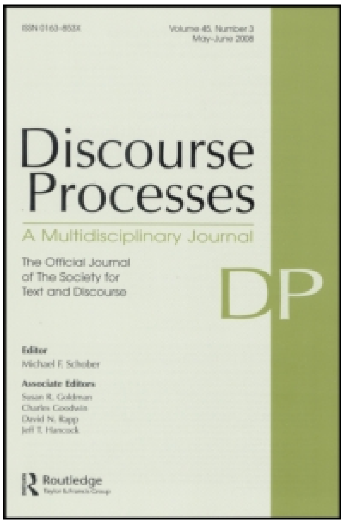

Discourse Processes
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653637

Modeling Semantic and Structural Knowledge in Web Navigation

Ion Juvina[a]; Herre van Oostendorp[b]
[a] Department of Psychology, Carnegie Mellon University, [b] Center for Content and Knowledge Engineering, Institute of Information and Computing Sciences, Utrecht University, The Netherlands

To cite this Article Juvina, Ion and van Oostendorp, Herre(2008) 'Modeling Semantic and Structural Knowledge in Web Navigation', Discourse Processes, 45: 4, 346 — 364
To link to this Article: DOI: 10.1080/01638530802145205
URL: http://dx.doi.org/10.1080/01638530802145205

Discourse Processes, 45:346–364, 2008

ISSN: 0163-853X print/1532-6950 online
DOI: 10.1080/01638530802145205

Routledge
Taylor & Francis Group

Modeling Semantic and Structural Knowledge in Web Navigation

Ion Juvina
Department of Psychology
Carnegie Mellon University

Herre van Oostendorp
Center for Content and Knowledge Engineering
Institute of Information and Computing Sciences
Utrecht University, The Netherlands

Research on cognitive modeling of information search and Web navigation emphasizes the importance of "information scent" (the relevance of semantic cues such as link labels and headings *to* a reader's goal; Pirolli & Card, 1999). This article shows that not only semantic but also structural knowledge is involved in navigating the Web (Juvina, 2006). This article argues for considering both semantic and structural knowledge in modelling Web navigation. A cognitive model is proposed that uses information scent to account for user's judgments of relevance (a semantic dimension) and "path adequacy" (the semantic similarity between a navigation path and a user's goal) to account for user's efficiency in traversing a Web structure (a structural dimension). Two empirical studies show that abilities to represent and manipulate spatial structures are complementary to semantic abilities in accounting for Web navigation performance.

The aim of this article is to argue for considering both semantic and structural knowledge in cognitive models of Web navigation. This section foregrounds our empirical and modeling arguments and places them within known theoretical frameworks.

Correspondence concerning this article should be addressed to Herre van Oostendorp, Center for Content and Knowledge Engineering, Institute of Information and Computing Sciences, Utrecht University, Padualaan 14, 3584 CH Utrecht, The Netherlands. E-mail: herre@cs.uu.nl

Several approaches to modeling Web navigation are inspired by the information foraging theory (Pirolli & Card, 1999). Cognitive models based on this theory assume that selections of users' actions are determined by utility assessments. Users assess meaning of proximal cues such as link labels and make predictions about distal related content. "Information scent" is a measure of this subjective assessment of how likely a proximal cue is to lead toward a desired distal content. In terms of Card et al. (2001), "information scent is the (imperfect) perception of the value, cost or access path of information sources obtained from proximal cues, such as WWW [World Wide Web] links. On a webpage information scent may be delivered by link descriptors, images, contextual clues, such as preceding headings, or by page arrangement" (p. 499). Although it is theoretically valuable, this definition is difficult to apply in user modeling or cognitive modeling. An operational definition has been considered necessary. According to this operational definition, information scent is the assessed semantic relevance of screen objects to users' goals (Kitajima, Blackmon, & Polson, 2000; Pirolli & Fu, 2003). Several cognitive models have been developed based on the concept of information scent (Blackmon, Polson, Kitajima, & Lewis, 2002; Kitajima et al., 2000; Miller & Remington, 2002; Pirolli & Fu, 2003). We focus here on a Comprehension-based Linked Model of Deliberate Search (CoLiDeS) developed by Kitajima et al.

CoLiDeS assumes that comprehension of texts and images is the core process underlying Web navigation. As its authors claim, CoLiDeS is inspired by the text comprehension theory of Kintsch (1998). The construction-integration theory of text comprehension (Kintsch, 1998) postulates a construction phase in which a mental representation is constructed from textual input, reader's goals and prior knowledge, and an integration phase that establishes coherence of the constructed representation via a spreading activation mechanism. CoLiDeS explains how users parse and comprehend the content of a Web page (analog of the textual input) and then select what action to perform next. This comprehension process is influenced by the user's goal and background knowledge. CoLiDeS compares the user's goal with link texts on Web pages and selects the link that best matches the user's goal. The selected link is clicked on and the process of judging link relevance (information scent), and selecting a link is repeated until the user's goal is attained or the user stops. The relatedness of screen objects to the user's goal (information scent) is measured based on three factors: semantic similarity, frequency, and literal matching. Semantic similarity is calculated based on co-occurrences between words and documents with the aid of a machine learning technique called latent semantic analysis (LSA; Landauer, Foltz, & Laham, 1998). Different LSA semantic spaces are used to model background knowledge for different user populations. Contents that literally match the user's goal and are relatively frequent are more likely to be selected.

CoLiDeS, as well as other models inspired by the concept of information scent, emphasizes the semantic dimension of Web navigation; that is, they assume that the process of relevance assessment is central to Web navigation, and they propose explanatory mechanisms and computational instruments to account for this process. However, several authors have mentioned the importance of users' spatial cognition for Web navigation (Czerwinski, van Dantzich, Robertson, & Hoffman, 1999; Tamborello & Byrne, 2005; Tavanti & Lind, 2001). In a study conducted by Farris, Jones, and Elgin (2002), participants had to explore a Web site and draw the Web site's information structure afterwards. Analysis of participants' drawings made the authors observe that the Web site's structure was not represented, but the conceptual relations between various information units were. Farris et al. concluded that participants' representations were nonspatial because the Web site structure was not accurately drawn and, instead, the drawings pictured conceptual relations. An alternative explanation could be that users represented the information space structure and not the Web site structure. The fact that participants' drawings were inspired by semantic relations does not imply that their mental representations were nonspatial. Participants were still able to draw these relations in spatial-like configurations. Most probably, users' representations of the information space being traversed are both spatial and semantic, and this would explain the correlations found between both semantic and spatial abilities, on one hand, and Web task performance, on the other hand (Chen, 2000; Gugerty, Treadaway, & Rubinstein, 2006; Neerincx, Pemberton, & Lindenberg, 1999).

If spatial cognition is indeed an important factor in Web navigation, then the concept of information scent is narrowly defined or insufficient. Models based on information scent tend to see each judgment of relevance in isolation and ignore the roles of context and history. They tend to abstract out the current position of the user within a particular page or across pages, the previously viewed information elements, and the viewing patterns such as order and frequency of (re)visiting particular information elements. Recent research shows that users' decisions are based not only on the assessed relevance of the currently available screen objects, but also on the relevance of objects that were encountered in earlier steps of the navigation session (Howes, Payne, & Richardson, 2002). Backtracking behavior is frequent (Cockburn & McKenzie, 2001) and is involved in the process of judging relevance. Wen (2003) coined the term "post-valued recall" (PVR) to refer to the interest a user may have in recalling information whose value is not recognized until some time after its initial retrieval. PVR occurs after the user has surfed enough to establish a context within which to judge the value of a Web page. CoLiDeS represents users' background knowledge as a static repository (i.e., a certain LSA semantic space). However, PVR would predict that a user's evaluation of an interface object (e.g., link text) is different at different stages of a navigation session, and the difference is made by the dynamic knowledge that is acquired during the interaction. A model that takes context and previous explorations into account has been proposed (Brumby, 2004), but it addresses only effects at a single Web page level.

To address these limitations of the existing models of Web navigation, we have made a number of amendments to CoLiDeS (Juvina, 2006; Juvina, van Oostendorp, Karbor, & Pauw, 2005). The altered model has been labeled CoLiDeS+ to indicate that it shares the main assumptions with the original model and is intended to be an augmented model.

CoLiDeS+ brings in the concept of "path adequacy" as a complement to the concept of information scent. Path adequacy is the semantic similarity between a navigation path and a user's goal. A navigation path is a succession of links that have been selected prior to a particular moment in a navigation session. Users are assumed to base their selections not only on goal relevance of incoming information but also on whether a candidate selection is consistent with past selections. Therefore, in CoLiDeS+, selecting a link on a specific Web page is a function of goal description, link description, and path description.

CoLiDeS+ also incorporates navigation strategies by maintaining a developing representation of the information space being navigated (navigation path) and checking at each step for impasses based on path adequacy. An impasse occurs when path adequacy does not increase after selecting a link, and it is a reason to switch the path. At this point, CoLiDeS+ reacts with a strategy that we called "next best," and it is to some extent similar with the opportunistic strategy of Miller and Remington (2002). Next best means that not only the link with the highest similarity to the user's goal is considered but also links with lower similarities provided that they contribute to an increase in path adequacy. Eventually, the options of backtracking one or more pages or going to index pages are available.

A short description of the algorithm used by CoLiDeS+, presented later, shows how the concept of path adequacy is considered in addition to link relevance (see also Figure 1; and see, for a more elaborate description, Juvina & van Oostendorp, 2006b):

- A task description is taken as input and assumed to be equivalent to the user's goal.
- A Web page is attended to, parsed in several areas, and a particular area is focused on (e.g., a menu).
- Menu entries are comprehended (based on how semantically similar to the user's goal they are), and one entry (the one that is most relevant to the user's goal) is selected and acted on (e.g., clicked on).
- A new page is attended to and, if the target information cannot be found, the cycle is reinitialized.

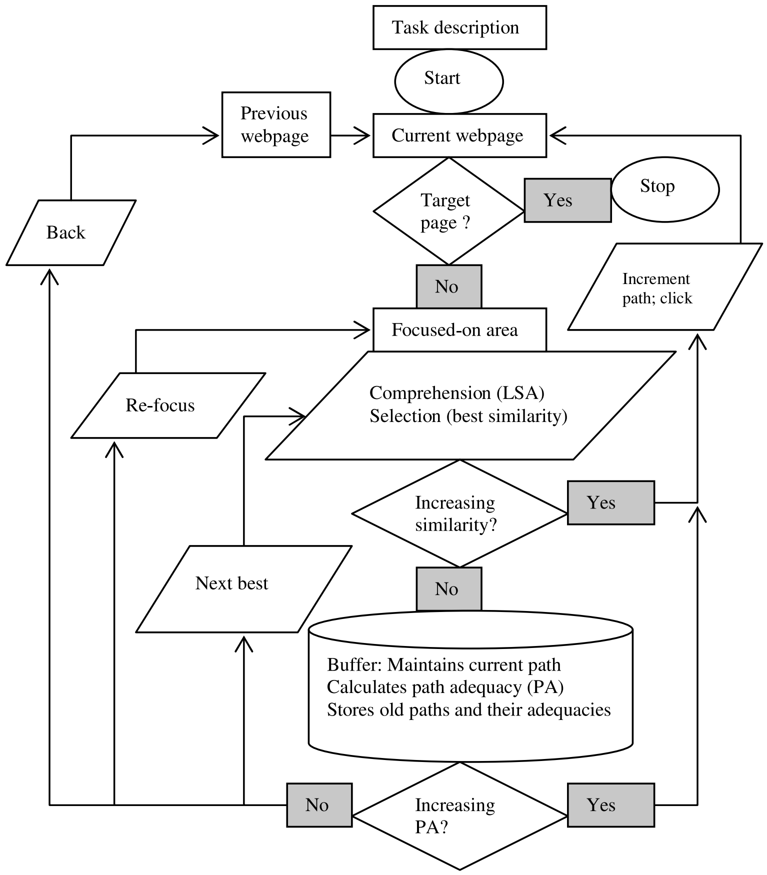

FIGURE 1 A diagram of the algorithm that implements the CoLiDeS+ model. *Note*. LSA = latent semantic analysis.

- The selected element is retained in a memory structure that maintains user's navigation paths.
- Starting with the second cycle, a navigation path is available, and the metric called path adequacy is computed. Selections of screen objects to be acted on are made if they contribute to an increase in path adequacy.
- Otherwise, an impasse is declared and dealt with by considering next-best options, changing the focused-on area and backtracking.
- The algorithm stops when the user decides that the current page contains the target information.

Based on its conditions of increasing link relevance (scent) and path adequacies, as well as its mechanisms of recovering from impasses, the algorithm is able to find the right path. (We return to this in a later section regarding the second study.) Based on this type of simulation, it is possible to determine at each step in the simulation process what is the model's successful path up to that moment and what are the model's unsuccessful trials (detours from the successful path).

The following section presents two empirical studies aimed at testing the validity of our arguments. A final section discusses the results of the empirical studies and concludes the article.

EMPIRICAL STUDIES

Two empirical studies were conducted with the aim of testing how relevant the structural–spatial knowledge is, besides the semantic knowledge, in modeling Web navigation behavior.

Study 1: Correlates of Web Task Performance

A correlational study has been conducted to test whether spatial and semantic knowledge is indeed relevant for Web task performance when controlling for possible confounding variables such as episodic memory and working memory capacity. Furthermore, it was tested whether the path adequacy measure described in the previous section captures the efficiency of traversing a knowledge space structure and is related to Web task performance.

Participants. The study was run with 30 participants in a single session. One half of the participants (7 women and 8 men) were registered as students in the Information Management Department of Twente University, and the other half (8 women and 7 men) were students in the Information Science Department of Utrecht University. Students were randomly selected out of the cataloges of both universities and invited to participate. Students who declined the invitation were replaced by making new random selections in the pool of registered students. The range of participants' age was 19 to 28 years, with a mean of 21.5 and a standard deviation of 2.29. Each participant received a compensation of 20€ at the end of the session.

Design. A within-subjects design was employed to investigate the correlations between a criterion—Web task performance—and a number of user characteristics as predictors—*spatial ability, domain knowledge, episodic memory,* and *working memory.* Metrics of Web navigation behavior and their correlations with user characteristics and task performance were used in the interpretation of the outcomes.

Materials and measurements. The task domain for this study was Web-assisted personal finance. Three real Web sites were used: financenter.com, thisismoney.co.uk, and amazon.com. Realistic tasks were constructed based on the collection of incidents published by Morrison, Pirolli, and Card (2001). The following is one of the tasks used in this study:

> Suppose you want to buy a car in 2 years. You have already saved $500. How much do you need to save on a monthly basis in order to make a down payment of $8000 for the car? Assume that the savings and tax rates are as listed. What is the most expensive car you can afford if you will be able to pay 40 monthly payments of at most $150 each after the down payment?

There were six tasks of this kind. Task performance was recorded by the participants during the navigation session on a dedicated form and coded afterwards by two experts based on a coding scheme specified in advance. Performance of each task was scored on a 4-point scale based on correctness and completeness of answers. The scale ranged from 0 (*not attempted task*), 1 (*erroneous answer*), 2 (*partly correct answer*), 3 (*correct answer*), to 4 (*correct and complete answer*). The proportion of initial agreements between the two experts was very high (0.98). Disagreements were resolved by discussion between the two experts. The internal consistency between the performance scores on the six tasks was moderate in size (Cronbach's $\alpha = 0.67$). The means and standard deviations for performance on each of the six tasks are presented in Table 1. The distributions of performance scores for the six tasks were reasonably normal. The mean of the six scores per participant was calculated and used as the task performance score in further analyses.

TABLE 1
Means and Standard Deviations for Performance Scores on Each of the Six Tasks

Task	*M*	*SD*
1	1.5	1.8
2	2.0	1.7
3	2.7	1.7
4	2.4	1.8
5	2.8	1.8
6	0.9	1.5

Domain knowledge was measured as self-assessed knowledge and interest in the field of Web-assisted personal finance on 5-point Likert scales with the following items:

- "Have you ever used a personal finance Web site (Yahoo Finance, MSN Money, etc.)?"
- "Have you ever used Internet banking services?"
- "Have you ever done transactions via the Internet (buying, selling, etc.)?"
- "How likely are you to use the Internet for your personal finances in the future?"

Measures of spatial ability, episodic memory, and working memory were provided by TNO Human Factors Institute (Neerincx et al., 1999).

- The spatial ability measure used the classical mental rotation task (Shepard & Metzler, 1971), and the spatial ability score was the number of correct solutions obtained by rotating three-dimensional objects (correct matches between objects and their rotated equivalents).
- The episodic memory measure presented three lists of 60 images each; the participants had to name the images in the first two lists; between Lists 2 and 3 there was a distraction task; List 3 contained images that were presented before in Lists 1 and 2 together with new images; the participants had to recognize the images that were presented in List 1.
- The working memory measure used a reading span task (Daneman & Carpenter, 1980; Linderholm, 2002): Participants were presented with series of sentences, the size of series increasing progressively from two to seven sentences; the participants were asked to read the sentences aloud and try to understand their content; after each series, the participants were asked to recall the last word of each sentence in that particular series; the working memory score was calculated based on correctness of recalls.

Metrics of Web navigation behavior were computed based on Web-logging data collected during the navigation sessions:

- *The amount of revisits* is calculated as the average frequency of revisits to a page.
- *Back button usage* indicates the percentage of back button clicks among all recorded navigation actions.
- *Fan degree* represents the average number of links followed per page (Rauterberg, 1996).
- *Path adequacy* was calculated based on the Web-logging data and the task descriptions that participants were provided with at the beginning of a

task. A navigation path was considered to be a concatenation of semantic objects that the user has encountered in her or his way. As semantic objects, one can consider link anchors, page titles, page contents, URLs, clickable icons, banners and images, and so forth. The following is a navigation path composed of page titles. If the user visited the pages titled, "Should I finance or pay cash for a vehicle? Calculators," "How much will my vehicle payments be? Calculators," "Glossary," and "What vehicle can I afford? Calculators," his or her navigation path was represented as a string of all words in these titles: ⟨*should, I, finance, or, pay, cash, for, a, vehicle, calculators, how, much, will, my, vehicle, payments, be, calculators, glossary, what, vehicle, can, I, afford, calculators*⟩. Then, path adequacy was determined as a coefficient of semantic similarity between a navigation path and a task description. The navigation path for a task included all titles of the pages visited during the execution of that task. LSA was used to measure the semantic similarity between navigation paths and task descriptions, on basis of the "general reading" semantic space available at http://lsa.colorado.edu. The degree of semantic similarity is measured by the cosine value, analogous to a correlation coefficient.

Procedure. The study was conducted in one session. The first part of the study was dedicated to questionnaires and cognitive tests aimed at measuring user characteristics. The second part consisted in execution of six Web navigation tasks. Participants were informed, and they agreed to have their navigation behavior logged with the aid of unobtrusive Web-logging software. The navigation sessions were restricted to a maximum of 40 min.

Results. Stepwise multiple linear regression was employed to check the influence of user characteristics on Web task performance as the criterion variable. The initial regression model included as predictors all user characteristics described earlier (domain knowledge, spatial ability, episodic memory, and working memory). The critical threshold of significance was set to $\alpha = 0.05$. Nonsignificant predictors were eliminated by the stepwise regression procedure. The final regression model retained only two significant predictors of Web task performance: spatial ability and domain knowledge. Table 2 presents the R^2 for

TABLE 2
Predictions of Web Task Performance Based on User Characteristics

Criterion	R^2	*Predictors*	β
Web task performance	0.39	Spatial ability	0.496
		Domain knowledge	0.385

this final regression model and the standardized regression coefficients (beta) for the two predictors. The effect size for regression was calculated with the following formula: $ES^2 = R^2/(1 - R^2)$; and it returned a value of 0.64—that is, a large effect size according to Cohen (1992).

To interpret the significant influence of spatial ability on Web task performance, bivariate correlations were computed between metrics of navigation behavior, on one hand, and spatial ability, domain knowledge, and task performance, on the other hand. More sophisticated techniques such as structural equation modeling were not applicable due to the relatively small size of the sample. Only significant correlations ($\alpha = 0.05$) are reported here:

- Spatial ability negatively correlates with metrics involving revisitation:
  - *Revisits:* $r(28) = -0.44$.
  - *Back button use:* $r(28) = -0.43$.
  - *Fan degree:* $r(28) = -0.43$.
- Path adequacy positively correlates with spatial ability, $r(28) = 0.36$; and task performance, $r(28) = 0.47$.

Discussion. Based on the results of this study, it appears that both spatial and semantic cognitive mechanisms are critical in adequately performing Web navigation tasks. The stepwise method of regression analysis made possible to rule out factors that were confounded with one another. For example, the influence of working memory on hypertext navigation as reported by Tucker and Warr (1996) might have not appeared as significant if spatial ability was included as a predictor in their model. Our results show that spatial ability is more important for Web navigation performance than working memory capacity.

The distinction between spatial ability and working memory capacity is not clear-cut from a theoretical point of view. As a matter of fact, one component of working memory—visuospatial sketchpad—is believed to be involved in storing and manipulating complex spatial patterns (Baddeley, 2000). This component is most probably involved in mental rotation—the test we have used to measure individual differences in spatial ability. To assess individual differences in working memory, we have used the reading span test, which addresses another component of working memory—the articulatory loop. Thus, we have taken into consideration both spatial and nonspatial aspects of working memory, and it turned out that the spatial ones are relatively more important for Web navigation than the nonspatial ones.

Correlations between spatial ability and navigation metrics helped us with understanding the relation between spatial ability and Web task performance. Spatial ability is negatively correlated with metrics involving revisitation. Conceivably, users with high spatial abilities are faster or better at representing the information space structure and, consequently, they need less revisitation.

Conversely, low spatial ability users need a high amount of revisitation and exploration to accurately represent the information space structure and make correct choices. Supposedly, common cognitive processes are used to represent and operate on an information space and to mentally rotate objects in a three-dimensional space.

The positive correlations between path adequacy, on one hand, and spatial ability and task performance, on the other hand, suggest that path adequacy might capture reasonably well the ability of users to represent and operate on an information space. The positive correlation between path adequacy and task performance underscores the relevance of contextual information for cognitive modeling of Web navigation.

The results of this study confirm the limitation of the existing cognitive models of Web navigation (Blackmon et al., 2002; Kitajima et al., 2000; Pirolli & Fu, 2003) mentioned in the introduction of this article. These models almost completely ignore the spatial dimension, and treat solely the semantic dimension of Web navigation (information scent). The model we have proposed (CoLiDeS+) has the potential to overcome this limitation because it uses path adequacy to indicate user's efficiency in traversing an information space.

Study 2: CoLiDeS+ and Spatial Ability

This study was intended to test the validity and value of CoLiDeS+. It was hypothesized that CoLiDeS+ would be able to simulate real user's navigation behavior, and the navigation support generated based on simulations of successful paths would have a positive influence on user's navigation behavior and task performance. This facilitation was expected because the model-generated support guides the reader through relevant contextual information (see also Juvina & van Oostendorp, 2005). This positive influence was expected to be bigger for users with a deficit of spatial ability because CoLiDeS+ took over the job of representing the information structure and remembering past selections.

Participants. Participants were 29 undergraduate and graduate students at Utrecht University (13 men and 16 women). They were recruited through on-campus advertisements. The ad required Web experience, but students from information and computing sciences were not allowed to participate. This was meant to avoid recruiting an over-specialized sample. Participants' age was not recorded for this study, but it was within approximately the same range as in the first experiment.

Design. The 29 participants were randomly assigned to two conditions: a *control* condition in which 15 participants had to perform as many tasks as possible from a set of six in 45 min, and a *support* condition in which 14 participants did the same while receiving model-generated navigation support (suggestions).

Materials and measurements. Six realistic Web tasks were constructed based on the collection of cases gathered by Morrison et al. (2001) and using the experience of Kitajima et al. (2000) concerning the size and elaboration of task descriptions. Each task had an associated Web site. The following is one of the six tasks used in this experiment:

> Motivation/context: You want to buy a new car but you don't have enough money. The Internet has made getting a loan much easier because it provides you with the resources you need to find the right loan, the approval process is quicker, there is less documentation that needs to be filled out, and you can do it in the privacy of your own home. You want to shop around for an auto loan lender, find an attractive interest rate, and find out how much your monthly payment will be. Also you wonder what will happen if you become unable to make your payments due to various conditions (sickness, etc.).

- Given facts: You cannot afford to pay more than 180 pounds per month for more than 48 months.
- Use http://www.alliance-leicester.co.uk/ to:
  - Calculate how much you can loan.
  - Calculate how much your monthly payment will be.
  - Look for one way to handle situations when you cannot pay.

The six tasks were simulated with CoLiDeS+ prior to the actual navigation sessions. The results of simulations were successful paths (i.e., successions of links leading to the target pages) and "dead-ends" (i.e., pages that are not linked with the target pages, making it necessary for the user to backtrack). The navigation path is here composed of link labels instead of page titles as in the previous experiment, and path adequacy is calculated at each step of a navigation session. Based on these results of the simulations, two types of suggestions were generated: link suggestions—when a link contained in a successful path was visible on the screen, the user received the suggestion "Click on ⟨link label⟩"; and path switch suggestions—when a dead-end page was downloaded, the user got the suggestion "Go back." Both the "click-on" and "go-back" suggestions were taken from the model simulation. Whenever a link leading to the target pages was visible on the screen, it was suggested; whenever the user visited a page that had been found by the model to be a dead-end, a go-back suggestion would be issued. The suggestions were provided in the auditory modality.

Solutions to tasks were reported on paper and evaluated afterwards for correctness. The answers were scored for correctness and completeness on a 4-point scale for each task ranging from 0 (*not attempted task*), 1 (*erroneous* *answer*), 2 (*partly correct answer*), 3 (*correct answer*), to 4 (*correct and complete answer*). The total correctness score for the six tasks ranges from 0 to 24. Calculating a general correctness score across the six tasks was justified by reasonably good consistency (Cronbach's $\alpha = 0.63$) and by the need to build a well-formed variable (continuous and normally distributed). The average duration of tasks per participant was calculated by dividing the total navigation time (45 min) to the number of tasks attempted. An overall estimate of task performance was calculated by dividing the total correctness score to the average duration of tasks. The natural logarithm of this ratio was taken to correct for a skewed distribution.

Participants' spatial ability was tested with a mental rotation task, the same task as described in the first experiment. Navigation actions of participants were automatically recorded with Web-logging software. Navigation sessions were recorded on video using the software tools "The Observer" and "Camtasia Studio."

Procedure. In the first part of the study, the mental rotation task was administrated to all participants. The second part of the study consisted of administration of the six navigation tasks. Participants in the *support* condition were instructed in advance that suggestions were generated by a robot, they were meant to help with task execution, and they were not mandatory: Participants could follow them or not at their discretion. The experimenter provided suggestions in a standard way. At the end, participants were debriefed and received 10€ for their participation.

Results. The first outcome was that CoLiDeS+ was able to generate successful paths and dead ends, although the way it navigated the Web sites was not as similar to real users as suggested by Kitajima et al. (2000). It made extensive use of next-best trials, refocusing, and backtracking. The number of steps to solutions was higher than the actual users took. However, even when the model took different decisions than the actual users, the correct paths and dead ends were correctly identified, due to the mechanisms of solving impasses. This result supports the decision to use the model's outcomes (successful paths and dead ends) in generating navigation support.

By making use of path adequacy, CoLiDeS+ simulated users' behavior slightly better than CoLiDeS. To show this, we have analyzed 10 user sessions recorded on video, randomly selected from the control condition. Only data from the control condition were used because the support condition was already manipulated based on CoLiDeS+ simulations. We looked at what selections users made and compared that with what selections CoLiDeS and CoLiDeS+ would have made based on the same set of options.

First, for each user, for each task, for each visited Web page, we recorded the (last) area of the page the user focused on, options available in this area, and which option was selected by the user. This latter information was recorded as the user selection. Second, the task description and the same options were entered in the LSA algorithm to compute semantic similarities between the task description and each option (scent). The option with the highest similarity was chosen and recorded as the CoLiDeS selection. This CoLiDeS selection can match or not match with what a user did. Third, the CoLiDeS+ selection was calculated on the same data by adding the path to each option and selecting the option with the highest path adequacy (goal similarity of a path; see the CoLiDeS+ algorithm presented earlier in this article). Also, here user selections can match or not match CoLiDeS+ selections. The overall results are summarized in Table 3. It shows that overall CoLiDeS+ simulates users slightly better than CoLiDeS ($\chi^2 = 3.52$; $p = .06$; marginally significant).

Providing navigation support made a significant difference in users' navigation behavior and task performance. The number of navigation steps was lower in the support condition than in the control condition, $t(27) = 3.86$, $p = .001$. It took an average of 30 steps ($SD = 9$) to execute a task in the control condition and only 19 steps ($SD = 5$) in the support condition. The average duration of tasks was shorter in the support condition than in the control condition, $t(27) = 2.16, p = .04$. It took an average of 10.26 min ($SD = 2.76$) to complete a task in the support condition and 12.49 min ($SD = 2.77$) in the control condition. Task performance was significantly higher in the support condition ($M = 1.16, SD = 0.48$) than in the control condition ($M = 0.68, SD = 0.69$), $t(27) = 2.16$, $p = .04$. The score on task performance (natural logarithm of correctness divided by time) ranges from minimum $= -1.12$ to maximum $= 2.07$, and the distribution is sufficiently normal.

As expected, the correlation between spatial ability and task performance was significant ($\alpha = 0.05$) for the control condition, $r(27) = 0.64$, $p = .01$; and not significant for the support condition, $r(27) = 0.15, p = .60$. The difference between the two r coefficients does not reach statistical significance

TABLE 3
CoLiDeS+ Makes the Same Selection as the User in 54.9% of Cases, Simulating Users Slightly Better than CoLiDeS

Total Valid Selections	*275*	*%*
CoLiDeS selection matches user selection	129	46.9
CoLiDeS+ selection matches user selection	151	54.9

($z = 1.45$; $p = .07$) because of the small number of cases in the two groups (14 and 15, respectively).

Participants were divided in two groups with high and low spatial abilities (the median of test scores was taken as a cutting point). The difference in task performance induced by navigation support was separately checked for low and high spatial ability participants. Results based on the means are depicted in Figure 2. One can see that the difference induced by navigation support is bigger for participants with low spatial ability—$t(10.9) = 2.27$, $p = .04$—than for participants with high spatial ability—$t(13.1) = 0.73$, $p = .48$. The interaction between condition and levels of spatial ability did not reach statistical significance—$F(1, 25) = 1.45$; $MSE = 0.5$; $p = .24$—because of the small number of participants in the two groups. Partial eta squared (a measure that does not depend on the number of participants) for this interaction is 0.055, corresponding to a medium effect size.

Discussion. CoLiDeS+ simulated users' task execution better than CoLiDeS based on information scent and path adequacy—a measure of contextual information involved in user navigation. When simulated solutions were offered as navigation support, users performed better on the given tasks. It seems that the offered navigation support prevented users from spending time and cognitive resources with those navigating actions that are not directly effective but are usually employed to accurately represent the information structure. Users

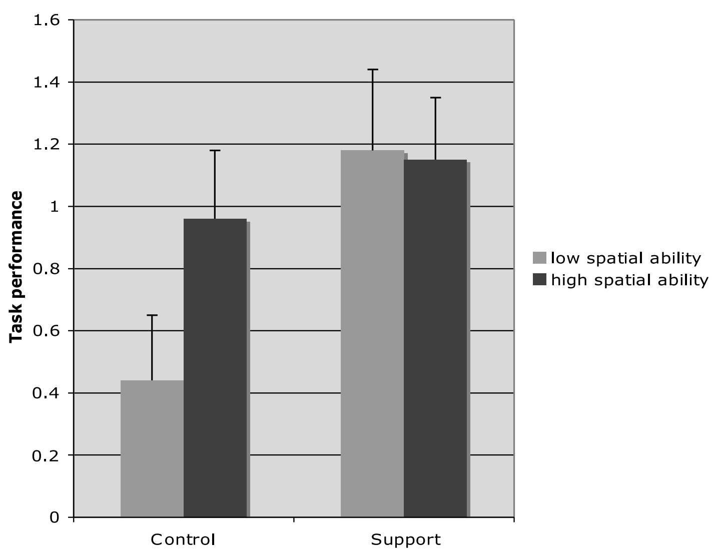

FIGURE 2 The effect of providing navigation support on (mean) task performance for users with low and high spatial ability.

engage in apparently useless navigation actions to get acquainted with the context of a particular piece of information, which is eventually useful in judging the value of that particular information. By making use of path adequacy, CoLiDeS+ gives an account for this type of behavior. Also, when the suggestions are given in another modality, with graphical suggestions, similar positive effects are found (Juvina et al., 2005).

The correlation between spatial ability and task performance indicates that users' ability to mentally represent and manipulate information spaces is critical for Web navigation tasks. When provided with navigation support, users with low spatial abilities had a higher performance increase than users with high spatial abilities. It follows that users with low spatial abilities are probably less able to represent the information space, and this is why they benefit considerably when the cognitive model is doing this job for them. We consider this result to be additional proof that CoLiDeS+ gives an account not only for the process of assessing relevance of link labels to users' goals but also for the ability of users to represent and manipulate an information space.

GENERAL DISCUSSION AND CONCLUSION

The two studies previously presented demonstrate a clear relation between spatial ability and Web task performance. More specifically, we have assumed that spatial ability was related to representing the information space and operating on it. In other words, we have related spatial ability to navigation across pages and not with the process of visual search within a page. This assumption was based on the correlation found between spatial ability and navigation metrics such as the revisits, back button use, and fan degree (discussed earlier, in Study 1). However, the alternative explanation could not be ruled out completely. In normal Web use, there are quite a few visuospatial features that are not related to navigation across pages (e.g., page layout, color, size, highlights), and that could determine the correlation between spatial ability and task performance. A complete proof would be to eliminate these features and show that the effect of spatial ability on task performance remains. A situation like this (i.e., lack of visuospatial features) exists in real Web use of visually impaired persons. They often use screen readers (i.e., tools that read out loud the content and options on Web pages in a nonselective way; Juvina & van Oostendorp, 2006a). In the case of using the Web by visually impaired persons via screen readers, an essential feature of Web navigation—handling a hyperstructure of linked documents—is isolated from other aspects of screen-based interaction. Visual features such as page layout, font color, and size are absent.

We have conducted a study in which visual impairment was simulated (Juvina, 2006). The screen was turned off, and users had to make use of a screen reader to navigate and complete a series of Web tasks. Due to space limitations, we can only report here the key result of this study: The significant correlation between spatial ability and Web task performance was once more replicated ($r_{14} = 0.564$). Thus, spatial ability is indeed related to navigation across pages and not to visual search within a page.

In conclusion, structural knowledge (i.e., representing the information space structure and operating on this representation) is critical for navigating the Web. Theories and models of Web navigation need to account for both semantic and structural dimensions of Web navigation behavior. In particular, our findings contribute to refining the information foraging theory (Pirolli & Card, 1999). We suggested that, to make adequate assessments of relevance, users need to be able to efficiently represent and operate on an information space.

Our empirical results showing the complementary roles of semantic and spatial abilities in Web task performance are in accord with previous findings (Chen, 2000; Gugerty et al., 2006; Neerincx et al., 1999). Although the role of spatial cognition in Web navigation is empirically recognized, the computational models of Web navigation tend to abstract out the spatial dimension and overemphasize the semantic dimension (information scent). For instance, CoLiDeS (the source of CoLiDeS+) focuses on modeling a local top-down dimension from the construction-integration theory of text comprehension: assessing text elements in the view of the user's goal based on user's background knowledge. This assessment of goal relevance is modeled by information scent. However, human comprehension is regarded as cycles of top-down and bottom-up processes (Kintsch, 2005). Construction is local (context-free), whereas integration is global (context-dependent). By adding contextual information (a global and bottom-up feature), we made the model more consistent with its theoretical grounds. Path adequacy indicates how adequately the information space has been represented and traversed up to the current point. As our correlational study suggests, this adequacy depends on users' spatial abilities. When the goal relevancy of the current option is not clear-cut, the current option is included in the path, and path adequacy is recalculated. An increase in path adequacy is an indication that the current option is goal relevant in the context of past selections. Thus, CoLiDeS+ takes a step toward modeling the role of spatial abilities in hypertext comprehension by including path adequacy next to information scent.

ACKNOWLEDGMENTS

The following undergraduate and graduate students helped with data collection: Poyan Karbor, Brian Pauw, Arnaud Lek, Martijn Abbing, and Richard van Yperen. We thank the editor and the two anonymous reviewers for their comments on an earlier draft of this article.

REFERENCES

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423.

Blackmon, M. H., Polson, P. G., Kitajima, M., & Lewis, C. (2002). Cognitive walkthrough for the Web. In D. Wixon (Ed.), *Proceedings of the SIGCHI conference on human factors in computing systems 2002* (pp. 463–470). Minneapolis, MN: ACM Press.

Brumby, D. P. (2004). A model of single-page Web-search: The effect of interdependence on link assessment. In *Proceedings of the 6th international conference on Cognitive Modeling* (pp. 402–403). Mahwah, NJ: Erlbaum.

Card, S. K., Pirolli, P., Van der Wege, M., Morrison, J. B., Reeder, R. W., Schraedley, P. K., et al. (2001). Information scent as a driver of Web behavior graphs: Results of a protocol analysis method for Web usability. In D. Wixon (Ed.), *Proceedings of the SIGCHI conference on human factors in computing systems 2001* (pp. 498–505). New York: ACM Press.

Chen, C. (2000). Individual differences in a spatial-semantic virtual environment. *Journal of the American Society for Information Science, 51,* 529–542.

Cockburn, A., & McKenzie, B. (2001). What do Web users do? An empirical analysis of Web use. *International Journal of Human–Computer Studies, 54,* 903–922.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Czerwinski, M., van Dantzich, M., Robertson, G. G., & Hoffman, H. (1999). The contribution of thumbnail image, mouse-over text and spatial location memory to Web page retrieval in 3D. In M. A. Sasse and C. Johnson (Eds.), *Human-Computer Interaction–INTERACT '99* (pp. 163–170). Edinburgh, England: IOS Press.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

Farris, J. S., Jones, K. S., & Elgin, P. D. (2002). Users' schemata of hypermedia: What is so "spatial" about a Website? *Interacting With Computers, 14,* 487–502.

Gugerty, L., Treadaway, C., & Rubinstein, J. S. (2006). Individual differences in Internet search outcomes and processes. In R. Grinter, T. Rodden, P. Aoki, et al. (Eds.), *Proceedings of the SIGCHI conference on human factors in computing systems 2006* (pp. 815–820). New York: ACM Press.

Howes, A., Payne, S. J., & Richardson, J. (2002). An instance-based model of the effect of previous choices on the control of interactive search. In W. D. Gray and D. Schunn (Eds.), *Proceedings of the 24th annual meeting of the Cognitive Science Society* (pp. 476–481). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Juvina, I. (2006). *Development of a cognitive model for navigating on the Web*. Unpublished doctoral dissertation, Utrecht University, The Netherlands.

Juvina, I., & van Oostendorp, H. (2005). Cognitive model working alongside the user. In T. McEwan, J. Guliksen, and D. Benyon (Eds.), *People and Computers XIX–The bigger picture. Human Computer Interaction 2005* (pp. 409–420). London: Springer.

Juvina, I., & van Oostendorp, H. (2006a). Enhancing Internet experience of visually impaired persons by means of dynamic highlighting and selective reading. *Information Design Journal, 14,* 71–81.

Juvina, I., & van Oostendorp, H. (2006b). Individual differences and behavioral metrics involved in modeling Web navigation. *Universal Access in Information Society, 4,* 258–269.

Juvina, I., van Oostendorp, H., Karbor, P., & Pauw, B. (2005). Toward modeling contextual information in Web navigation. In B. G. Bara, L. Barsalon, and M. Bucciarelli (Eds.), *27th meeting of the Cognitive Society* (pp. 1078–1083). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Kintsch, W. (2005). An overview of top-down and bottom-up effects in comprehension: The CI perspective. *Discourse Processes, 39*, 125–129.

Kitajima, M., Blackmon, M. H., & Polson, P. G. (2000). A comprehension-based model of Web navigation and its application to Web usability analysis. In S. McDonald, Y. Waern, and G. Cockton (Eds.), *People and computers XIV* (pp. 357–373). New York: Springer.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes, 25,* 259–284.

Linderholm, T. (2002). Predictive inference generation as a function of working memory capacity and causal text constraints. *Discourse Processes, 34*, 259–280.

Miller, C. S., & Remington, R. W. (2002). Effects of structure and label ambiguity on information navigation. In D. Wixon (Ed.), *Proceedings of the SIGCHI conference on human factors in computing systems 2002* (pp. 630–631). New York: ACM Press.

Morrison, J. B., Pirolli, P., & Card, S. K. (2001). A taxonomic analysis of what World Wide Web activities significantly impact people's decisions and actions. In J. Jacko and A. Sears (Eds.), *Proceedings of the SIGCHI conference on the human factors in computing systems 2001* (pp. 163–164). New York: ACM Press.

Neerincx, M. A., Pemberton, S., & Lindenberg, J. (1999). *U-WISH Web usability: Methods, guidelines and support interfaces*. Soesterberg, The Netherlands: TNO Human Factors Research Institute.

Pirolli, P., & Card, S. K. (1999). Information foraging. *Psychological Review, 106,* 643–675.

Pirolli, P., & Fu, W.-T. (2003). SNIF-ACT: A model of information foraging on the World Wide Web. In P. Brusilovsky, A. Corbett, and F. de Rosis (Eds.), *Proceedings of the 9th international conference on User Modeling* (pp. 45–54). Berlin: Springer-Verlag.

Rauterberg, M. (1996). How to measure cognitive complexity in human–computer interaction. In R. Trappl (Ed.), *Cybernetics and Systems 1996. Proceedings of the 13th European meeting on Cybernetics and Systems Research* (pp. 815–820). Vienna, Austria: Austrian Society for Cybernetic Studies.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171,* 701–703.

Tamborello, F. P. I., & Byrne, M. D. (2005). Information search: The intersection of visual and semantic space. In W. Kellogg, S. Zhai, G. van der Veer, and C. Gale (Eds.), *Proceedings of the SIGCHI conference on the human factors in computing systems 2005* (pp. 1821–1824). New York: ACM Press.

Tavanti, M., & Lind, M. (2001). 2D vs. 3D. Implications on spatial memory. In K. Andrews, S. Roth, and P. Chung Wong (Eds.), *Proceedings of IEEE's annual conference on Information Visualization* (pp. 139–148). Washington, DC: IEEE Computer Society.

Tucker, P., & Warr, P. (1996). Intelligence, elementary cognitive components, and cognitive styles as predictors of complex task performance. *Personality and Individual Differences, 21,* 91–102.

Wen, J. (2003). Post-valued recall Web pages: User disorientation hits the big time. *IT & Society, 1,* 184–194.