Hierarchical histories for hypertext navigation: a pilot study of individual differences

Hierarchical histories for hypertext navigation: a pilot study of individual differences*

Jon Oberlander and **Richard Cox** and **Robert Inder** and **Richard Tobin**
Human Communication Research Centre
University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, Scotland
{J.Oberlander, R.Cox, R.Inder, R.Tobin}@ed.ac.uk

Abstract

Two hypermedia navigation aids are virtually ubiquitous—the history list, and the graphical overview. However, it can be argued that there is an increasing need for a half-way house between the linear record of the individual's session, and the hierarchical or network representation of abstract document structure. We are therefore investigating 'hierarchical histories'; one particular approach to this enables backtracking to be based on the structure of the user's dialogue with the hypermedia system. We believe that such 'discourse-structural backtracking' (DSB) avoids various difficulties that have been pointed out in the literature, and improves a hypertext system's usability. To test this, we have implemented DSB in *Emacs Info*, and run some simple experiments; the pilot results confirm the importance to navigation of individual differences in cognitive style. In particular, we believe that DSB may help alleviate the navigational problems suffered by users with poorer spatial visualisation skills.

Introduction

Two hypermedia navigation aids are virtually ubiquitous—the history list, and the graphical overview. However, we may also need a half-way house between the linear record of the individual's 'previous personal use of the system' (Neilsen 1990), and the hierarchical or network representation of abstract document structure. We are therefore investigating 'hierarchical histories' and 'discourse-structural backtracking', aiming at avoiding some of the difficulties that have been observed in the literature.

The rest of this document is structured as follows. First, we discuss the notion of hierarchical histories, emphasising that the need for them has already been recognised in the literature. In sections 3 and 4 we quickly sketch the relevant ideas from discourse theory, and the implementation we built to explore their efficacy. These two sections rehearse material that has been reported elsewhere (Inder and Oberlander 1994, 1995), and can be skipped by the informed reader. The next section outlines the method and materials pursued in the empirical evaluation originally reported by Davis (1994). The results from that study initially appeared inconclusive, but the following section argues that the data can be re-analysed, taking into account individual differences between users, focussing on known indicators of cognitive style. Hence, we then report this pilot re-analysis, and the suggestive results that have emerged. We conclude by considering some further data, and indicate the likely future direction of full studies.

Hierarchical histories

The need for an 'intermediate' navigation tool is particularly apparent on the World Wide Web. The most popular browsers, Netscape and Mosaic, provide histories, which are *linear* lists of recently visited nodes, and use the same data structure to support a basic form of backtracking—chronological, or single visit (Bieber and Wan 1994)—in which users may retrace their steps through a space of documents.[1] However, offering a graphical overview would help to make the *hierarchical*, or network, structure of the documents available to the user. Graphical overviews are therefore on the Web project's invited contribution list (Berners-Lee 199?), and are already supported in Hyper-G. The problem remains that straightforward ways of building such tools (Halasz, Moran and Trigg 1987) fail with long sessions or with large or ill-structured documents—both of which are common on the Web. The key difficulty is that the graphs just get too big, or too messy. Obviously, differential magnification of the graph (Furnas

*Submitted for the MIT Press CD-ROM book based on the first international workshop on Intelligence and Multimodality in Multimedia Interfaces, held in Edinburgh, July 1995.

[1] They also offer a simple footprinting mechanism, to indicate which links have already been followed, together with manually-added bookmarks, which in Netscape can be hierarchically organised. Bookmarks, of course, typically tell you where you might go, not where you have just been.

1986; Misue, Eades, Lai and Sugiyama, 199?) can greatly increase the ability of a graphical representation to convey useful detail within the context of a 'whole'. But the scale of the problem confronting Web browsers requires enormous information suppression, and attempts at aggregation (that is, information hiding) based on the content of the node (Mukherjea, Foley and Hudson, 1994; Zizi and Beaudouin-Lafon, 1994) may ultimately prove useful.

Nonetheless, there is a third way, and it should be of help in any browser, not just the ones pointed at the Web. Instead of linear history lists, and hierarchical document overviews, we can aim at HIERARCHICAL HISTORIES (cf. Foss 1988). But to give histories that extra dimension, we need to recognise the task-oriented structure that individual users impose on the information. Dömel's (1994) WEBMAP represents one possible approach, while HYPERFLEX uses more sophisticated techniques to recognise user goals from analysis of node content (Kaplan, Fenwick and Chen, 1993). Bieber and Wan (1994) raise the possibility of recognising user task structure from access sequences, and using it to guide backtracking.

In particular, they propose 'detour-removing' backtracking, which avoids re-visiting nodes which were only visited by mistake. They suggest that task structure can be used to tell when the user has returned to the main navigation sequence after a detour, and that the detour can then be eliminated from the history—and hence, from any subsequent backtrack. The history would be hierarchical in the sense that node-visits would categorised as main or subordinate, orthogonally to their sequential position. However, once the categorisation was made, the history would be destructively altered, and converted back into a linear list—albeit a 'relevant' one.

As Neilsen (1995:252) observes, the problem with detour-removal 'is obviously to detect the detours and there is currently no empirical evidence to suggest that this is possible in the general case'. Elsewhere, in fact, Neilsen (1990) experimented with a backtrack mechanism that effectively removed irrelevant nodes from the history. As well as a standard backtrack button, his HYPERTEXT87 system also supported an 'express return' button, in one particular location. This allowed users to make a double backtrack, 'providing a shortcut through the straight backtrack path' [p309]. However, he found that 'almost no test users understood this facility' (Neilsen 1995:32).

Inspired by the study of what makes a body of language cohere (cf. for instance, Grosz and Sidner 1986), we have developed an idea broadly similar to Bieber and Wan's, albeit extracting information from document structure rather than window operations. Retrieving this information allows us to offer a principled mechanism for detecting detours, and allowing backtracking to bypass them, thereby providing a systematic method for express returns. We believe it thereby avoids Neilsen's objections, and may be worth deploying more widely.

Discourse-structural backtracking

Elsewhere, we have observed that discourse theory's aim is to model the construction and evolution of the structures underlying extended discourses (Inder and Oberlander 1995). Most approaches agree that discourse is hierarchically structured, and that a limited set of relations link its sub-parts. Within this framework, a theory must explain how new information is integrated into the existing structure of a discourse. For instance, it is important to model the process of 'discourse popping', which occurs whenever the topic under discussion changes, or reverts back to an older topic.

Where approaches diverge is over the number of relations (and levels of representation) proposed by the theories. At one extreme, Rhetorical Structure Theory (Mann and Thompson 1987) recruits some 23 relations, including Explanation and Parallel. At the other, Grosz and Sidner (1986) support just two relations; so discourse purposes are related by either immediate dominance (a part-whole, subordination relation); or satisfaction precedence (a sequencing, coordination relation).

Applying a two-relation theory of discourse structure to hypertextual discourse suggests a number of more-or-less obvious theoretical identifications.

On the one hand, we can look at the static document structure. First, `Next` and `Previous` nodes can be seen as coordinated with the current node. Secondly, sub-nodes, and hyper-linked nodes can be seen as subordinated to the current node. Finally, linking in general is equivalent to discourse attachment.

On the other hand, we can see how an individual's specific dialogue through their browser will map this hierarchy or network into another discourse structure. This hierarchical structure is clearly richer than a linear history, since it encodes the relationships between embedding discourse contexts. With this structure, we can stipulate that the rules for discourse attachment should apply as in conversation. As with co-operative conversation, there ought to be navigational features that make it obvious when a discourse pop or push is possible, and where the topic would switch to.

This additional structure is sufficient for a hierarchical history mechanism. The history could be displayed as an actual graph of session structure, as in

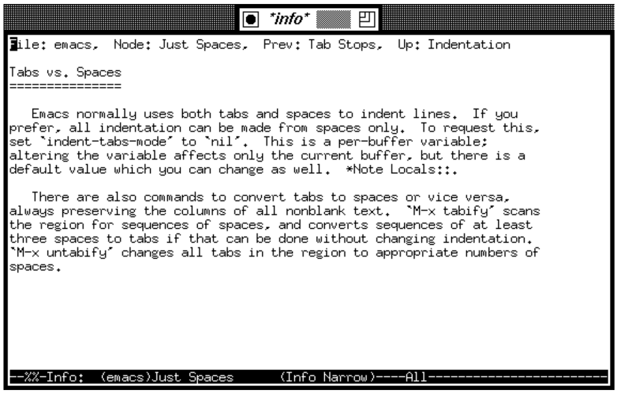

Figure 1: A node presented by *Info*. Anchors at the top indicate the destinations of single-key commands to move to `previous` and `up` nodes. '*Note Locals::' is an example of an embedded link.

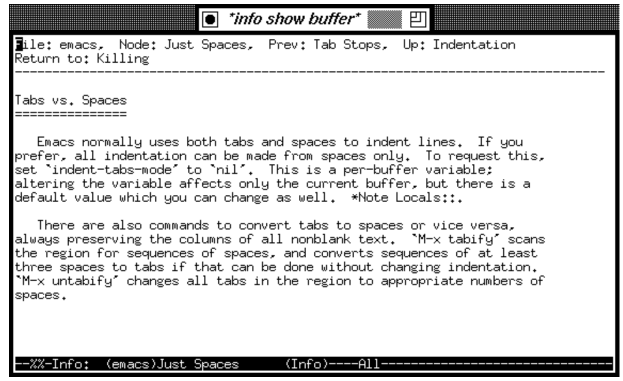

Figure 2: Figure 1's node as presented by *DS-Info*, which has added a `Return to` anchor.

Dömel (1994), and this is our ultimate aim. However, our initial work has focussed simply on showing that *any* access to this hierarchical history would help navigation. Hence, we have implemented a DISCOURSE-STRUCTURAL BACKTRACKING mechanism. If hierarchical history can make a positive difference *without* a graphical display, then it should be even more effective with one.

Implementation: DS-INFO

In Inder and Oberlander (1994, 1995), we have described *DS-Info*, and we here lay out its key features as briefly as possible. *DS-Info* is based on the *Info* hypertext help system embedded within the *Emacs* family of editors. *Emacs* 18 *Info* handles only plain ASCII text, and expects keyboard commands (see Figure 1). However, it supports several link types, allowing document structure to be made explicit.

How do we end a digression, and return to the original point (cf. Foss 1988)? In *Info*, `Note`s often initiate digressions, and there is only one way back from them: backtracking. But this can be inefficient and distracting if a sequence of nodes has been followed after a `Note`. Thus, *DS-Info* offers a command, `Return to`, which closes off a digression and returns to the previous context—it functions as a discourse pop, just like saying *Anyway!* in a conversation. Following a `Note`, or jumping to an arbitrary node, opens a new discourse context, a daughter to the current one. When a node is displayed as part of such a nested discourse, *DS-Info* creates a link to the point of departure from the mother discourse and adds it as a `Return to` anchor on the target node.[2] Compare Figures 1 and 2.

This, then, is the essence of discourse-structural backtracking: the user's discourse structure is used automatically to locate places from which they have digressed, and the most recent of these is presented as a navigation option. As well as this, however, *DS-Info* exploits the idea that if we follow a digression, the target text is usually presented in a different context from the one for which it was written. This converts whatever discourse information the author has incorporated into an irrelevant distraction. *DS-Info* handles this by removing any 'discourse context' anchors from nodes which are reached out of context: the reader can still see links to daughter nodes, but those to sister or mother nodes are suppressed. In this way, a node will bear different anchors, depending on whether it's reached in or out of context. See Figure 3.

In fact, informal evaluation quickly revealed that suppressing mothers was a bad idea. *DS-Info* frequently presented text that built on other nodes that had not themselves been presented, and provided readers with no way to access the prerequisite text. We therefore added a new anchor, leading to information `About` a node that is reached out-of-context. The destination of this link is the node that was originally `Up` from the current node, but the link signals a different purpose within the discourse, and the reader following it arrives at this mother node out of context. This facility resembles a limited version of Instone, Teasley and Leventhal's (1993) bidirectional 'incoming links'. In their system, HYPERHOLMES II, users were presented with 'a list of other nodes in the document that have links pointing to the current node' [p502]. `About` offers a single, 'standard' incoming link to the new node. See Figure 4.

[2] In the current *Emacs* 19 implementation, the anchors are active, so navigation commands can be moused as well as keyed. In the *Emacs* 18 version evaluated, however, only keyboard commands were supported.

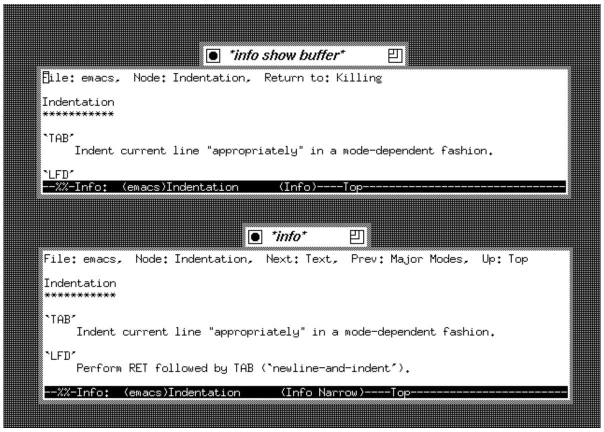

Figure 3: Different headers for the same node. The upper window shows how *DS-Info* has modified the header specified in the document, shown in the lower window, by removing the structural links and adding a `Return to` anchor.

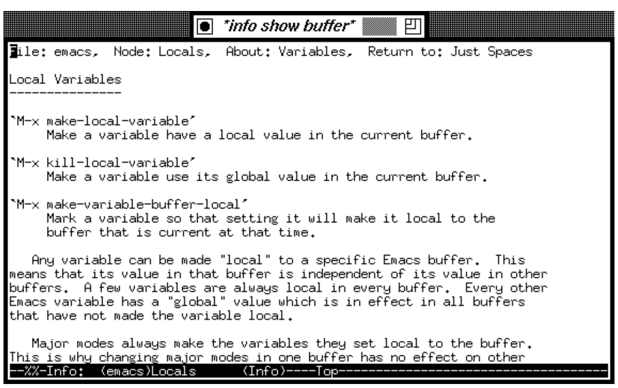

Figure 4: A node showing an `About` anchor. This is the result of using *DS-Info* to follow the link in the text of the node shown in Figure 2: note how the `Return to` anchor now indicates that node.

Cognitive evaluation

So, we have suggested that discourse structure supports a hierarchical history, and that backtracking can be designed to exploit it. Is this form of backtracking any use? To test the system, a controlled laboratory experiment was conducted; the original results were reported in Davis (1994).

There were 12 subjects, all postgraduate students who were regular users of *Emacs*, but novice users of *Info*. Their task was to solve four problems, using a constructed information-base, about the facilities in a fictitious city called Lance, using either *Info* or *DS-Info* as hypertext engines. The problems were like this:

> I am currently at the Ski Centre and have to get to a friend at the Pelican Cafe in 15 minutes. It is currently 3pm. I have only got $3.00 left. How do you recommend I get there? I know I can because I have done it before. I get no concessions.

On average, a problem could take between 10 and 15 minutes to solve. Each subject used each system for two questions, and the experimental design was controlled for order of system- and question-presentation. A number of measures were taken, including time-to-solution, accuracy-of-solution, and full time-stamped system-logs.

From the system-logs, we can compute various measures. For instance, we can calculate the visits-per-node executed by a subject on a problem. This is just the total number of node-visits divided by the number of distinct nodes visited, and is the inverse of the proportion of visits which are made to previously unvisited nodes. Or we can count the top node visits executed by a subject on a problem, which is the number of visits to the information-base's initial node. To illustrate, the fragment of system-log in Table 1 yields 2.11 visits-per-node, and 4 visits to top.

Davis compared subjects' performance on the problems, using the two systems. He found no effect from order of system-presentation, but also no significant differences on time or accuracy. However, there were some encouraging trends. For instance, it appeared that *DS-Info* use was associated with fewer errors in subjects' solutions. And when *DS-Info* was used after *Info*, use of its `Return to` command apparently superceded `Last`, which it still supported—so subjects were at least not confused about it.

Most interestingly, for current purposes, however, was the indication that top-node visits were reduced. On the fourth problem, the reduction approached significance. When attempted using *Info*, mean visits to top = 4.67, $n = 6$. When attempted with *DS-Info*, mean visits to top = 2.33, $n = 6$. A one tailed t-test

Table 1: A fragment of a subject's timestamped system log

```
Key:  reset    Time:  14:13:30   Node:  "Top"
Key:  menu     Time:  14:14:23   Node:  "City"
Key:  follow   Time:  14:14:31   Node:  "Travel"
Key:  menu     Time:  14:14:39   Node:  "Rush Hour"
Key:  last     Time:  14:14:46   Node:  "Travel"
Key:  menu     Time:  14:15:06   Node:  "Metro"
Key:  menu     Time:  14:15:08   Node:  "Metro Map"
Key:  last     Time:  14:15:18   Node:  "Metro"
Key:  top      Time:  14:15:31   Node:  "Top"
Key:  menu     Time:  14:15:34   Node:  "City"
Key:  menu     Time:  14:15:54   Node:  "Entertainment"
Key:  last     Time:  14:16:01   Node:  "City"
Key:  menu     Time:  14:16:06   Node:  "Sports"
Key:  menu     Time:  14:16:21   Node:  "Skiing"
Key:  top      Time:  14:16:59   Node:  "Top"
Key:  menu     Time:  14:17:01   Node:  "City"
Key:  follow   Time:  14:17:05   Node:  "Travel"
Key:  top      Time:  14:17:30   Node:  "Top"
Key:  menu     Time:  14:17:33   Node:  "City"
  ⋮              ⋮                  ⋮
```

indicated that the fall approached statistical significance: $t = -1.56, df = 10, p = 0.074$.

Though not conclusive, this apparent trend is suggestive in the light of previous work on individual differences in navigation. Our recent work on the cognitive effects of multimodal learning environments has also focussed on the influence of cognitive style (cf. Stenning, Cox and Oberlander, 1995); we therefore briefly review the issues concerning individual differences in navigation, in the next section.

Individual differences

Vicente, Hayes and Williges (1987) examined individual differences among users of a hierarchical file system. They used a battery of psychometric tests, and found that vocabulary and spatial visualization tests were the best predictors of task performance, accounting for 45% of the variance in the data. The Educational Testing Service's VZ-2 test of spatial visualization requires subjects to indicate the appearance of an unfolded piece of paper, which, while folded, has had holes punched through it. Ability on the VZ-2 test was the best predictor, and it was found that subjects with low spatial ability took twice as long to perform an information retrieval task, compared with highly spatial subjects. Examination of command use frequencies indicated that low spatial subjects were getting lost in the file hierarchy, and having to retrace their steps.

Campagnoni and Erlich (1989) applied similar techniques in a hypertext context. Using the VZ-2 test, they found that spatial visualization ability correlated positively with efficiency of information-location in hypertext; low spatial subjects took around 1350 seconds for a task, compared with highly spatial subjects, who were almost twice as fast, taking around 750 seconds. They also found that there was a highly significant negative correlation between subjects' spatial ability and the number of visits they made to the top node of the hypertext ($r = -0.75, p < 0.005$); low spatial subjects made around 10 visits to top on a task, compared with highly spatial subjects, who made around 6 visits.

From this, Campagnoni and Erlich argue that 'individuals with good spatial visualization skills can construct a better internal model of the information architecture of a hierarchical hypertext system' [pp287–8]. It follows from this that navigation facilities, which aim to help subjects build better models of information architectures, will be of particular help to the subjects with low spatial skills.

Re-interpreted results

The design of Davis (1994) did not include psychometric tests. However, in view of Campagnoni and Erlich's findings, we can use a *surrogate* measure to split the population into two spatial ability groups. We can do this because the Lance information-base is also hierarchical, with a top node; in this architecture, and with standard navigation facilities, there is a strong negative correlation between spatial ability and number of visits to top. Hence, we can perform a median split, dividing the group into two groups of 6: those who made relatively few visits to top when using the standard system; and those who made relatively many visits to top when using the standard system. We term the former, more spatially able, group TopLo; we term the latter, less spatial, group TopHi.

We should emphasise here that this TopLo/TopHi categorisation is only a surrogate for a categorisation based directly on spatial ability. It is not a genuine substitute for the latter, but it does at least provide a means of re-interpreting Davis's (1994) data in the light of individual differences. The following results must obviously be viewed as pilot indications—providing reasons for pursuing further investigation—rather than as definitive results.

So, recall that taking the population of 12 as a whole, we did not find any significant differences between their task performances on *Info* and *DS-Info*. Now, however, we can consider whether TopLo and TopHi subjects differ in the way they respond to the two systems. If Campagnoni and Erlich are right, then better navigation will help the low spatial TopHi subjects most. We therefore have a specific, directional, hypothesis to test: does *DS-Info* alleviate the task performance of TopHi subjects in any way? Table 2 summarises our findings.

There are two main points to be made. On the one

Table 2: Individual differences and response to discourse-structural navigation.
Info is the standard system; *DS-Info* adds discourse-structural navigation. TopLo subjects make relatively few visits to the top node; TopHi make relatively many. Top v: mean visits to top; v/n: mean visits per node; w/m: mean non-top visits to non-top node. t = mean time in seconds on problem.

	Top v	v/n	w/m	t
INFO				
TopLo	2.83	2	1.96	735
TopHi	7.5	2.85	2.65	1059
DS-INFO				
TopLo	3.17	2.19	2.13	931
TopHi	5.75	2.72	2.57	846

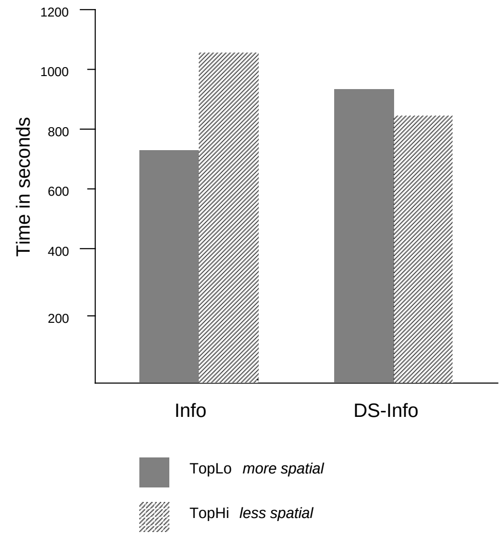

Figure 5: Mean time to solution. The difference in speed is significant when the groups use *Info* ($t = -1.97, df = 10, p < .05$), and insignificant when they use *DS-Info*.

hand, the analysis indicates that, for a given ability group, there are no direct differences between their performance on the two systems. On the other, however, there are differences on the between-group measures, and a suggestive result which pertains to differing system use.

First, then, it should be noted that TopLo subjects (the highly spatial) show no significant differences between their performance on *Info* and *DS-Info*, on any of the selected parameters. Equally, TopHi subjects (the low spatial) only approach a significant difference on their performance on *Info* and *DS-Info* on one parameter. Their top-visits decline from 7.5 to 5.5; $t = 1.67, df = 5, p = 0.08$.

Secondly, separate 2 × 2 analyses of variance (ANOVAs) were conducted for the four dependent measures—visits to top, visits per node, visits per non-top mode and time per problem. The ANOVAs were of mixed design. Groups was the between-groups factor with two levels (TopLo, TopHi). System used (*Info*, *DS-Info*) was treated as a two-level, within-subject, repeated measures factor, since each subject used both systems (note that order of exposure to systems was counterbalanced across subjects in the study).

Visits to top: top v The ANOVA results for visits to top revealed a significant main effect for group $F(1, 10) = 9.71, MS = 78.84, p < .02$. TopHi subjects made significantly more visits to top than TopLo subjects on both systems. This is not too surprising, since visits to top was the basis of dividing subjects into TopLo and TopHi groups. On the *Info* system, the mean visits to top for TopLo subjects was 2.83, compared with 7.5 for TopHi. On the *DS-Info* system, the means were 3.17 and 5.75 for TopLo and TopHi, respectively.

The main effect for system and the group-by-system interaction were not significant.

Visits per node: v/n Again the main effect for group was significant ($F(1, 10) = 5.07, MS = 2.88, p < .05$). On *Info*, the means for TopLo and TopHi respectively were 1.99 and 2.85. On *DS-Info*, they were 2.19 and 2.71. As the means suggest, the main source of significance in the groups main effect was on *Info* and not *DS-Info*. Post-hoc t-tests revealed a significant difference between the group means on *Info* ($t = -2.12, df = 10, p < .05$), but no significant difference between the group means on the *DS-Info* system.

The main effect for system and the interaction of group-by-system were not significant.

Visits per non-top node: w/m No significant effects were found in the ANOVA. However, a t-test between TopLo and TopHi mean visits per non-top node (means were 1.96 and 2.65 respectively) closely approached significance ($t = -1.17, df = 10, p = .057$).

Time: t No significant effects were found in the ANOVA. However, a t-test between the TopLo and TopHi means on *Info* revealed a significant difference ($t = -1.97, df = 10, p < .05$). Using *Info*, TopLo took a mean 735 seconds per problem, compared with TopHi, who took 1059 seconds. Using *DS-Info*, TopLo took a mean 931 seconds, compared with TopHi's 846 seconds, and this latter difference was not significant. See Figure 5.

ANOVA is a multivariate analysis technique which permits the analysis of interactions between variables, as well as main effects. The use of ANOVA also overcomes the problem of chance significance that is associated with multiple t-tests. In the results reported here, post-hoc t-tests were primarily used as a means of partitioning significant main effects detected by the ANOVAS.

Discussion

The re-interpreted results offer indirect evidence for the hypothesis that *DS-Info*'s implementation of discourse structural backtracking aids most those with low spatial abilities. In effect, it could be ameliorating their poorer spatial visualization skills, and pulling them up to the navigation standards of those with better spatial skills.

Is there any way of understanding what is changing, as the subjects adapt to the different systems? At the moment, we can only speculate, and the discussion in this section should be read in this light—we do not claim any definitive results. However, we do have two further sources of evidence which may give some clues as to what adaptation involves: correlational analysis, and command selection vectors.

First, correlational analysis provides some insight. Keeping the two ability groups separate, we can compare the extent to which the parameters (top-v, v/n, w/m, t) are intercorrelated, contrasting this across the two systems. It seems from this analysis that the highly spatial subjects don't change much between *Info* and *DS-Info*: their individual v/n seem to correlate well across the systems, and with w/m. However, this is *not* so for low spatial subjects. One might conclude that it is the TopHi who are changing. However, it also seems that for highly spatial subjects using *DS-Info*, both v/n and w/m correlate very well with time: visits per node and time spent on the problem are highly correlated ($r = 0.98, p < 0.01$), as are visits per non-top node and time ($r = 0.98, p < 0.01$).

In itself, the idea that mean visits per node correlates well with time may not sound so surprising. But what is odd is that the level of correlation holds *only* for highly spatial subjects, and *only* when they use *DS-Info*. So the surprising thing is, perhaps, that visits per node, as an index, isn't significantly correlated with time for low spatial subjects, whatever system they use, or for highly spatial subjects using *Info*.

Secondly, preliminary analysis of the commands selected by subjects is intriguing. We follow Vicente et al., and Elkerton and Williges (1985), in constructing a vector to summarise command usage across the subject groups and systems. Each subject is given a vector of 11 votes, one for each of the mnemonic one-key commands available in *DS-Info*. The commands are: `follow`, `last`, `menu`, `next`, `top`, `up`, `previous`, `goto`, `search`, `return`, and `about`. The last two of these, of course, are *only* available in *DS-Info*. A vote is cast for a command if it selected at least once on a problem. Keeping the two groups of 6 subjects separate, the votes are summed for the 2 problems tackled with each system. Thus, for instance, every TopLo subject uses the menu command `m` on both questions they tackle with *Info*, so the command scores 12. This polling procedure prevents the bias that would arise from repeated patterns of behaviour. The results are summarised in Table 3, where commands are identified by their single key invocations.

Table 3: Individual differences and command selection vectors.
Info is the standard system; *DS-Info* adds discourse-structural navigation. TopLo subjects make relatively few visits to the top node; TopHi make relatively many. The eleven mnemonic one-key commands are spelt out in the text. A maximum score of 12 indicates that all subjects used the command on both problems attempted with that system; 0 indicates that no-one used that command on either problem.

	COMMANDS										
	`f`	`l`	`m`	`n`	`t`	`u`	`p`	`g`	`s`	`r`	`a`
INFO											
TopLo	10	2	12	6	7	12	7	8	4	0	0
TopHi	12	11	12	5	9	9	3	0	4	0	0
DS-INFO											
TopLo	11	1	12	8	8	9	6	10	6	12	6
TopHi	12	7	12	3	8	10	0	3	5	9	2

Seven commands have similar usage patterns across systems: `f`, `l`, `m`, `t`, `p`, `g` and `s`. Four commands have usage patterns that vary across systems: `n`, `u`, `r` and `a`; of these, `r` and `a` are only available in *DS-Info*. Most interestingly, four commands have usage patterns that vary across *subject groups*: `l`, `p`, `g` and `n`. `last` actually remains constant across systems, although it is much less used by the highly spatial TopLo—but these subjects adopt `return` enthusiastically, scoring 12, against the 9 from the TopHi subjects. Use of `previous` and `goto` is also more prevalent amongst the TopLo subjects, but again is not influenced by system.

The most revealing patterns concern `next` and `up`. TopLo subjects use `next` more widely on *DS-Info* than on *Info*; TopHi subjects use `next` about as much as

TopLo, when they use *Info*. However, when TopHi use *DS-Info*, they use `next` *less* widely than on *Info*. For this command, then, the subject groups apparently respond to *DS-Info* in opposite ways. And while the usage pattern for `up` is broadly similar between the subject groups, we find that TopLo subjects use `up` less wiedly on *DS-Info* than on *Info*, and that by contrast, TopHi subjects use `up` *more* widely on *DS-Info* than on *Info*. Again, the behaviour diverges.

This suggests that spatial visualisation ability might influence how subjects respond to a system incorporating discourse-structural backtracking (DSB). Highly spatial subjects didn't use ordinary backtracking (through `last`), but they all adopt DSB. They don't need to use `up` as much, to get them out of deeply embedded nodes, but they use `next` more than before, perhaps because `return` will always get them back to the point. Their command usage is still 'deep', but seems somehow 'broader' than before. By contrast, low spatial subjects respond in a different manner. First, DSB augments ordinary backtracking—it does not replace it. The subjects are using `up` a little more than before, and using `next` rather less. Their usage pattern seems, if anything, 'deeper' than before.

So, even if *DS-Info* removes performance differences between subjects of differing spatial ability, it may still be changing both types of subject. Low spatial subjects may be navigating the information base with deeper browsing patterns. High spatial subjects—already more confident, with less conventional backtracking—are able to navigate with patterns that may be no less deep than before, but do have broader portions. And these navigation patterns allow highly spatial subjects to solve problems at a rate directly proportional to the number of visits they make to each node in a session.

Conclusion and Future Directions

These pilot results indicate that in a full study, it will be worth testing for a clear effect based on individual differences in cognitive style. The discussion of command selection suggests how such an effect could arise. We conclude that there are preliminary indications that: (i) discourse-structural backtracking could help users with poorer spatial skills approximate the search performance of those with better skills; and (ii) it may affect people with good spatial skills, in ways that deserve further investigation.

There is therefore provisional evidence for the utility of the hierarchical history that underpins discourse-structural backtracking. We can avoid the two obvious pitfalls observed by Neilsen, discussed in section . *DS-Info* has shown that there is a way of detecting structure that can be used to inform backtracking (cf. Neilsen 1995:252); and that there is a way of implementing this 'express' backtracking which is readily understood by the users (cf. Neilsen 1995:32). In our most recent work, we have developed a graphical version of *DS-Info*, known as *GDS-Info*, which directly displays the hierarchical history of the session, exploiting animated graphing, and permitting graph-based navigation actions. We aim to carry out larger-scale evaluations of this system, prior to porting its capabilities into a Web browser. These evaluations will again focus on individual differences, and build on the experience reported here.

In the first instance, we are currently running new tests on *DS-Info*, administering direct psychometric tests for individual differences in spatial and visual skills. Two tests from the Educational Testing Service's factor-referenced cognitive test kit (Ekstrom, French and Harmon, 1976) are being employed. The first is the VZ-2 'paper-folding' test, a measure of spatial ability. The ETS test manual (1976:173) states that the VZ-2 and related tests measure 'The ability to manipulate or transform the image of spatial patterns into other arrangements'. The second test is the SS-1 'maze-tracing' speed test. This measures subjects' speed in exploring visually a wide or complicated spatial field. Both tests might be expected to measure some components of a 'visualizer-verbalizer' dimension of cognitive style (Jonassen and Grabowski, 1993)

Secondly, given the suggestive results from the command selection vectors, we need to consider how better to characterise user navigation strategies, building on the analysis of Canter, Rivers and Storrs (1985). We are already refining our measures of user activity, to discriminate better between types of re-visit to nodes.

Finally, in evaluating *GDS-Info* itself, we would aim to use equally controlled but more naturalistic materials for our information-base; through the World Wide Web, much larger naturally-occurring hypertexts should be accessible for experimentation.

Acknowledgements

The support of the Economic and Social Research Council for HCRC is gratefully acknowledged. The first author is supported by an EPSRC Advanced Fellowship. Special thanks to Matt Davis, whose original data is reanalysed here, and to Corin Gurr for his good advice. This paper is based on the presentation given at the first international workshop on Intelligence and Multimodality in Multimedia Interfaces 1995, in Edinburgh, July 1995. We are grateful to members of the audience for their valuable feedback, and to anonymous reviewers of a draft of this paper.

References

Berners-Lee, T. J. (199?) "Software Projects for the Enthusiastic". WWW project document. http://www10.w3.org/hypertext/WWW/Bugs.html.

Bieber, M. and Wan, J. (1994) Backtracking in a Multiple-window Hypertext Environment. In *Proceedings of the ACM European Conference on Hypermedia Technology*, pp158–166. ACM Press, New York.

Campagnoni, F. R. and Erlich, K. (1989) Information Retrieval Using a Hypertext-Based Help System. *ACM Transactions on Information Systems*, **7**, 271–291.

Canter, D., Rivers, R. and Storrs, G. (1985) Characterizing user navigation through complex data structures. *Behaviour and Information Technology*, **4**, 93–102

Davis, M. (1994) *Addition and Evaluation of Discourse Structure to Hypertext Navigation.* MSc thesis, Cognitive Science, University of Edinburgh. September, 1994.

Dömel, P. (1994) Webmap—A Graphical Hypertext Navigation Tool. In *Proceedings of the 2nd International WWW Conference*, Chicago, Fall 1994. http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/Searching/doemel/www-fall94.html.

Ekstrom, R. B., French, J. W. and Harmon, H. H. (1976) *Manual for Kit of Factor Referenced Cognitive Tests.* Princeton, NJ: Educational Testing Service (ETS).

Elkerton, J. and Williges, R. C. (1985) A performance profile methodology for implementing assistance and instruction in computer-based tasks. *International Journal of Man-Machine Studies*, **23**, 135–151.

Foss, C. L. (1988) Effective browsing in hypertext systems. In *Proceedings RIAO'88 Conference on User-Oriented Context-Based Text and Image Handling*, pp82–98, Cambridge, Ma., 1988.

Furnas, G. (1986) Generalised Fisheye Views. In *Proceedings of CHI '86*, pp16–23.

Grosz, B. and Sidner C. L. (1986) Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, **12**, 175–204.

Halasz, F. G., Moran, T. P. and Trigg, R. H. (1987) NoteCards in a nutshell. In *Proceedings of CHI '87 Conference on Human Factors in Computing Systems and Graphics Interface*, pp45–52.

Hyper-G and Harmony http://info.iicm.tu-graz.ac.at/

Inder, R. and Oberlander, J. (1994) Applying Discourse Theory to aid Hypertext Navigation. Poster presented at the ACM European Conference on Hypermedia Technology, Edinburgh, 1994. http://www.cogsci.ed.ac.uk/hcrc/wgs/graphics/publications/echt94.html

Inder, R. and Oberlander, J. (1995) Using Discourse to Aid Hypertext Navigation. In *Proceedings of HCI International*, pp195–200, Yokohama, July 1995.

Instone, K., Teasley, B. M. and Leventhal, L. M. (1993) Empirically-based Re-design of a Hypertext Encyclopedia. In *Proceedings of InterCHI '93*, pp500–506.

Jonassen, D. H. and Grabowski, B. L. (1993) *Handbook of individual differences, learning and instruction.* Hillsdale, NJ: Lawrence Erlbaum & Associates.

Kaplan, C., Fenwick, J. and Chen, J. (1993) Adaptive Hypertext Navigation Based On User Goals and Context. *User Modelling and User-Adapted Interaction*, **3**, 193–220.

Mann, W. C. and Thompson, S. A. (1987) Rhetorical Structure Theory: A theory of text organisation. Technical Report ISI/RS-87-190, USC/ISI, June, 1987.

Misue, K., Eades, P., Lai, W. and Sugiyama, K. (199?) Layout Adjustment and the Mental Model. *Journal of Visual Languages and Computing* (In Submission).

Mukherjea, S., Foley, J. and Hudson, S. (1994) Interactive Clustering for Navigating in Hypermedia Systems. In *Proceedings of the ACM European Conference on Hypermedia Technology*, pp136–145. ACM Press, New York, 1994.

Neilsen, J. (1990) The art of navigating through hypertext. *Communications of the ACM*, **33**, 296–310.

Neilsen, J. (1995) *Multimedia and Hypertext: The Internet and Beyond.* AP Professional, Cambridge, Ma..

Stenning, K., Cox, R. and Oberlander, J. (1995) Contrasting the cognitive effects of graphical and sentential logic teaching: reasoning, representation and individual differences. *Language and Cognitive Processes*, **10**, 333–354.

Vicente, K. J., Hayes, B. C. and Williges, R. C. (1987) Assaying and Isolating Individual Differences in Searching a Hierarchical File System. *Human Factors*, **29**, 349–359.

Zizi, M. and Beaudouin-Lafon, M. (1994) Accessing Hyperdocuments through Interactive Dynamic Maps. In *Proceedings of the ACM European Conference on Hypermedia Technology*, pp126–135. ACM Press, New York, 1994.