Drop-Down Democracy: 
Internet Portal Design Influences Voters’ Search Strategies

Scott P. Robertson  Christine E. Wania  George Abraham  S. Joon Park  
University of Hawaii  Drexel University  Drexel University  Drexel University  
Info. & Comp. Sci.  iSchool  iSchool  iSchool  

scott.robertson@hawaii.edu, {christine.wania, george.abraham}@ischool.drexel.edu

Abstract
In this study we report how different interfaces for searching the internet changed the strategies and information opportunities of voters as they searched for political information. A drop-down interface that helped users define specific queries was compared to a traditional typed-query interface. The drop-down interface resulted in a broader set of information opportunities. Voters using the drop-down interface relied less on candidate websites, visited a broader sampling of website types, and examined issues more closely. Portal design heavily influences search strategies and information opportunities.

1. Introduction

1.1 Politics and the Internet
As the internet becomes more important in political campaigns, it is essential to understand how users search for information and inform themselves in anticipation of voting. In 2006, 60 million people in the United States (31% of the adult population) used the internet for some political purpose [1]. “Campaign internet users,” as the Pew Internet and American Life Project labeled them [1], used the internet to find information about candidates, talk to others about political issues, and post and forward their own and other people’s political commentary. An increasing number of campaign internet users are watching political and audio recordings on the internet and they are being pulled away from traditional sources of information such as television and radio.

Table 1 shows the percentage of campaign internet users who reported getting information from various internet sources. Politicians and political organizations are responding by increasing their presence on the internet and using more diverse internet resources such as blogs and virtual meeting spaces.

<table>
<thead>
<tr>
<th>Source</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>News portals</td>
<td>60%</td>
</tr>
<tr>
<td>TV network websites</td>
<td>60%</td>
</tr>
<tr>
<td>Local news organizations’ websites</td>
<td>48%</td>
</tr>
<tr>
<td>Websites of major national newspapers</td>
<td>31%</td>
</tr>
<tr>
<td>State or local government websites</td>
<td>28%</td>
</tr>
<tr>
<td>Issue-oriented websites</td>
<td>24%</td>
</tr>
<tr>
<td>Blogs</td>
<td>20%</td>
</tr>
<tr>
<td>International news orgs’ websites</td>
<td>20%</td>
</tr>
<tr>
<td>Websites created by candidates</td>
<td>20%</td>
</tr>
<tr>
<td>News satire websites</td>
<td>19%</td>
</tr>
<tr>
<td>Websites of radio news organizations</td>
<td>19%</td>
</tr>
<tr>
<td>Websites of alternative news orgs</td>
<td>10%</td>
</tr>
<tr>
<td>Email listervs</td>
<td>10%</td>
</tr>
</tbody>
</table>

Table 1. Sources of information for campaign internet users (from [1]).

It is not clear how campaign internet users find the information sources that they eventually use, but it is reasonable to assume that, like most internet users, they take advantage of search tools such as Google or Yahoo. As voters move to the internet, they have to figure out how to construct good queries and understand how to browse and filter query result lists. In our study, we examined how people in a mock voting situation approach query construction, result list consideration, and website browsing. We tested a novel front-end to a search tool that was designed to help users formulate better queries by taking advantage of recognition memory and forcing more specific searches.

1.2 Searching the Web
Large scale studies of search tool users have shown that people are generally not sophisticated
searchers. Spink, Wolfram, Jansen, & Saracevic [2] studied over one million queries by users of Excite and found that they tended to create queries with an average of only 2.16 terms. A study of over one billion AltaVista queries [3] similarly found that searchers created short queries (2.35 terms on average) and rarely modified their queries.

Query length and complexity, however, have been shown in a variety of studies to vary with searcher experience and domain expertise [4][5][6]. Holscher and Strube [7] found that search experts created longer queries than average users (3.64 terms versus 1.66 terms, respectively). In a second study they showed that experts in the search domain used shorter queries than domain novices (1.96 versus 2.96 terms, respectively) and conjectured that domain experts can be more focused and accurate.

Once users formulate a query, they are faced with the task of browsing the results. There are many browsing and information seeking strategies that users employ [8][9], and they vary depending on how directed searchers are [10], how much they integrate browsing and searching [11][12], how comprehensive they wish to be in their understanding of the domain [13][14], their intended use of the information [15], and even how the information influences searchers’ affective state [16].

1.3 Political Information Seeking

Voters have some special information needs and they are exploring a unique information space [17][18][19]. Voters are typically in “decision-making mode,” and some models of voter information browsing stress this aspect by suggesting that voters are updating checklist schemas about candidates (“online model”) as opposed to trying to learn all they can about the candidates [27]. Voters are faced with juggling many different types of information such as advocacy and issue-oriented information, persuasive and argumentative content, opinion, news and other media [20].

There have been a number of studies that examine voters’ political information seeking behaviors [21] or that ask voters questions about their political information seeking behaviors on the web [23], although these studies do not look directly at voters searching freely for information on the internet. Redlawsk and Lau [17][18][19] used a dynamic information board in which information flows over time in order to simulate the random and haphazard encounters with information that many voters experience as a campaign progresses. While this is an important type of information experience for voters, as they turn more and more to the internet they will become more in charge of their information exposure and they will be exposed to many different types of information that may not have been available in the past [1]. Researchers need to understand how voters will seek and filter information, how they utilize information that they obtain actively, and what they remember from such information encounters.

1.4 Drop-Down Recognition Searching

In several studies, we have been examining how voters search the internet for information about how to vote [24][25][26]. We are interested in how various information technologies might fit into different aspects of voters’ decision-making processes and how a thorough understanding of voter decision-making processes could guide the design of voter portals [25]. New candidates and ballot issues present voters with the problem of searching in a domain where their knowledge is limited. In typed-query search environments like Google,
Yahoo!, and most other search tools, voters must generate queries based on their own knowledge of the ballot and election issues.

In this study, we were interested in examining whether the search tool itself could help users to formulate more targeted searches that produce more relevant results. Following from an earlier study in which ballot items served a dual purpose of reminding voters of the candidates and issues and serving as navigation aids [25], we designed an interface in which ballot items appeared in a drop down list and, when selected, automatically initiated internet searches (Figure 1). Our interface also included a list of 24 issues in another drop-down list (Figure 2) and the option of viewing Web Pages or News in another drop-down list. Users could combine the three boxes to create specific searches such as web pages on “Dianne Feinstein and taxes.” In the background the search tool also added more specific terms to the search (e.g. “Senate and California” in the example above).

We were interested in examining how the recognition-based, drop-down interface might change users’ search strategies. We hypothesized that users might perform more searches and more issue-based searches if they were prompted by the drop-down menus.

Following Lodge [27, 28] and Huang & Price [15], we also contrasted conditions in which subjects were told that they would actually vote with conditions in which subjects were told that they needed to “learn as much as possible” about candidates in order to interview them later. Voting is often associated with “impression formation” tasks and is considered to be less cognitively challenging than an explicit learning task. Huang & Price [15] showed that subjects forming impressions were less likely to do within-candidate searching.

2. Method

2.1 Participants
Twenty-nine participants were recruited, using information flyers, from areas around Drexel University in Philadelphia, PA. Data was collected from August 10, 2006 to November 2, 2006. Each participant was paid $35 for their time. The participants ranged in age from 18 to 58, with a mean age of 29 years old. Three participants identified themselves as Republicans, 11 as Democrats, 8 as independents, 2 as socialist, and 3 as “other.” Seventeen participants had a four year degree, 5 had a Graduate degree, 1 had a Doctorate, 1 had a two year degree, and 1 had a high school degree.

2.2 Procedure
The independent variables were Search Interface (Drop-Down interface versus Traditional Query interface) and Information Task (Voting versus Non-Voting). Participants were randomly assigned to one of the four groups created by the 2x2 pairing of the independent variables. All participants were given a scenario and instructions on how to use the search interface. The scenarios asked...
subjects to imagine that they had just moved to California and that an election was coming up.

In the Non-Voting scenario, participants were told that they were journalists, that they would be “interviewing some local political figures in addition to some experts in the area of a proposition,” and that they should “learn as much as you can about the candidates and the proposition so that you will be able to thoroughly interview each participant and mediate the conversations.” In the Voting scenario, participants were told that they were going to “vote for one candidate for controller, one candidate for senate and yes or no for one proposition” and that they should “find information to help you make voting decisions.”

Participants used either the Yahoo! search engine's traditional interface or the drop-down interface depicted in Figures 1 and 2 (which obtained results from the Yahoo!).

All participants were informed that there were two candidates for U.S. Senate: Dianne Feinstein and Richard Mountjoy; two candidates for California Controller: John Chiang and Tony Strickland; and one proposition: “Proposition 87 – Alternative Energy Research, Production, Incentives, Tax on Oil. Initiative Constitutional Amendment and Statute.” These were actual ballot items in a current election at the time the study was conducted.

While carrying out the tasks described in the scenario participants were encouraged to think aloud. Software was used to capture and integrate the search behavior and verbalizations of each participant. An experimenter remotely tagged the capture file while the participant was searching for information. These tags were adapted from a previous study we conducted on online political information seeking behavior [26].

At the end of the experiment, regardless of the group assignments, participants were asked to vote. They were also asked to provide a free recall for each ballot item and answer some debriefing questions.

3. Results

In this section we report results of several dependent measures. In all cases where means are reported, we performed a 2x2 analysis of variance (ANOVA) using Search Interface and Information Task as the independent variables.

3.1 Queries

Subjects made 356 queries overall. Out of all queries, 103 (30%) consisted of a candidate’s name only (e.g. “Diane Feinstein”), 110 (31%) consisted of a candidate’s name and an issue (e.g. “Diane Feinstein taxes”), and 5 (1.4%) contained only issue information. Out of the 138 remaining queries, 39 (11%) contained combinations of the candidate name and proposition name, or office, or other candidates; 65 (18.25%) contained just the proposition name and its variants, e.g. “CA and Prop 87”, or the full name of the proposition; 21 (8.7%) were comprised of the state name, office, or a combination of both; and 13 (3.7%) were general queries.

It was possible to make four types of queries using either interface, specifically: candidate-only, candidate-plus-issue, issue-only, and proposition-name (other query types were possible only in the query interface). Figure 3 shows the number of queries observed of each of the four types possible in both interfaces for users of the drop-down interface and traditional query interface. The distribution of these four query types was significantly different across the two interfaces, \( \chi^2(3)=67.93, p<.01 \), with the drop-down interface resulting in many more candidate-plus-issue queries. The distribution of query types was not significantly different across the two voting conditions.

There was an average of 12.31 unique queries observed per subject. There was a marginally significant interaction between the voting and search interface conditions (Figure 4), with voting subjects making more queries when they were using the drop-down interface and non-voting subjects making more queries when they
were using the query entry interface, $F(1,28)=3.37, p<.08$.

Subjects revisited previous queries an average of 9.59 times, but users of the query interface revisited previous queries considerably more than users of the drop-down interface (13.71 versus 5.73 revisits, respectively), $F(1,28)=10.08, p<.01$.

The average length of queries was greater for the traditional query interface versus the drop-down interface (3.07 versus 2.50 words, respectively), $F(1,28)=6.61, p<.01$, although it is important to remember that query length was not under the direct control of subjects using the drop-down interface.

### 3.2 Time, Number of Visits, and Information Opportunity

Subjects spent an average of 42min, 10sec on the entire task, and this did not differ across conditions. The total time can be partitioned into mean time entering queries (1min, 36sec), mean time viewing results lists (5min, 12sec), and mean time browsing websites (30min, 41sec) (the remaining 5 minutes or so is time spent in activities unrelated to the task such as asking questions, taking breaks, etc.). Time spent looking at websites and results lists did not vary across conditions, however there was an interaction between conditions for the time spent entering queries, $F(1,28)=4.37, p<.05$. As Figure 5 shows, voters spent a greater amount of time in the query entry phase when using the drop-down interface, but nonvoters spent a greater amount of time in the query entry phase when using the traditional query interface. This mirrors the results for number of unique queries in Figure 4.

By far the largest percentage of subjects’ time (73%) was spent browsing websites. We grouped the types of websites that subjects looked at into the following seven categories, and here we also report the proportion of total website browsing time spent in each category:

- **Candidate Websites (28%)**: Official website of the candidate, e.g. www.chiangforcalifornia.com
- **Government Websites (19%)**: Official U.S. government websites with URLs ending with dot (. ) gov, e.g. www.ca.gov
- **Political Organizations (18%)**: websites that either reflect party lines e.g. www.freerepublic.com (conservative) or take a clear stance on issues, e.g. www.prolife.com
- **Voter Guides (6%)**: Portals created to consolidate and deliver factual information about the election and candidates, e.g. www.vote-smart.org
- **Wikipedia (6%)**: Online peer-reviewed and authored encyclopedia at http://en.wikipedia.org
- **Blogs (5%)**: Personal blogs hosted by individuals about elections, proposition, and candidates; not part of candidate sites or political organizations.
- **Other (9%)**: Websites that were not related to the task but show up in the results because of similar distribution of keywords, basically noise.

We also examined the number of websites of each category that were available to subjects in their results lists across the conditions. On each results page that a subject looked at we recorded the number of items that were visible (more if the participant scrolled). We describe this as “information opportunity.” The proportion of websites visited relative to the information opportunity was also calculated for each website type.
3.2.1 Candidate Websites
The drop-down interface generated more opportunities to view candidate websites than the query interface (12.73 versus 5.29 opportunities, respectively), $F(1,28) = 20.47, p < .01$. Drop-down interface users looked at a smaller proportion of available candidate websites than query interface users (.28 versus .87, respectively), $F(1,28) = 5.61, p < .05$. Subjects using the drop-down interface revisited candidate websites fewer times than subjects using the query interface (4.66 versus 9.18 revisits, respectively), $F(1,28) = 12.64, p < .01$, and subjects using the drop-down interface spent half as much time overall at candidate websites than subjects using the traditional query interface (6.6 min. versus 12.01 min., respectively), $F(1,28) = 7.87, p < .01$.

3.2.2 News
The drop-down interface offered more opportunities to view news items than the query interface (27.6 versus 15.0 opportunities, respectively), $F(1,28) = 4.03, p < .05$. This did not result in a difference in the number of visits or revisits to news sites, however subjects using the drop-down interface indeed spent more time at news sites than subjects using the query interface (3.07 minutes versus 1.21 min., respectively), $F(1,28) = 4.14, p < .05$.

3.2.3 Voter Guides
The drop-down interface offered more opportunities to view voter guides than the query interface (14.40 versus 5.86 opportunities, respectively), $F(1,28) = 7.85, p < .01$. Subjects using the drop-down interface viewed voter guides more than subjects using the query interface (2.0 versus .79 times, respectively), although the result was marginally significant, $F(1,28) = 3.43, p < .08$. Also, subjects using the drop-down interface spent more time with voter guides overall than subjects using the query interface (3 min. versus 42 sec., respectively), although the result was also marginally significant, $F(1,28) = 3.13, p < .09$.

3.2.4 Political Organizations
The drop-down interface offered more opportunities to view political organization sites than the query interface (28.27 versus 16.71 opportunities, respectively), $F(1,28) = 4.77, p < .05$. Subjects using the drop-down interface viewed political organization websites more than subjects using the query interface (6.4 times versus 2.0 times, respectively), $F(1,28) = 10.18, p < .01$. Subjects using the drop-down interface spent more time at websites of political organizations than subjects using the query interface (7.43 minutes versus 3.77 min., respectively), $F(1,28) = 4.65, p < .05$.

3.2.5 Government Websites
There were no significant effects involving government websites of either the interface or voting variables with regard to information opportunity, number of visits or revisits, or time spent.

3.2.6 Wikipedia
Subjects who were not voting had a tendency to revisit Wikipedia more than subjects who were voting (2.07 times versus 0.47 times, respectively), $F(1,28) = 3.89, p < .06$.

3.2.7 Blogs
Although the average number of blog visits was very low (1.52 visits per subject), a marginal interaction suggested that there was a tendency for subjects to spend more time in blogs when they were not voting and using the traditional query interface, $F(1,28) = 3.49, p < .07$. Subjects using the query interface spent an average of 28 sec. looking at blogs when voting but 3 min. 32 sec. when not voting. In contrast, subjects using the drop-down interface spent an average of 1 min. 34 sec. looking at blogs when voting but 58 sec. when not voting.

3.2.8 Other Websites
The query interface offered more opportunities to view other websites not related to the election, although the difference was not significant. There was a corresponding tendency for users of the query interface to spend more time at websites in the “Other” category than users of the drop-down interface (4 min. 8 sec. versus 1 min. 32 sec., respectively), $F(1,28) = 3.49, p < .07$.

3.3 Use of the Results Page
We counted the number of times subjects jumped to a web page from the results list and from another web page. We conducted a 2x2x2 mixed design ANOVA using origin of a hypertext jump (results list versus web page) as a repeated measures factor and voting condition and interface condition as between-subjects factors.
Subjects relied heavily on the results list to find new web sites rather than following links between web sites. On average, subjects jumped to 20.97 web sites from the results page but only 1.90 web sites from other web sites, \( F(1,25)=98.99, p<.001 \). A significant interaction between search method and origin of the hypertext jump indicated that the number of jumps from the results list differed depending on the type of interface being used, \( F(1,25)=8.46, p<.01 \). Subjects jumped to web sites from the results list an average of 15.93 times when using the query interface and 25.67 times when using the drop-down interface, but the small number of jumps between pages was the same for both interfaces.

3.4 Candidate-Centered versus Issue-Centered Searching

We examined sequences of search query terms in order to determine if participants tended to do a lot of candidate switching or to stay with one candidate at a time and look at issues relevant to that one candidate. Candidate switching was indicated when participants made sequential queries consisting only of different candidate names. Issue switching within candidates was indicated when participants first searched a candidate name and then followed it by adding an item to the name and executing a second query.

Users of the drop-down interface performed much more issue switching within candidates than users of the query interface (6.73 versus 1.57 instances, respectively), \( F(1,28)=5.31, p<.05 \). There was no difference between interface conditions in terms of the amount of candidate switching, however non-voting subjects did more candidate switching than voting subjects (4.43 versus 2.93 instances, respectively), \( F(1,28)=4.09, p<.05 \).

Payne [22, 29] describes an index to measure how much a person is candidate-centered versus issue-centered when browsing websites. We adapted this index to characterize search strategy by calculating the following ratio:

\[
\frac{\text{Issue Switches} - \text{Candidate Switches}}{\text{Issue Switches} + \text{Candidate Switches}}
\]

The index ranges from -1 for searches consisting of all candidate switches to +1 for searches consisting of all issue switches. Subjects using the drop-down interface tended to do both types of searches whereas subjects using the query interface tended to be candidate switchers (indices = 0.00 and -0.45, respectively), \( F(1,28)=4.80, p<.05 \).

3.5 Comments

While searching and browsing, the subjects were encouraged to think-aloud which resulted in 2080 individual comments. The comments were coded into 10 categories (adapted from [26]) by two coders independently. Cohen’s Kappa for assessing inter-coder reliability was initially 0.52, which translates to moderate agreement [30]. The coders reconciled differences and eventually assigned each comment to a final category as follows:

- **Goal (11%)**: A statement about what the participant plans to do, e.g. “I would like to find a government website that explains what the proposition is.”
- **Action (6%)**: A statement describing what the participant was doing, e.g. “I’m checking wiki.”
- **Question (3%)**: An interrogative statement, e.g. “Who is endorsing him?”
- **Evaluative General (18%)**: A general evaluative remark but not related to the ballot item, e.g. “I don’t see anything helpful so I’ll search for others.”
- **Evaluative about a ballot item (16%)**: Evaluative comment but cannot be determined positive or negative about the ballot item, e.g. “It looks like he is more popular because he has a double digit lead” OR “This (prop87) may be better for researchers but maybe not for consumers.”
- **Positive about a ballot item (6%)**: A good evaluative comment in support of the ballot item, e.g. “I get a good vibe from him, I’d vote for this guy.”
- **Negative about a ballot item (4%)**: A bad evaluative remark about the ballot item, e.g. “I don’t like this guy, I can’t find anything about his stance on issues.”
- **Fact Discovery (16%)**: A statement of a non-evaluative piece of information about one of the candidates, e.g. “He served in the Navy”
- **Issue (7%)**: A statement about a particular political issue, e.g. “She combats crime, violence, cancer.”
- **General Statement (13%)**: A non-evaluative comment not specifically about a candidate, e.g. “I don’t really follow politics.”

Subjects using the query interface made more
evaluative comments than subjects using the drop-down interface (30.93 versus 18.8 comments, respectively), $F(1,28)=5.30, p<.05$.

Subjects made more positive comments about the ballot items when they were voting compared to not voting (6.4 versus 2.0 comments, respectively), $F(1,28)=4.75, p<.05$. They also made twice as many negative comments about ballot items when they were voting compared to not voting (4.27 versus 2.07 comments, respectively), although this difference was not statistically significant.

Subjects asked more questions when they were not voting versus voting (3.43 versus .8 questions, respectively), $F(1,28)=10.12, p<.01$.

3.6 Recall
Overall, subjects recalled an average of 15.0 total items and their recalls did not vary across voting or interface conditions.

3.7 Summary of Results

3.7.1 Information Task
In contrast to the non-voting subjects, subjects who thought they were voting:

- Asked more questions while browsing
- Showed less candidate switching in their query patterns
- Made more positive and negative comments about the candidates and issues while browsing
- Visited Wikipedia less often

This is consistent with a view that the task of deciding who to vote for is more evaluative, more specific, less fact-oriented, and less generalized. The larger number of questions, the tendency to explore more within each candidate site, and the relative lack of interest in Wikipedia suggest that search is more directed for voters as opposed to more generalized for non-voters.

3.7.2 Search Interface
In contrast to the query interface, subjects using the drop-down interface:

- Had more opportunities to view candidate websites but actually viewed a smaller proportion of what was available
- Had more opportunities to view political organization websites and took advantage of that by actually viewing a greater proportion of what was available
- Used the results list more to jump to websites
- Showed more issue switching within candidates and a balanced use of candidate switching and issue switching in their query patterns.
- Revisited previous queries less
- Returned to candidate websites less often and spent half as much time at them
- Spent more time at news sites
- Viewed voter guides more often and spent more time with them
- Viewed websites of political organizations more often and spent more time with them
- Spent less time in websites unrelated to the election issues
- Made fewer evaluative comments

The results are consistent with the view that the drop-down interface allowed subjects to make more focused searches. The longer time spent in query formulation by voters using the drop-down list reflects time spent browsing the issues and deciding which ones to select. The drop-down interface provided a recognition list of items that helped subjects think of issues to explore.

The results are also consistent with a view that the drop-down interface resulted in less reliance on candidates’ own websites to gain information and guide search. The drop-down interface gave voters more opportunities to view information from media, political organizations, and non-aligned voter organizations, and voters took these opportunities. Once drop-down interface users began exploring non-candidate information sources, they spent relatively more time with them and less time at irrelevant websites.

Although we have no direct measure, the fact that subjects went to many more websites from the results pages in the drop-down condition than the query condition suggests that the results obtained from the drop-down interface may have been more interesting and relevant than the results obtained using the query interface.
4. Discussion

Our results demonstrate that the form of a search portal can have dramatic effects on what a voter is exposed to and how he or she searches. Voters using a traditional query interface generate impoverished queries consisting essentially of the candidates’ names. They consequently generate a single result set for each candidate and use these single sets as their points of entry to the internet. Candidate websites are naturally high in the result lists and are often used as the organizing point for finding further information. The overall strategy of these users is candidate-centered.

In the drop-down interface condition, voters had the option of simply choosing the candidate name and searching the same way as query-interface users. However, the presence of the issue drop-down list prompted them to make more queries and be more specific. This in turn generated many different result lists for each candidate and provided many more information opportunities. Voters in this condition wound up relying less on candidate websites and sampling a broader variety of other types of websites, especially news, voter guides, and political organizations. They changed the queries frequently, returned to the results sets often, and spent more time considering what to look at. These users demonstrated a balance of both candidate-centered and attribute-centered search strategies.

Our results leave us with many questions. First, how malleable are search strategies? Huang & Price [15] demonstrated that candidate-centered and attribute-centered searching can vary as a function of searchers’ goals, and our results further suggest that voters will switch strategies depending on the characteristics of the search tool.

Second, how far should search aids go in restructuring searchers’ queries? We chose to add terms to the query that users constructed using the drop-down interface in order to make them more specific and targeted, but users were not aware of this added specificity. Should users be aware of the exact query that is being generated and should they have the option to change it? What dangers are there in adding terms and not making them visible to users? To what degree would seeing and being able to change automatically-generated queries help voters learn?

Third, to what degree should designers choose the items that will be available in a search interface aid? Should the items be customizable by users, how would this change their search strategies, and how would novices and experts differ in their ability to personalize?

Finally, who would benefit from front-end “query prostheses” such as the one studied here, and who would be hindered? What other enhancements to search tools would be beneficial in the context of voting and democratic deliberation? Our users’ informal comments, for example, suggested that they would like to see categorized results sets and be able to perform “who’s for and who’s against?” searches.

It is important for developers of tools in this area to understand how they are influencing the goals, actions, information opportunities, and ultimate decisions that voters make.

5. Acknowledgement

This work is supported by the National Science Foundation under grant number IIS-0535036 to the first author.

6. References


Assessing the Value of a Pattern Structure For Communicating Design Advice

GEORGE ABRAHAM
Drexel University

Short Abstract (50 words):
This paper presents findings from an experiment evaluating the impact of a pattern structure on the worthiness of design advice, and explores whether patterns facilitate communication of design knowledge. The study includes claims as an alternate structure. Findings indicate that each structure has strengths that can be combined to yield a more robust representation.
Assessing the Value of a Pattern Structure For Communicating Design Advice

Abstract

This paper presents findings from an experiment evaluating the impact of a pattern structure on the worthiness of design advice. It explores the broader issue whether patterns facilitate communication of design knowledge. This study differs from past research by focusing on the benefits of using a pattern structure on design advice, and not on the efficacy of available patterns in collections. In addition to patterns, the study includes claims as an alternate structure. Design advice prepared using pattern or claim structure was judged by experts using the following four criteria: context-of-use described, rationale provide, usefulness of advice, and overall quality. Our findings indicate that patterns and claims each have certain strengths that can be combined to yield a more robust representation, and invites a rethink of the ubiquitous pattern structure. This research responds to the call for much needed empirical work in the area of patterns in HCI.

Keywords

Patterns, claims, benefits, experiment, interaction design

ACM Classification Keywords

H.1.2. User/Machine Systems, H.5.2. User Interfaces
Introduction
A pattern is a relatively familiar concept in HCI, yet we are not as familiar with what or how does it facilitate design or evaluation of interactive systems. A pattern, simply defined, attempts to capture the essence of solutions to a recurring interaction design problem in a given context.

Since its introduction, interest in this topic has prompted the HCI community to document patterns in interaction design (e.g., [21]). Consequently, patterns in HCI lead in numbers when compared to other disciplines (e.g., software engineering, pedagogy), and possibly growing. In contrast, Alexander's proposal, consisting of 253 patterns [2], were an attempt to address recurring design problems in architecture ranging from design of rooms to urban planning! It is still unclear whether we, in HCI, are better off with more patterns or whether documenting more patterns will bring us closer to understanding the potential value of patterns.

Dearden and Finlay [9], in their review of pattern languages in HCI, draw our attention to insufficient evidence demonstrating benefits of patterns or pattern languages either to the design process or the designed product. We are aware of several studies exploring this question in the area of ubiquitous computing [7, 17], participatory design [8], teaching HCI [4, 14], and information retrieval [22]. The controlled studies focused on the efficacy of using patterns for improving the quality of the design product. The evidence, however, is not conclusive whether patterns help. Nevertheless, the studies seem to support the view that patterns may help communication during the design process.

"Patterns support communication" appears to be a recurring theme in the pattern literature. Saponas et al. [17] briefly explored this premise in their study, where they looked at how patterns were being used during design discussions. It is theorized that the concrete and contextual nature of advice presented as patterns will facilitate communication [3, 10]. This study explores the communication aspect further by asking whether patterns add value when communicating interaction design advice. It differs from past research in two significant ways:

- It evaluates possible benefits of structuring design advice in a pattern form (i.e., problem, solution, and context), and not the efficacy of existing patterns or collections
- It studies what impact such a structure may have on the quality of design advice, which can be considered to be a meaningful and specific subset of design communication (patterns are after all meant to provide design guidance)

We understand patterns to be a complex concept (i.e., invariance, shared values, pattern relationships), and believe that by isolating these embedded concepts we may begin to understand the possible value or issues surrounding each. With this goal, we suggest studying the pattern structure as a first step. The counter-argument may be that the value of patterns cannot be determined by focusing on the individual pattern properties, and that it is a combination of these individual properties that lend strength to patterns. While this sounds reasonable, we are yet to see evidence to merit either perspective.
The following sections briefly describe why a claim structure was selected for comparison in addition to our present focus on patterns. We then discuss the experimental design and define the criteria used to evaluate design advice. Based on the preliminary findings, we then discuss why we may need to reconsider the pervasive pattern structure.

Choosing Claims
There is sufficient literature describing the use of claims in design of interactive systems [5, 6]. Our intent here is not to provide a detailed review of claims or redefine claims, but rather to briefly highlight reasons for including claims in this study.

Thomas and Moran [16] present a reasonably detailed overview of different approaches available for documenting design rationale, both about the process or the designed artifact. Among them, claims analysis involves identifying the positive and negative implications on usability, or use, stemming from a design choice. A scenario of use that involves the task and artifact forms the basis for this reflective exercise. A form of rationale is explicated when debating the pros and cons; the understanding being that this rationale may allow others to comprehend why a designed artifact is the way it is.

Discussing rationale is relevant to patterns because no definition of a pattern is complete without a discussion of why it is a good solution in a given context. Design rationale in the pattern sense consists of justifications for the solution (i.e., make obvious some of the reasoning behind the design decision). In addition to including rationale, a pattern also attempts to guide the reader towards a feasible solution for a design problem. But unlike patterns, where the justifications are presented mainly to support a solution, claims attempt to present a balanced view. It seems to emphasize the design trade-offs, whereas patterns emphasize relevancy of design advice.

As for similarities, both claims and patterns appear to be a reflective form of rationale compared to other design rationale approaches, which seem geared towards capturing design history. It has been suggested that the documented claims may promote re-use of design knowledge by making the design advice more usable [20]. Similar proposals have also been made about patterns in HCI. And like pattern collections, there also exists a claims library [11, 19]. The basic claim structure (scenario of use, pros, and cons) is similar to the pattern structure in terms of complexity. This is important to consider because it helps minimize the amount of training that might be required by participants. Furthermore, claims is a familiar concept in HCI that is relatively easy for the participants to understand and use.

For a pattern, we chose the problem-solution-context structure, which is considered to be the essence of a pattern [1].

Method
A between-subjects design was used to randomly assign 57 participants into three groups: pattern structure, claim structure, and control. All participants were undergraduate students from an information and technology school, and had completed coursework in HCI. Participants were familiar with the concept of usability, design principles, and heuristics (completing
at least an introductory course in HCI was a prerequisite for the study).

Table 1. Number of participants assigned to each condition

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Claim</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>19</td>
<td>19</td>
</tr>
</tbody>
</table>

The instructions stated that a room planner will enable users to add and remove furniture from an online catalog to a room plan, and allow them to try out different arrangements. To focus the participants on capturing design advice, instead of asking each of the to come up with their own menu design, they received a sheet with a randomized ordering of three menu options: Directory-tree, Drop-down, and Accordion (shown in Fig.1). All three options are fairly common in interactive systems. These options were presented side-by-side and included a brief set of features describing the menu. That is, the Directory-tree menu listed the following: Expand/Collapse different sections using the +/-; Multiple sections can be opened at a time; Product information and image displayed inside each menu; Fixed height for the menu; Scroll (outside) to view the open catalog. The other two menus also listed identical five features, but the features were customized for the menu type (i.e., for the Accordion menu it read that clicking on the arrows open/closes the section)

Participants were first asked to rank order the three menu options, after which they were instructed to explain their top pick. Depending on the assigned condition, participants were given additional instructions; participants in the pattern and claims condition were asked to address three questions when explaining their menu choice, shown in the following table 2:

In the study, each participant was asked to adopt the role of a designer and recommend a menu design for a room planner interface. Participants received a stripped down version of room-planner as shown in Fig.1.
Pattern
What is the user trying to do?
How will the recommended menu option be used?
Explain why it’s the best option in this given situation?

Claims
How will the recommended menu option be used?
Advantages of using this design choice?
Disadvantages of using this design choice?

Table 2. Additional instructions provided to the participants. The pattern condition questions attempt to address the problem, solution, and context structure of a pattern. Similarly, the questions for the claims condition was derived from claim structure.

All participants were asked to explain their design decision as if they were teaching someone else to make a similar choice, and that others reviewing this design advice may not see the menu options they rejected. Their recommendations were rated for quality by two expert judges (each having over 20 years experience in HCI education and research; an approach adopted in other studies on patterns ([17, 22]). We note that no absolute measures exist to evaluate design advice because no representation in itself complete or sufficient to fit all purposes; nonetheless, the following four criteria were adapted from other studies [13, 15]:

- **Context of Use Described** or how well does the design advice describes the task context and the concreteness with which it is described (i.e., usage scenario, specific examples)
- **Rationale** or whether reasons were offered for calling the menu option a good fit, along with consequences of using this option
- **Usefulness** or how well does the design advice address the interaction (fit) between the design choice and the task context (i.e., relevancy of rationale)
- **Overall Quality** or in general how good is the provided design advice

The recommendation or design advice was rated on these four criteria using a 7-point Likert type scale (e.g., **Context of Use**: 1 No Description -2-3-4-5-6-7 As good as it gets)

Preliminary Findings
For each recommendation, the ratings from the two judges were averaged. Treating the rating data as interval-type, intraclass correlation coefficient was computed to assess the reliability of ratings between the two judges. ICC (3, k) has been suggested in the literature [18] when each target is rated independently by a fixed number of judges (i.e., a two-way mixed model, where k=2). Since the ratings were averaged, the consistency option was used instead of perfect agreement. The following table 3 shows the average measure of ICC coefficients for each criteria showing good consistency between the judges:

<table>
<thead>
<tr>
<th></th>
<th>Context</th>
<th>Rationale</th>
<th>Usefulness</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.81</td>
<td>0.79</td>
<td>0.80</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Table 3. Average measures for ICC (3, 2) using consistency option

An analysis of variance (ANOVA) of the rating data found the effect of structure on **usefulness** to be significant, F(2, 18)=6.783, p <0.05. A post-hoc analysis using Scheffe’s post-hoc criterion showed that
the advice prepared using the claim structure received higher mean ratings.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Pattern</th>
<th>Claims</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td>3.00, 1.07</td>
<td>2.71, 1.13</td>
<td>2.68, 0.96</td>
</tr>
<tr>
<td>Rationale</td>
<td>3.50, 0.88</td>
<td>3.84, 1.29</td>
<td>3.39, 1.08</td>
</tr>
<tr>
<td>Usefulness*</td>
<td>3.13, 0.99</td>
<td>3.95, 1.12</td>
<td>2.79, 0.85</td>
</tr>
<tr>
<td>Quality</td>
<td>3.18, 0.87</td>
<td>3.34, 0.90</td>
<td>3.00, 0.83</td>
</tr>
</tbody>
</table>

Table 4. Descriptive statistics showing means (M) and std. deviations (SD); * significant at p<0.05. A 7-pt Lickert type scale was used.

Claim structure was also awarded higher mean ratings for rationale and overall quality although the differences were not significant. Design advice prepared using the pattern structure received higher mean ratings for context of use described than those using the claim structure or control group, again not significant.

Discussion & Future Work

Patterns have been argued as a format to capture and disseminate design advice. Hence, it’s reasonable to examine whether the ubiquitous pattern structure (Problem + Solution + Context) affects the worthiness of the design guidance.

Preliminary findings suggest that when it comes to capturing design advice, the claim structure elicited a description that was perceived to be more useful. Usefulness was defined as relevancy of the rationale provided. When authoring design guidance, the challenge is to determine whether the captured description is sufficient (usually prepared to benefit people other than the author). The importance of context-of-use in such descriptions has become increasingly apparent (i.e., in a world where mobile platforms co-exist with desktop based solutions, the design guidance may need to be contextualized). The ability to contextualize or reinterpret design advice in newer situations appears to reside with the designer, but this skill cannot be taken for granted. The design-advice prepared using a pattern structure appears to perform better in terms of capturing the context-of-use. Comparison of word counts (for advice) showed that both pattern and claims condition had higher means compared to the control group, and the differences were statistically significant. The mean difference between pattern condition and claim condition was not. A more detailed content-analysis of design-advice may reveal how different or similar these groups were in terms of focus. This phase of analysis is ongoing.

Griffiths and Pemberton [12] assert that teaching a process by which patterns are identified and explicated may prove to be more useful to students than a familiarity with pattern collections. It is, however, unclear what needs to be documented. Our findings indicate that describing the problem, solution and context may be inadequate. Claims seem to benefit from a discussion of negative consequences of a particular design decision. The higher usefulness rating for the claim structure prompts us to consider whether including cons in a pattern structure will be valuable. This seems counter-intuitive because it could be argued that patterns stem from addressing the negatives in a design situation (a problem), and hence an explicit focus on cons may be unnecessary. However, we theorize that patterns for
interactive systems may require a different treatment (documenting) compared to architecture. Pointing out the cons of a design choice may prove to more helpful in the context of interactive systems, where we may not have the benefit of lived-in experiences to understand the negative consequences of our design choices when compared to patterns in architecture.

This paper reports only on the experimental design, preliminary findings and some implications, as a detailed analysis is ongoing. As stated earlier, content analysis to delve deeper into the similarities and differences between design-advice prepared using the pattern and claim structure is pending. Also pending is a review to see whether existing patterns in collections address cons of a design solution, and if so, how. We look forward to extending the discussion on patterns new directions and further examining the potential value of patterns to HCI.

References


Evaluating a Pattern Language as Shared Language for Interaction Design

George Abraham
Drexel University
3141, Chestnut St, Philadelphia, PA-19104
George.Abraham@iSchool.drexel.edu

Abstract
Few studies have empirically evaluated pattern languages as a shared language for interaction design. In the HCI community, interaction patterns proliferate, but consensus on what constitutes a pattern language is missing. We evaluate four criteria required of a pattern language, and its impact on shared understanding.

Keyword
Pattern-Language
Evaluating a Pattern Language as Shared Language for Interaction Design

George Abraham
Drexel University
3141, Chestnut St, Philadelphia, PA-19104
George.Abraham@iSchool.drexel.edu

ABSTRACT
Building shared understanding, between users and designers, through a common language for design appears to be a fundamental assertion shared between HCI and Architecture about pattern languages. Despite the popularity of interaction patterns, there are very few empirical studies evaluating this premise of a shared language for design. In this paper we briefly summarize our experimental design for a planned study involving users, designers, and pattern languages. We propose four criteria (cohesiveness, generativity, design guidance, coherence) to qualify a pattern language, and evaluate it using a pattern-sorting exercise. Preliminary findings show that a pattern language influences shared understanding, or a shared organization, irrespective of an explicit design task.

Author Keywords
Pattern Language, shared understanding, empirical study, interaction design

INTRODUCTION
Interaction patterns are a hot topic in HCI [13]! An interaction pattern captures the essence of solutions to a recurring design problem, in a specific context. While an interaction pattern targets a specific type of interaction problem, a pattern language consists of a cohesive network of related interaction patterns that might help design an interactive system of some kind, e.g. website, interactive exhibit, kiosk [10, 17].

Despite the popularity, there is a lack of empirical studies evaluating whether the premise of shared understanding through pattern languages is valid. Lack of empirical work on using pattern languages in HCI is its primary weakness [14]. The only three known controlled studies explored the benefits of pattern languages only with designers (or novice designers), not users [12, 27, 31]. It could be argued experts already have access to a pattern language of some kind, and hence may not be the right audience. [15, 16, 20] report the use of pattern languages within a participatory setting with users, but even though users reported a positive experience, it is difficult to differentiate the impact of pattern language from the facilitator’s expertise.

In this paper we briefly describe the experimental design for a larger controlled study involving both users and designers, and the impact of a pattern language on the design of interfaces. The study would evaluate if patterns languages are usable to both users and designers. The focus of this paper is on validating four criteria to qualify a collection of interaction patterns as a pattern language: cohesiveness, generativity, design guidance, and coherence, using a pattern-sorting exercise. This exercise evaluates whether exposure to a specific design task affects how pattern-users organize a collection, which satisfy the four criteria. Findings show that, to a large extent, users (of the language) are able to see a shared organization, independent of the specific task. This validated pattern language will be used in the main study.

RELATED WORK
This section reviews the origins of the pattern concept in architecture, and its subsequent adaptation by HCI and other design communities, notably Software Engineering (SE). This is followed by a summary of empirical studies on evaluating pattern languages in HCI.

Pattern Adaptations
The concept of a pattern is popular, and actively discussed, within SE and HCI communities. Origins of these adaptations, software design patterns and interaction patterns, lie in architecture. Christopher Alexander and his colleagues compiled a set of 253 patterns, which they published as “A Pattern Language” for architecture [2]. They claimed that each living space is designed to support a pattern of activities that takes place within this space. In more general terms, the form we design depends on the context. Furthermore, they argued, it was possible to extract a spatial arrangement of elements in the space that help minimize (remove) the misfit between the form and context, or resolves the forces that exist. The forces could refer to the opposing requirements we have from a space, e.g. communal rooms seek a balance between a feeling of community and privacy. In response, the Alcoves pattern ([2], p. 253) describes how we could design the space to resolve these opposing forces.

Software design patterns and interactions patterns are two notable adaptations of patterns in Architecture. Although
the pattern concept has inspired adaptations in other domains, e.g. groupware design patterns, pedagogical patterns, socio-technical patterns, SE and HCI are prominent having devoted more research attention to the discovery of patterns in their respective domains. This is evident in the number of published patterns [22].

Software design patterns have taken up the explicit goal of re-use of best practices in software design; object-oriented software design to be specific. These best practices describe how to build code that works, is efficient, and easy to maintain. Gamma, Helm, Johnson, & Vlissides [21], who are attributed for coming up with software design patterns, state the purpose of design patterns as a way to, "… help the designer to get a design right faster" (p.2).

In comparison, patterns and pattern languages in HCI have been proposed as a lingua franca for design [6, 18], a common language that could be used by designers and users to communicate. Carroll in his comments on [14], refers to patterns as a “powerful, yet simple idea that could be explored.” He states patterns possess a “moderate level of abstraction”, but are concrete enough to facilitate communication within, and among, users and designers. Importantly, “… they help to anchor design rationale and design values, making it easier for designers to articulate and manage the embedded meanings and consequences of their designs” ([11], p.2). The HCI community has asserted that HCI is more similar to architecture than software engineering. Fincher et al [19] suggest that "It’s relatively easy to make an analogy between the domains of architecture and UI design, based on concern for the effect of a constructed artifact on personal and social behaviors …” (p. 1044).

[7] & [9] propose that interaction patterns should be readable to users, and usable to anyone with a stake in the design. Here, the patterns not only aim to serve as a format, but also aim to deliver design advice usable to designers and non-designers.

Author co-citation (ACA) studies in HCI [32] and design in general [4] have shown that Alexander’s work is viewed as contributing to Design Complexity and Theory research. ACA belongs to a larger body of techniques devoted to the quantitative study of academic literature. ACA implies relationships between authors based on the number of times two authors are cited together in one document (reference section). These studies show a degree of similarity between Alexander and Donald Schon (Design Theorists), Nigel Cross (Design Taxonomists), Horst Rittel and Chris Argyris (Design Complexity). This seems to validate that HCI cares about pattern languages as an approach to design.

Empirical Studies

Despite all that has been written about patterns and pattern languages in HCI, there is little empirical support verifying benefits of patterns and pattern languages in design, or validity of published patterns and pattern languages themselves [14]. Even fewer studies have explored the premise of pattern languages as a shared language for interaction design. Patterns have been frequently discussed for more than a decade in the HCI community. There is merit in the concept of patterns for the broad HCI community. But, while descriptions of pattern collections proliferate, discussions of pattern languages are fragmented. Dearden and Finlay [14] in their survey of pattern languages report that empirical studies evaluating impact of pattern languages in design, or design process, remains its biggest weakness. To our knowledge only three controlled studies exist [12, 27, 31], but all of them evaluate the impact of a pattern language on the quality of the designed product with designers.

Dearden, Finlay, Allgar and McManus [15, 16] focused on a participatory design process, using pattern languages as a way to involve the users. Although not a controlled study, it used a pattern language, along with a facilitator, for designing paper prototypes of web applications by users. The users ranged in experience from no-experience to some experience in design of similar systems. Despite positive feedback from the participants about the exercise, because of the case study setting, the authors are unable to separate the impact of using Interaction patterns from the influence of facilitator expertise. The authors acknowledged that there was insufficient data to discuss quality.

In a controlled study, Chung et al. [12] explored if pre-patterns were helpful for communication between designers, and if using pre-patterns yielded better designs in the area of Ubiquitous computing (Ubi-Comp). Pre-patterns by definition are not Interaction patterns, but represent candidate Interaction patterns emerging in Ubi-Comp. This controlled study compared use of pre-patterns by experienced and novice designers. They reported novice designers found pre-patterns helpful in understanding a new domain. The experienced designers rated pre-patterns higher in terms of allowing them to finish task quickly and overall usefulness for the task.

Saponas, Prabhaker, Abowd and Landay [27] empirically evaluated the impact of pre-patterns on early stage designs for digital home applications and communication of design ideas. The participants were professional designers with varying years of design experience. They determined pre-patterns lead to higher quality designs on the basis of fewer heuristic violations. The study found difference in the ratings for level of completeness and quality not significant.

Wania and Atwood [31] compared the impact of using pattern language on design of interfaces for an information retrieval system in a controlled study. This study also compares pattern language with heuristics. The primary contribution of this work is in the discovery of a information retrieval pattern language, which is being tested in the study. They showed that when provided with a pattern language, participants were able to see common
relationships among interaction patterns, validated by a pattern-sorting exercise.

Pattern Languages and Shared Understanding
A shared language for design and shared understanding could be considered as the primary contribution of a “A Pattern Language.” Pattern languages in HCI have been proposed as a way to compete with a growing diversity and complexity of interactive systems [6], to serve as a common language for interaction design.

In interaction design, the concrete and contextual nature of design advice in patterns and the generative nature of the language have the potential to embolden users to participate in the design process [18]. Providing design methods usable to users would allow them, if they choose so, to step out of the traditional informative and evaluative roles they are unintentionally cast into, to play a more participative role in design of systems being designed for them [3]. Such benefits cannot be easily overlooked by HCI, especially when interactive systems are pervading multiple contexts of use and spanning diverse stakeholders [5].

Given the significance of shared understanding to pattern languages, none of the controlled studies, described earlier, have explored this aspect of pattern languages. With the exception of [15, 16], these studies rely on evaluating pattern languages with designers, not users.

The availability of over 400 interaction patterns in pattern libraries or collections (web and print) [17, 22, 28], paired with reasonable maturity in how websites are designed and used, sets the stage for re-visiting the discussion on pattern languages and shared understanding. Research attention has persisted long on discovery of interaction patterns and not on building pattern languages. As a result, interaction patterns proliferate [22], but consensus on what constitutes a pattern language is missing.

In the following section we propose four criteria to qualify a pattern language, and describe how interaction patterns were chosen for the pattern sorting exercise.

SELECTING INTERACTION PATTERNS FOR THE STUDY
There are different perspectives in the literature on what constitutes a pattern language [23, 24, 29]. Based on a summary of views in the literature regarding pattern languages in HCI, this study proposes four criteria required of pattern languages: cohesiveness, by restricting the domain or narrowing the scope [24]; generativity, ensuring the connected nature of patterns [1, 8]; design guidance, guiding pattern users to feasible combinations of interaction patterns or qualifying conditions [26, 30]; and coherence, ability to yield concrete artifacts that could be validated against experience [1].

For this study, based on the proposed criteria, a pattern language for supporting an e-commerce buy transaction task was selected from a published collection of interaction patterns [17]. The E-commerce domain, and related buy transaction were chosen for two reasons: First, a reasonable degree of maturity of the domain ensured patterns were available in pattern collections; second, given the trend in e-commerce (B2C) transactions, there is a greater chance that the study participants have performed a buy transaction, thus allowing the participants to validate the presented interaction patterns against experience (concrete artifacts).

Design of Sites (DoS) [17], and van Welie’s Patterns in interaction design (PID) collections have interaction patterns clearly marked for E-commerce. Between the two, DoS collection was selected as the source for interaction patterns since it satisfies the four criteria mentioned earlier. During the selection process, some patterns were omitted if the referenced patterns did not pertain to the buy transaction. Using this process, thirty-nine interaction patterns were selected. For ease of sorting, as encouraged by similar studies [6], a one-page abridged version of each pattern was used instead of the two-three page long narratives available in print. These 1-page versions were edited so that references to patterns not included in this study were omitted.

METHOD
We first provide a brief overview of the larger controlled study before describing the pattern sorting exercise reported in this paper, to give readers better understanding of planned research.

The main study would assess the impact of a pattern language on shared understanding between users and designers. It will do so empirically by evaluating the outcomes (pay-offs) of a paper-prototyping exercise, completed by user pairs and designer pairs [12, 27], for a design task. This study argues, for a shared language, the users should be able to use it independent of designers, i.e. be an equal opportunity method. Two experts would judge the quality of paper prototypes using criteria derived from past studies [12, 27, 31]. The prototypes delivered by users would be compared to that created by designers to assess whether pattern languages in HCI are usable to users, and whether it contributes to a better design (comparison with control group). The pattern sorting exercise described below serves as an evaluation for the pattern language that would be used in the larger study.

Participants
Seventeen doctoral students in information studies participated in this study. All participants had prior experience buying products, e.g. books, on the Internet, and hence were familiar with the buy transaction. The larger study would make use of designer-designer pair and user-user pairs, but since this was a pilot experiment, some subjects were run as individuals and some as pairs. The breakdown is shown in the table 1 below. We understand mixing pairs and individuals may signal less than desired internal validity, we believe considering each pair as one
unit of analysis would help counter some of the bias in the presented findings.

<table>
<thead>
<tr>
<th>Group I</th>
<th>Group II</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 pairs</td>
<td>3 pairs</td>
</tr>
<tr>
<td>3 individuals</td>
<td>2 individuals</td>
</tr>
</tbody>
</table>

**Table 1.** Number of participants assigned to each condition

**Procedure**

As illustrated in Table 1, participants were divided in two groups. Both groups I & II received an alphabetical list of the 39 pre-selected patterns (one-page version). All participants received a brief tutorial on what was meant by an interaction pattern, by explaining the kind of information captured in the different sections of the Location Bread-Crumbs pattern [17]: what, use when, how, why, and related interaction patterns. Participants were then requested to sort the patterns into piles based on similarity (as perceived by them) between the patterns. They were permitted to have as many piles as necessary, even if a pile contained only a single pattern. They were provided sticky notes to name their piles as they go along, but were told not to worry if they could not find a suitable name for the pile.

The only difference in instructions between groups I & II was, Group-I participants were asked to first review a usage scenario before sorting the patterns. The scenario describes a user performing an online purchase. Following this, they were asked to sketch paper prototypes for interfaces that would support such a transaction. After drawing, Group I participants were asked to sort the interaction patterns such that someone new to this collection could use of the patterns for completing the described scenario. The sort data, which patterns were grouped into what piles, was recorded for data analysis to see if the participants have a shared understanding of the presented patterns.

**Data Analysis**

The sort data was input into a pattern co-occurrence matrix for analysis. Two such aggregate co-occurrence matrices were constructed, one for each group (exposure to task/no exposure) and subjected to cluster analysis and Multi-Dimensional Scaling (MDS) in parallel: SPSS procedure CLUSTER with complete linkages, or furthest neighbor, produced clusters of highly similar patterns; and SPSS procedure ALSCAL (non-metric method) was used to generate a two-dimensional representation of similarity between interaction patterns. An MDS map helps visualize the relationships between individual patterns in a two-dimensional space. In MDS, R Square (percentage variance explained) and Stress value (<0.2 usually acceptable [25]) are considered indicators of goodness of fit.

The overall congruence between the two co-occurrence matrices (group I & II) was measured using Quadratic Assignment Procedure (QAP), which reports a Pearson correlation along with a test of significance. QAP analysis is a procedure used for determining the similarity between networks, and can be performed using UCINET-X analysis tool (QAP correlation procedure).

**PRELIMINARY FINDINGS**

Earlier, we proposed four criteria based on suggestions in the literature that we believe are required of pattern languages and which, we further believe, capture most of the criteria proposed by others.

For these criteria to be satisfied, we would expect to see two results in the sort data reported here. First, the sorts should be reliable. That is, they should show a great deal of coherence and little variability between subjects. Second, there should be minimal differences between the sort performed by subjects who considered a specific interaction design task and those who sorted with no task information.

That is exactly what our findings show—participants see a common organization for interaction patterns irrespective of the task.

Figure 1 (Appendix A) shows the cluster-enhanced MDS map of the pattern-sort data for the group-I who were exposed to the design activity. Hierarchical cluster analysis gave a four-cluster solution where patterns were grouped by WEB-PAGE ELEMENTS/STRUCTURE, USER-ASSURANCE & HELP, TASK-FOCUS, and TASK-CONTEXT. (For reasons of space, the dendogram showing the results of cluster analysis is omitted.) In this case, plotting the co-occurrence data in two dimensions yields a good fit, with a high percentage of variance explained (RSQ=0.97146) and a low stress value (0.077). It shows that the participants in this group see a common arrangement or organization of patterns.

Figure 2 (Appendix A) shows a cluster-enhanced MDS map of the pattern-sort data for the Group-II. Group II participants were not exposed to the design activity, but were asked to sort patterns independently of a given application. Cluster analysis of the sort-data generated a four-cluster solution where patterns were grouped by WEB-PAGE ELEMENTS/STRUCTURE, USER-ASSURANCE & HELP, TASK-FOCUS, and patterns related to design of a PRODUCT PAGE. Plotting the co-occurrence data in two dimensions yields a very good solution, with a high percentage of variance explained (RSQ=0.958), and a low stress value (0.089). In this case too, the participants see a common arrangement or organization of patterns, with two dimensions for distinguishing between patterns.

---

1 Patterns with similar co-occurrence patterns are placed near each other. Those with many links are placed closer to the center of the map, and dissimilar patterns are further apart. Patterns in close proximity but belonging to different clusters have important secondary links or boundary spanners.
Irrespective of whether they were focused on a specific interaction activity, subjects show very cohesive groupings of interaction patterns. But, how similar are these groupings? The degree of similarity indicates the extent to which pattern languages are independent of a given context rather than just a convenient way to describe a given context. Of the 39 patterns, 28 (72%) of them retained their cluster membership when sorts by groups I and II were compared. The Pearson correlation value returned by Quadratic Assignment Procedure (QAP) is used as a measure of the overall similarity between the two networks, or in this case, co-occurrence matrix. The Pearson correlation was found to be 0.62, significant at p < 0.05, which indicates that there’s a significant degree of overlap in the mental models, irrespective of an explicit task.

The high degree of variance explained in the two MDS maps, and a significant correlation between the two co-occurrence matrices, show that a pattern language influences the shared understanding of how the interaction patterns are related.

FUTURE WORK
Preliminary findings from the pattern-sorting exercise show that pattern languages influence a shared understanding, i.e. how participants make sense of the collection. It also reveals that four criteria: cohesiveness, generativity, design guidance, and coherence, we proposed earlier in paper, help qualify a pattern language, which in turn promotes a shared understanding. We believe using a somewhat cohesive language, which in turn promotes a shared guidance, and coherence, we proposed earlier in paper, help evaluate whether pattern languages have the potential to be a shared language for interaction design between users and designers. This planned study would make use of the validated language, reported here, to study the impact on design of interfaces for an online buy transaction. We look forward to completing this research, and welcome any feedback and suggestions.

REFERENCES


APPENDIX A

Figure 1. MDS plot for task based sort group-I. RSQ= 0.97146, Stress= 0.077 (very good if < 0.1)

Figure 2. MDS plot for non-task based sort group-II. RSQ= 0.958, Stress= 0.089 (very good if < 0.1)