < 0.05$, but not in the random menu, $F < 1$.

Lostness. Results from the lostness score showed no significant main effects. More important for our purposes, data revealed a close to significant interaction between Expertise and Menu, $F(1, 32) = 3.82$, $MSe = 0.04$, $p = 0.059$. Simple effects analyses were consistent with our predictions, showing that experts get lost fewer times than novices in the semantic menu, $F(1, 32) = 5.87$, $p < 0.05$, but not in the random menu, $F < 1$. Again, planned comparisons were made considering only data from block 3. As with the omnibus analysis, experts get lost less times than novices in the semantic menu, $F(1, 32) = 5.61$, $p < 0.05$, but get similar scores in the random menu, $F < 1$.

Number of targets found. Data from number of targets found showed a main effect of block, $F(2, 64) = 20.03$, $MSe = 0.77$, $p < 0.001$. Participants found less targets in the first block ($M = 8.4$), than in the second (M = 9.4) and in the third (M = 9.7). Regarding our predictions, the interaction between Expertise and Menu did not reach significance, F (1, 32) = 2.22, MSe = 1.67, $p = 0.14$. However, considering the previous pattern of results for the other two dependent variables, we considered the possibility that under that non-significant interaction relevant results could arise. For that reason, we performed two exploratory analyses as done for the previous dependent variables. Data showed that experts found more targets than novices in the semantic menu, F (1, 32) = 5.87, $p < 0.05$, but there were no differences in the random menu, $F < 1$. When only data from block 3 was considered, results showed no difference for expertise neither for the semantic menu, F (1, 32) = 1.31, MSe = 0.38, p = 0.26, nor for the random menu, $F < 1$.

-- Insert Table 1 about here --

General discussion

Results from the experiment support the hypothesis that the effect of domain expertise on IR depends on the combination of user knowledge and system characteristics. When the latter are compatible with expert mental representation (i.e. semantic grouping), knowledgeable users find more information, in less time and in a more efficient way than novice users. The expertise effect disappears when the characteristics of the system are not compatible with the knowledge of the expert user (i.e. random grouping). These conclusions are based on three different IR dependent variables (response time, lostness and number of targets found).

Results from the reported experiment challenge the robustness of the expertise effect found in previous literature on IR (Hollands and Merikle 1987, Shin *et al.* 1994, McDonald and Stevenson 1998a, 1998b, Patel *et al.* 1998, Last *et al.* 2001). In these experiments a random organized interface was not used, which can be the reason for the absence of a null effect of expertise on them. Although it can be argued that the use of a random organized interface is an unnatural setting, it allows us to better understand the mechanisms underlying expert performance in IR tasks, and has been used successfully as a research strategy in other IR experiments (e.g. Halgren and Cooke 1993, Niemela and Saariluoma 2003, Niemela *In press*).

In addition, the results of the experiment question the CFT that describes expert interaction in hypertext on the basis of the flexibility of the mental representation of the expert user (Spiro *et al.*, 1988, 1991). This theory can not explain why experts perform at novice level with the random organized interface, since in this case the expert should have restructured his / her flexible mental representation in order to map the way the information is organized on the interface. Following our results on response time and lostness, that restructuration does not appear even after the participant has gained some experience with the system, mainly after having been induced to visit all main pages at least twice.

Therefore, we propose that the CAH (Vicente and Wang 1997) can be considered as an alternative framework in order to explain expert performance in IR tasks. This theory predicts that the expertise effect would appear when the interface constrains the complexity of the task (e.g. the way the items on a menu are organized) in a way that can be matched with the expert mental representation. This can explain why experts outperform novices when interacting with a semantically organized interface: the expert can use the structure provided by the interface in order actively use his / her mental representation in order to find an item. In addition, the CAH predicts that when the organization of the information on the interface is incompatible with those of the expert, he / she would not be able to use his / her knowledge in the IR task. That explanation can account for the result of the experiment showing that experts perform at novice level in a random menu.

This study has benefited from the application of a theory that emphasizes the interactive nature of human cognition. User interaction with a system can not be explained by only focusing on his / her characteristics, nor only on system characteristics. On the contrary, this behaviour must be explained by the interaction of both human and system characteristics (Cañas *et al. In press*). Previous results reporting the isolated effects of expertise and semantic grouping in IR has been widely reported on the literature (see introduction). However, the complete picture of expert behaviour in IR is not explained until both human and system characteristics are explored in conjunction.

Results reported in this study have relevant implications for the design of hypertext systems. A well established design guideline asserts that system interfaces must map the way the user organizes its knowledge of the system in memory (his / her mental model of the system) (Norman 1988). Our results support this guideline in hypertext IR, but only if the system is devoted to domain expert users. If the system is mainly dedicated to novice domain users, the effort of engaging an analysis of the user's mental model can be avoided. In the experiment reported here, only expert users have benefited from mental model mapping, but it has had no influence on novice users. As exemplified in the above experiment, a simple way of mapping the user's mental model onto a hypertext could be to use the Gestalt laws for facilitating perceptual groupings of related elements (e.g. Niemela and Saariluoma 2003). The grouping of related elements can be done by using the same colour for each related group, by arranging related elements close to each other, or by connecting related elements by a graphical element (lines, boxes...). However, in the case that the mapping of the user's mental model is not done, attention must be paid to the use of design elements in the interface which can possibly induce an arbitrary grouping of elements. As shown in our experiment, the performance of expert domain users in IR can be hampered.

References

CAÑAS, J.J., QUESADA, J., ANTOLÍ, A. and FAJARDO, I., 2003, Cognitive flexibility and adaptability to environmental changes in dynamic complex problem-soling tasks, *Ergonomics*, **46**, 482-501.

CAÑAS, J.J., SALMERÓN, L., and FAJARDO, I., *In press*, Toward the analysis of the interaction in the joint cognitive system. In *Future interaction design*, edited by A. Pirhonen, H. Isomäki, C. Roast and P. Saariluoma (London: Springer-Verlag).

CHEN, C., and RADA, R., 1996, Interacting with hypertext: a meta-analysis of experimental studies. *Human-Computer Interaction*, **11**, 125-156.

HALGREN, S. L., and COOKE, N. J., 1993, Towards ecological validity in menu research. *International Journal of Man-Machine Studies*, **39**, 51-70.

HOLLANDS, J. G., and MERIKLE, P. M., 1987, Menu organization and user expertise in information search tasks. *Human Factors*, **29**, 577-586.

LAST, D., O'DONNELL, A., and KELLY, A., 2001, The effects of prior knowledge and goal strength on the use of hypertext. *Journal of Educational Multimedia & Hypermedia*, **10**, 3-25.

NIEMELA, M., *In press*, Layout arrangement and pop-up labels: effects on search. *Scandinavian Journal of Psychology*.

NIEMELA, M., and SAARILUOMA, P., 2003, Layout attributes and recall. *Behaviour & Information Technology*, **22**, 353-363.

MCDONALD, S., and STEVENSON, R. J., 1998a, Effects of text structure and prior knowledge of the learner on navigation in hypertext. *Human Factors*, **40**, 18-27.

MCDONALD, S., and STEVENSON, R. J., 1998b, Navigation in hyperspace: An evaluation of the effects of navigational tools and subject matter expertise on browsing and information retrieval in hypertext. *Interacting with Computers*, **10**, 129-142.

NORMAN, D. A., 1988, *The psychology of everyday things* (New York: Basic Books).

PATEL, S. C., DRURY, C. G., and SHALIN, V. I., 1998, Efectiveness of expert semantic knowledge as a navigational aid within hypertext. *Behaviour & Information Technology*, **17**, 313-324.

SHIN, E. C., SCHALLERT, D. L., and SAVENYE, W. C., 1994, Effects of learner control, advisement, and prior knowledge on young students' learning in a hypertext environment. *Educational Technology Research and Development*, **42**, 33-46.

SMITH, P.A., 1996, Towards a practical measure of hypertext usability. *Interacting with Computers*, **4**, 365-381.

SPIRO, R. J., COULSON, R. L., FELTOVICH, P. J., and ANDERSON, D. K., 1988, Cognitive flexibility theory: advanced knowledge acquisition in ill-structure domains. In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (Hillsdale, NJ: Lawrence Erlbaum Associates), pp. 375-383.

SPIRO, R. J., FELTOVICH, P. J., JACOBSON, M. J., and COULSON, R. L., 1991, Cognitive flexibility, constructivism and hypertext: random access instruction for advanced knowledge acquisition in ill-structured domains. *Educational Technology*, **31**, 24-33.

VALERO, P., and SANMARTIN, J., 1999, Methods for defining user groups and user-adjusted information structures. *Behaviour & Information Technology*, **18**, 245-259.

VICENTE, K. J., and WANG, J. H., 1998, An ecological theory of expertise effects in memory recall. *Psychological Review*, **105**, 33-57.

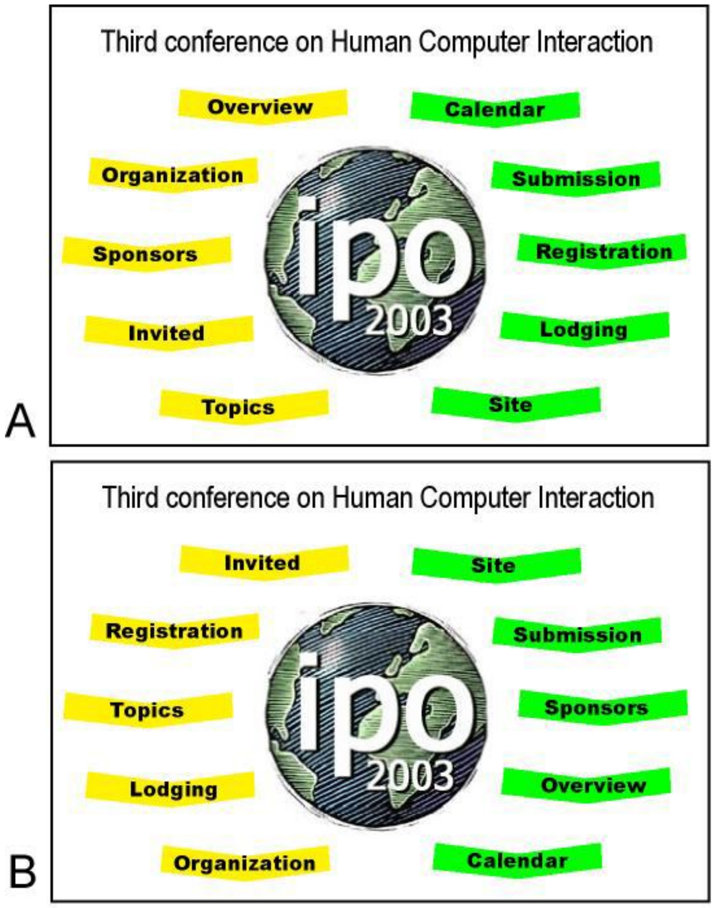

Figure 1. Different menu interfaces used in the experiment (the originals were constructed in Spanish).

(A) Semantic menu; (B) Random menu.

Table 1. Means and standard deviations (in parenthesis) for performance variables.

	Expertise & Type of menu			
	Novice		Expert	
	Semantic	Random	Semantic	Random
Targets found	8.9 (1.2)	8.9 (2.9)	9.7 (0.4)	9 (0.9)
Response time	13.3 (5.7)	12.2 (5.6)	8.2 (3)	11.9 (3.2)
Lostness	0.23 (0.18)	0.16 (0.12)	0.12 (0.12)	0.19 (0.15)

Author Notes

This research was funded by a research grant from the Spanish ‘Dirección General de Universidades del Ministerio de Educación’ to the first author.