Searches and Recommendations: Item-finding in Complex Information Environments

Juha Leino

University of Tampere
Department of Computer Sciences
Interactive Technology
M.Sc. Thesis
Supervisor: Kari-Jouko Räähä
May 2008
Information environments have grown vastly in complexity. Constantly added new features, such as various recommender systems, both offer alternative use approaches and compete for user attention. We studied the actual use of such complex information environments to see how users are using them and what strategies have emerged from that use.

Research has thus far largely focused on studying parts of such environments in isolation, and we wanted to see the whole picture and how the parts are used in the context of the whole. We used applied ethnography, observation with verbal protocol and semi-structured interviews, to study user behavior in Amazon, the world’s largest online retailer. Our six participants were all genuine users.

Our data shows how genuine users actually find items of interest in feature-rich information environments. Furthermore, we discuss how the participants experienced social presence in the environment. Our data is qualitative, and thus we focus on describing and discussing different aspects of and patterns in the item-finding process in addition to formulating tentative theories about user strategies that have emerged in complex information environments.

Our results underscore the need for studying information environments as whole in addition to studying them in parts because the context of the whole affects how users use the features. Neither level can be neglected when striving to build information environments that truly serve their users.

**Keywords:** Information environments, recommenders, applied ethnography, item-finding, strategy, social presence, user study.
Contents

1 Introduction ..........................................................................................................................1

2 Background.........................................................................................................................6

2.1 Recommenders in Amazon ...........................................................................................6

2.1.1 Overview ......................................................................................................................6

2.1.2 Delivering recommendations ....................................................................................8

2.1.3 Item-based collaborative filtering ...........................................................................11

2.1.4 Collecting user preferences .....................................................................................12

2.2 Related work .................................................................................................................14

2.2.1 Overview ...................................................................................................................14

2.2.2 Algorithmic recommenders ....................................................................................16

2.2.3 Customer reviews ...................................................................................................18

2.2.4 Complex information environments ........................................................................23

2.2.5 Social presence .........................................................................................................26

3 Method and participants .................................................................................................31

3.1 Method ..........................................................................................................................31

3.2 Participants ...................................................................................................................35

4 Results and discussion ....................................................................................................38

4.1 Finding items of interest ..............................................................................................38

4.1.1 Overview ...................................................................................................................38

4.1.2 Item-finding process in Task 1 ................................................................................40

4.1.3 Searching vs. recommendations vs. categories .......................................................47

4.1.4 Opportunistic and strategic use of recommenders ...................................................51

4.1.5 Best item strategy vs. Good enough item strategy ................................................53

4.1.6 Non-Amazon sources and their influence: Google and others ...............................55

4.2 Star ratings and Helpful ratings ....................................................................................57

4.2.1 Overview ...................................................................................................................57

4.2.2 Star ratings and other factors on the list page .........................................................58

4.2.3 Helpful ratings on Customer Reviews ....................................................................61

4.3 Reviews, Search Inside and Customer Reviews ...........................................................63

4.3.1 Overview ...................................................................................................................63

4.3.2 Search inside .............................................................................................................64
1 Introduction

“If I have 3 million customers on the Web, I should have 3 million stores on the Web”
- Jeff Bezos, CEO and founder of Amazon.com™

In the olden days of the Internet, we had two ways to find items of interest: Searching by typing keywords and searching by categories. However, explosion in the number of available items has made it challenging to find items of interest [Sarwar et al., 2000] with these two approaches, as the amount of information is growing faster than our capacity to process it [Sarwar et al., 2001]. For instance, a search without a keyword in Amazon.com’s book section returns 17,561,392 items (Feb. 29, 2008). Many other online services, such as Flickr¹, YouTube², and Del.icio.us³, do not pale in comparison when it comes to the number of items. We can use keyword searches to cut the humongous cake into smaller slices, but even the slices remain so big that finding an item that is of interest to us out of all the returned items remains challenging. Additionally, the more thinly we slice the cake, the more likely we are to crop some items of interest out of our slice.

This information overload has in part forced us to find better ways to assist users in finding items of interest [Sarwar et al., 2000; Kumar and Benbasat, 2006]. We have begun to harness collective intelligence [Weiss, 2005; O’Reilly, 2005; Riedl and Dourish, 2005] to build new ways to navigate to items of interest. Coined social navigation [Dourish and Chalmers, 1994; Dieberger et al., 2000], collective intelligence as a navigational tool takes many shapes and forms. One of the most prominent shapes it has assumed today are the omnipresent recommender systems that try to predict which items a particular user finds interesting and how well he or she likes individual items [Sarwar et al., 2000; Schafer, Konstan, and Riedl, 2001].

Recommenders are also a significant part of another prominent phenomenon in today’s online information environments, namely personalization [Sarwar et al., 2000; Linden, Smith, and York, 2003; Kumar and Benbasat, 2006; Anand and Mobasher, 2007]. Jeff Bezos, CEO of Amazon.com, crystallized the drive for

¹ http://www.flickr.com/
² http://www.youtube.com/
³ http://del.icio.us/
personalization in the modern information environments when he said that if he had
three million customers, he should have three million stores [according to Schafer,
Konstan, and Riedl, 2001]. Today, it is virtually impossible for two users to see the
exactly same pages with the exactly same content over any prolonged visit to Amazon.
The personalization starts immediately from the first choices that users make [Linden,
Smith, and York, 2003; Kumar and Benbasat, 2006]. For example, the books that we
are recommended depend on the items that we and other customers have viewed and
bought [Brusilovsky, 2007; Anand and Mobasher, 2007].

Because collective intelligence in the online information environments
depends on the user input, be it collected explicitly or implicitly [Dieberger et al.,
2000], the features work the better the more they are used [Resnick and Varian, 1997;
O’Reilly, 2005; Svensson, Höök, and Cöster, 2005]. The more is known about a
particular user and all the users, the better the environment can be tailored to the user.
However, the relentless collecting of user information has also led to justified privacy
concerns [Resnick and Varian, 1997; Kobsa, 2007; Story, 2008b].

Another aspect of Web 2.0 is social networking. Even if we are not talking
about purposely social networking services, such as Facebook⁴ or LinkedIn⁵, the
harnessing of collective intelligence has made other users an inextricable and inherent
part of our experience of information environments. Even the names of some features,
such as Amazon’s “Customers Who…” recommendations, directly imply the presence
of other users in the environment, even if we are not directly dealing with them at
personal level. The social texture generated by such features is thought to give users a
sense of social presence, of not being alone in the information environment [Svensson,
Höök, and Cöster, 2005]. Thus, information environments are becoming increasingly
social environments, and this affects our behavior in them.

As a result of the numbers of items and users growing to millions and even
billions, harnessing collective intelligence to assist users to find items of interest, and
the drive for personalization, among other factors, online information environments
have grown vastly in complexity. The ways to navigate information environments
abound and compete for the one thing that cannot increase: Our attention and other

⁴ http://www.facebook.com/
⁵ http://www.linkedin.com/
cognitive capacities. We as users of these environments have to cope with this increased complexity. [Juvina and van Oostendorp, 2004; Freyne et al., 2007]

Consequently, we call information environments that offer their users numerous features and alternative ways to use them and find items of interest and where the number of items introduces information overload, such as today’s e-commerce and other Web 2.0 sites, complex information environments. Users not only have to figure out which features work the best for their particular need and learn how to use and understand them but also need to understand how the features can act together and influence each other.

While many individual features that are employed in complex information environments today are results of meticulous and still on-going research, and others subject to great research interest, the user strategies to which these environments give birth have not been researched to the same degree of thoroughness. Consequently, we were interested to study what kind of strategies users have developed to cope with actual, existing complex information environments, how items of interest are being found today, and how users perceive the presence of other users in these environments.

In order to get a true view of the actual use, we used applied ethnography, in this case a combination of on-location interviewing and observation with verbal protocol. We chose Amazon, the world’s biggest online retailer, as the complex information environment for the study because it is the archetypical Internet store that has consistently been an early adopter and innovator of new e-commerce approaches [Kotha, 1998; Chevalier and Mayzlin, 2003; Kumar and Benbasat, 2006]. In particular, Amazon has used a wide array of recommender approaches [Kumar and Benbasat, 2006] in addition to search features and category-based approaches, thus constituting an interesting study ground for seeing how people find items of interest in complex information environments today.

In Amazon, we chose books as the item because books cannot be experienced online the way they are experienced in a brick-and-mortar store. For example, an mp3 song can be previewed online in the exactly same way as it is used. With other items, such as books or cameras, we have to decide based on second-hand information made

---

6 We use the term “Amazon” to refer to both Amazon.com and Amazon.co.uk, the sites the participants used in this study
available, notwithstanding such features as Amazon’s “Search Inside” as they do not show us the whole book or the binding.

Because of our qualitative research approach, our goal was to understand and describe the existing reality, and find patterns and strategies rather than prove any hypothesis right or wrong. In fact, we started with a research focus without any preconceived hypotheses. Our purpose was to summarize, describe, and formulate theories about user behavior in complex information environments based on what we observed and what the participants told us.

Because we wanted to observe the whole item-finding process in its entirety, our method limited us to six participants. This meant that we could only focus on big trends and not draw final conclusions about alternative approaches as the subgroups were by necessity very small. We had one 2-3 hours long observation-interview session with each participant. The sessions took place at the participants’ homes where they used their own computers.

The data that we collected enabled us to see clear trends in the item-finding process. Recommenders have become an integral part of the process. They are used both strategically and opportunistically in the item-finding process. Strategic use means that users know beforehand how they are going to use recommender(s) in the process while opportunistic use means that users use recommendations as they come upon them without a preconceived idea to use them in a particular way in the process.

While recommenders are now an important part of the item-finding process, keyword searching is also alive and well. In many ways, the co-existence of recommenders and query searching is a peaceful one. A process that begins as a keyword search might evolve into opportunistic use of recommenders while strategic use of recommendations might include query searching as a part of the process. A keyword search is a natural way to find items that we know while recommenders offer possibilities of discovery [Hangartner, 2007], and today users are combining the two in many creative and efficient ways. In contrast, searching by category appears to be dying out although the use of categories itself is still meaningful.

Item-specific recommendations, such as star-ratings on the list page, affect which items, be it books or reviews, some users pick for a closer scrutiny, and furthermore influence the overall perception. We also found some evidence that omnipresent Google colors how search-related features and the list page are perceived.
and used in other information environments. At the very least, Google raises the bar for other services by creating expectations.

The participants seemed to have two major approaches to finding items: Finding the best item and finding a good enough item. The approach that the participant took was reflected on many aspects of the process, such as tab use and need for comparisons. The participants appeared rather consistent in their choice of approach although we cannot declare based on such a small sample size that users consistently use one or the other approach.

The perception of the inherent sociality of complex information spaces divided the participants in two: Half of them felt that Amazon had social presence while the other half felt that the social texture was not enough for that. We also saw evidence that the actions of others made visible in the social texture affect the behavior of at least some users in complex information environments.

Some of the results discussed in this thesis were reported in [Leino and Räihä, 2007]. That paper focused on the use of recommenders in complex information environments. Here we review those results in greater detail and place them in the larger context.

The thesis is organized as follows. In Chapter 2, we look at the recommenders in Amazon and discuss related work. In Chapter 3, we discuss our study method and participants before moving on to discussing our results in detail in Chapter 4. In Chapter 5, we summarize our work and findings.
2 Background

2.1 Recommenders in Amazon

2.1.1 Overview

We are all more or less familiar with keyword/query searching and searching by categories. Query searching, be it Google’s front page\(^7\) or Amazon’s quick search, is an example of information retrieval system [Park and Pennock, 2007] while the way books are organized in libraries represents a pure category-based organization of information.

Recommender systems, on the other hand, are the new kid on the block. They have very quickly become so omnipresent in and integral to e-commerce sites [Mohan, Keller, and Ramakrishnan, 2006; Park and Pennock, 2007] that they have and still are re-shaping information environments [Schafer, Konstan, and Riedl, 1999]. Both the click-through and conversion rates, two important measures of effectiveness, show that as targeted marketing tools they out-perform untargeted content, such as top-seller lists [Linden, Smith, and York, 2003]. Nor is their use restricted to e-commerce. Recommenders are becoming integral parts of all kinds of information environments, as the amount of information keeps growing at an incredible rate and we increasingly need intelligent tools to search through it [Rashid et al., 2002; Kim, Kim, and Cho, 2008].

Recommenders are also an important part of personalization of information environments, of creating personalized user experience [Schafer, Konstan, and Riedl, 1999; Linden, Smith, and York, 2003], and Amazon’s personalization is essentially recommender-driven [Linden, Smith, and York, 2003; Kumar and Benbasat, 2006].

We look at recommenders here mainly from the viewpoint of how they are used in Amazon online store although we do discuss some relevant general aspects, too. We further restrict the discussion to those recommender features that the participants used, as those are of interest to us here. Thus, the discussion is not to be construed as a complete introduction to recommender systems or even to the

\(^{7}\)http://www.google.com/
recommender systems used in Amazon. In our discussion here, we use the recommender categories defined by Schafer, Konstan, and Riedl [1999; 2001].

The fact that Amazon has consistently been an early adopter and developer of new e-commerce approaches [Kotha, 1998; Kumar and Benbasat, 2006] also means that its interface is constantly undergoing many changes. In fact, Tim O’Reilly [2005] considers this constant updating and development cycle one of the hallmarks of Web 2.0 services.

Consequently, during the period between conducting the observation-interview sessions and analyzing and writing down the results, Amazon’s interface underwent several changes, some bigger and some smaller. Furthermore, Amazon.co.uk and Amazon.com interfaces have several differences. For instance, the “Customers Who…” recommender shown in Figure 1 is the current one from Amazon.com. During the sessions, Amazon.com did not allow adjusting the recommendations by categories but otherwise the interface in Amazon.com was as in Figure 1. In contrast, Amazon.co.uk interface to “Customers Who…” was text-based during the observation-interview sessions (Figure 2).

![Figure 1. Current “Customers Who…” in Amazon.com. At the time of the observation-interview sessions, recommendations could not be adjusted by categories in Amazon.com.](image1.png)

![Figure 2. “Customers Who…” in Amazon.co.uk at the time of the observation-interview sessions.](image2.png)
2.1.2 Delivering recommendations

Recommender systems typically suggest the user items that they predict to be of interest to that particular user, often in a particular context. While the recommendations can be given individually, they are often presented as lists of items, as they are in Amazon [Linden, Smith, and York, 2003]. They can also predict the rating the user would give an item, although Amazon does not do this. Instead of predicted rating and unlike MovieLens⁸, for instance, Amazon only gives the average customer rating (Figure 1) [Goy, Ardissono, and Petrone, 2007].

Furthermore, recommender environments can provide the ratings of individual items by individual community members. Often, these are accompanied by text comments, such as reviews. These text recommendations are typically given as raw data to other users to facilitate their decision-making process as do not lend themselves well to machine-summaries or personalization [Schafer, Konstan, and Riedl, 1999; Schafer, Konstan, and Riedl, 2001]. Users need to read the text and interpret it as positive or negative recommendation [Schafer, Konstan, and Riedl, 2001].

The recommendations can be delivered to users in various ways, using push technology, pull technology or organically [Schafer, Konstan, and Riedl, 2001]. Push technologies reach out to users when they are not actively interacting with the environment [Schafer, Konstan, and Riedl, 2001]. For instance, Amazon sends email recommendations to users who have not opted out. The recommendations include discounts, new releases, and seasonal offerings.

Pull technologies allow users to determine when they want recommendations [Schafer, Konstan, and Riedl, 2001]. Recommendations are displayed when users wish to see them [Schafer, Konstan, and Riedl, 2001]. For instance, Amazon provides a link to “Recommended for You” section that is not displayed unless the link is clicked.

Organic or passive delivery refers to recommendations being displayed in the natural context as part of the interface features [Schafer, Konstan, and Riedl, 2001]. For instance, on each item page, Amazon gives “Customers Who…” recommendation(s) (Figure 1) based on the currently viewed item. Another example

---

⁸ http://movielens.org/
of organic delivery is “Perfect Partner” recommendation that offers the current item together with another item. Although the interface strives to give an impression that “Perfect Partner” (Amazon.co.uk) involves discount, this is typically not the case (Figure 3). When the same feature is called “Best Value” in Amazon.com, there is a discount.

**Figure 3. “Perfect Partner” recommendation in Amazon.co.uk.**

In addition, Amazon has both “Reviews” and “Customer Reviews” in its arsenal of recommendations that the participants actively used. “Reviews” section is a rag-tag collection of item-related reviews, back-flab texts, and synopses, among others. This is where Amazon puts whatever material it has available on the item. “Customer Reviews” are exactly what the name says (although the word “customer” in Amazon.com means a user who has purchased an item while in Amazon.co.uk it means a user who has signed in). Customers can add text comments on the items and rate them on a 1-5 stars scale. In addition, users who are signed in can rate “Customer Reviews” by others as “Helpful” or “Not helpful.” Both ratings and reviews are non-personalized recommendations in that they are shown the same way to all visitors and thus do not adapt to each visitor [Schafer, Konstan, and Riedl, 2001].

Amazon.com’s customer reviews tend to be longer and more detailed than those of its arch-rival, Barnes & Noble’s. Furthermore, in 2003 only 12.6% of Amazon.com’s books did not have reviews while 54.2% of Barnes & Noble’s books did not have them. The median of the number of reviews was 11 on Amazon.com. On the whole, Amazon’s star rating and reviews are very positive, although not as positive as on Barnes & Noble. [Chevalier and Mayzlin, 2003]

Although Amazon boasts a plethora of other recommender features, such as “Listmania!” lists and “Continue shopping: Customers with Similar Searches Purchased”, the participants in this study did not use them or even pay any attention to them. Thus, they are outside of this discussion.

---

9 [http://www.barnesandnoble.com/]
One important aspect that is receiving increasing attention is the transparency of the recommenders [Herlocker, Konstan, and Riedl, 2000; Sinha and Swearingen, 2002; Goy, Ardissono, and Petrone, 2007]. While many recommender systems are black boxes that spew out suggestions and predicted ratings at users without any explanation as to why the item was recommended, Amazon has clearly strived to provide users with the logic behind the recommendations [Goy, Ardissono, and Petrone, 2007]. For instance, when we receive promotional emails from Amazon, they tell us why the item was recommended: “We’ve noticed that customers who have purchased or rated books by … have also purchased … by …. For this reason, you might like to know that … will be released on ….”

Likewise, the names of features, such as “Customers Who Bought/Viewed This Item Also Bought/Viewed,” tell us about why the items are recommended. In case of “Personalized Recommendations” (“Recommended For You”), we also get a short overall explanation, “Recommendations for you are based on items you own and more,” and an explanation under each recommended item why that particular item

---

10 From a promotional email sent to the author by Amazon.co.uk.
was recommended: “Recommended because you purchased…” In addition, Amazon allows us to “fix” recommendations by marking explicitly that we do not want to be recommended items based on certain items. Figure 4 shows the “Recommended For You” page with the “Fix it” window open for the first item in the recommendation list.

Transparency, or at least perception of transparency [Sinha and Swearingen, 2002], is important for building trust and confidence, and allowing users to evaluate a recommendation to be able to make an informed decision about it [Herlocker, Konstan, and Riedl, 2000; Adomavicius and Tuzhilin, 2005; Goy, Ardissono, and Petrone, 2007]. Users appreciate transparency for both new items and items that they already know [Sinha and Swearingen, 2002]. However, the explanations have to be succinct, as initial research suggests that over-involved explanations leave users feeling overwhelmed [Goy, Ardissono, and Petrone, 2007]. Overall, Amazon’s explanations are short and to the point.

2.1.3 Item-based collaborative filtering

There are various approaches to algorithmic recommendations. Collaborative filtering-based recommenders have proven both effective and popular [Herlocker et al. 2004]. They divide into two types, item-based (item-item) and user-based (user-user) [Sarwar et al., 2001]. Both form neighborhoods of either users or items (hence the names) by computing the similarity between all pairs. Predictions for individual item ratings are formed by aggregating the ratings of the user’s neighbor users for the target item (user-based) or the user’s ratings of the item’s neighbor items. [Mohan, Keller, and Ramakrishnan, 2006]

Amazon uses item-based collaborative filtering for various reasons [Linden, Smith, and York, 2003]. First, it is less sensitive to the sparsity problem [Sarwar et al., 2001; Linden, Smith, and York, 2003]. Sparsity refers to a situation where the proportion of missing ratings over all the possible ratings is high. Consequently, neighborhoods are hard to build and so recommendations either cannot be made or are of low quality. If we consider Amazon’s situation, a user who would have purchased just 0.5% of books would have had to purchase close to 90 000 books [Sarwar et al., 2001].

The sparsity problem is compounded by the new-user problem. A new user has not purchased or rated any items, and so forming neighborhoods is impossible. User-based approach would require the new user to rate certain number of items. For
instance, in MovieLens, new users have to rate 15 movies before they are given recommendations to guarantee the quality of the recommendations. Item-based approach, on the other hand, produces “high-quality recommendations based on as few as two or three items” [Linden, Smith, and York, 2003].

In addition, item-based approach generates consistently slightly better predictions [Sarwar et al., 2001; Linden, Smith, and York, 2003]. However, because the improvement is not significant, other factors can determine which approach to use.

Another significant reason for Amazon to opt for the item-based approach is that the user-based approach has a serious performance bottleneck because it has to calculate the neighborhoods online [Sarwar et al., 2001; Linden, Smith, and York, 2003]. With 9335 movies (March 4, 2008) and merely tens of thousands of users, MovieLens can do fine with user-based collaborative filtering, but for Amazon with tens of millions of users and millions of items, it is simply computationally too expensive [Linden, Smith, and York, 2003]. With the item-based approach, the similar-items matrix forming calculations can be performed off-line, which makes it very scalable [Sarwar et al., 2001; Linden, Smith, and York, 2003]. The online computation in Amazon depends only on the number of items purchased and rated, which is never too high for instantaneous response [Linden, Smith, and York, 2003].

2.1.4 Collecting user preferences

As a rule, the more is known about the user’s preferences, the better recommenders perform [Svensson, Höök, and Cöster, 2005]. User preferences can be collected implicitly or explicitly [Dieberger et al., 2000; Schafer, Konstan, and Riedl, 2001; Fox et al., 2005]. Implicit collecting refers to collecting preferences without the user having to do anything other than use the application. Purchases and items viewed, for instance, are used to generate preference information. The main problem with this approach is that we need to interpret what different actions mean. Does the fact that somebody views or buys a book mean that he or she actually likes it? However, implicit collecting is effort-free for the user.

Explicit collecting means that users have to give preferences actively, for instance by rating items. It is also called very descriptively “hand-raiser data” [Story, 2008a] because people have to do something, for instance type or click, to give it. While explicit preferences are generally considered more reliable, they do have some reliability problem, too, for instance the difference between the true preferences and
the preferred self-image or the image that we wish to present to others [Gadanho and Lhuillier, 2007]. Furthermore, explicit collecting of preferences means that users have to be active before they can enjoy the recommendations, and motivating users and thus getting the preferences has proven challenging [Resnick and Varian, 1997; Cosley et al., 2005; Fox et al., 2005].

Not surprisingly, research community has taken great interest in devising ways to improve the quality of both implicit and explicit gathering of preferences and to lower the amount of effort required from new users [for instance, Rashid et al., 2002; Fox et al., 2005; Gadanho and Lhuillier, 2007; Rubens and Sugiyam, 2007].

Amazon uses both methods of preference collecting. Users are invited to rate items, report which items they own, and to contribute “Customer Reviews,” among many other explicit means of gathering preferences. At the same time, Amazon records for instance what items the user views and purchases, and uses these to generate recommendations.

In addition to new users not having preference profiles, many returning users do not necessarily log in to identify themselves. For instance, in our study, only two participants were logged in, as the rest typically only logged in when they wanted to pay for the items. Consequently, Amazon has developed means suitable for making recommendations also for customers who are not logged in [Kumar and Benbasat, 2006].

The relentless gathering of potentially sensitive information continues to raise serious privacy concerns [Resnick and Varian, 1997; Schafer, Konstan, and Riedl, 2001; Fox et al., 2005; Anand and Mobasher, 2007; Story, 2008b]. Privacy International [2007], an England-based human rights group “formed in 1990 as a watchdog on surveillance and privacy invasions by governments and corporations”, states that Amazon’s privacy practices have “notable lapses”: “Amazon has improved much over the years but consumers should be informed on how their clicking, reading, and purchase habits are profiled and used.”

Kobsa [2007] reports that a vast majority of Internet users are uncomfortable about being tracked across site, and over half of them are concerned about being tracked at all [Berendt and Teltzrow, 2005]. While Internet users have not widely complained about tracking, it might simply be because they do not understand the implication of recording their “hopes, worries and fears” [Rotenberg, executive director of the Electronic Privacy Information Center, according to Story, 2008b].
However, the trend today appears to be that of collecting as much data as possible even if it has no current use and keep it in stock [Kobsa, 2007]. While privacy issues are not the focus of this study, their importance in the drive for personalization should not be underestimated.

Not are privacy issues the only dicey factor in the equation. Recommender systems and preferences data also allow other ethically questionable measures to be taken, including pricing measures to maximize “the lifetime value of the customer to the site” [Schafer, Konstan, and Riedl, 1999]. In fact, Amazon used the information it had collected for price targeting, that is, for charging higher prices from people who had shown themselves to be price-insensitive and lower prices from those whose behavior had marked them as price-sensitive. What the customer had bought did not only determine what they were offered but also at what price. Amazon’s strategy was revealed and as a result of public outcry, it has now promised not to resort to this type of price targeting anymore. [Harford, 2006]

All in all, the privacy and ethical questions related to recommenders and personalization in the complex information environments are likely to haunt us for a long time to come [Schafer, Konstan, and Riedl, 1999].

2.2 Related work

2.2.1 Overview

Studying complex information environments is challenging for a variety of reasons, no matter what kind of methodology is used [Kumar and Benbasat, 2006]. Furthermore, the environments need to be used by numerous users for a prolonged time before they even start to function properly [Svensson, Höök, and Cöster, 2005]. This causes building complete information environments, an onerous task in itself, to be of questionable value as finding suitable participants to use the environment for a prolonged period is also complicated [Kumar and Benbasat, 2006]. Most of the existing complex information environments, such as Amazon, function as businesses and are not open to researchers because of trade secrets. However, the commercial systems tend to be superior to the systems that researchers have resources to build [Kumar and Benbasat, 2006].

Consequently, a significant amount of the research on recommender systems centers on MovieLens, a fully functioning movie recommender system developed and
run by GroupLens Research at the University of Minnesota. Even researchers not connected to GroupLens directly often use its datasets [for instance, Rubens and Sugiyam, 2007; Anglade, Tiemann, and Vignoli, 2007; Nguyen, Denos, and Berrut, 2007] that GroupLens researchers have generously made public. While MovieLens does incorporate searching features, it is still to a large extent a recommender environment rather than a fully-fledged complex information environment. Thus, while ideal for studying different aspects of recommenders, it is not an ideal environment for studying complex information environments and the strategies that emerge there.

Much of recommender research has focused on algorithms and different accuracy metrics to measure them [Herlocker et al., 2004], perhaps even to the detriment of the whole field [McNee, Riedl, and Konstan, 2006]. Currently, the focus is enlarging to embrace different areas, such as user-centric perspectives and interface issues [McNee, Riedl, and Konstan, 2006]. However, much of the research still focuses on different aspects, and not on the whole picture. In contrast, our study clearly represents user-centric perspective to the complex information environments, and it strives to offer a picture of how these environments are actually used and what part recommenders play in the whole.

Customer/user reviews are being studied increasingly as their importance in the online shopping environment is undeniable [Chevalier and Mayzlin, 2003; Chen, Dhanasobhon, and Smith, 2006; Kumar and Benbasat, 2006; Hu, Liu, and Zhang, 2007; Gretzel, Yoo, and Purifoy, 2007]. However, few studies have focused on how they are used in decision-making [Ludford, 2007] or what role they play in the whole process. As with algorithmic recommender systems, customer reviews have mainly been studied in isolation of the environment and process context.

Perhaps the most significant effort on studying user behavior in complex online information environments to date is Kalas, a social navigation system for food recipes [Riedl and Dourish, 2005]. Kalas was custom-built for the study and had 302 users for six months [Svensson, Höök, and Cöster, 2005].

Social presence research has to a large extent focused on social presence as a precedent to trust and loyalty in e-commerce [Gefen and Straub, 2004; Cyr et al., 2007]. As lack of trust is seen as one of the greatest hindrances to the growth potential of e-commerce [Gefen, 2000; Gefen and Straub, 2004; Grabner-Kräutera and Kaluschaya, 2003], this is hardly surprising. However, our focus is more on how users
perceive social presence in complex information environments that are not specifically created for social interactions; whether the social texture in Amazon creates a perception and experience of other users being psychologically present [Hassaneina and Head, 2007], and how that affects users’ behavior.

Next we look at the research work on recommenders in more detail, first on algorithmic recommenders and then on customer reviews, before discussing the work done on complex information environments. Finally, we take a short look at the social presence research before moving on to discussing the method and the results of this study.

2.2.2 Algorithmic recommenders

As mentioned, recommenders have become an integral and crucial part of e-commerce [Schafer, Konstan, and Riedl, 1999; Sarwar et al., 2000; Mohan, Keller, and Ramakrishnan, 2006] and their benefits are increasingly seen in connection to offline commerce, too [Linden, Smith, and York, 2003]. Given the vast numbers of items, recommenders are important both in suggesting us items and predicting how and why we would like them [Schafer, Konstan, and Riedl, 2001; Mohan, Keller, and Ramakrishnan, 2006]. Recommender systems have produced business strategies and approaches online that could never exist in the physical world [Schafer, Konstan, and Riedl, 2001], and e-commerce is likely to continue to drive their development [Hangartner, 2007].

As we have already discussed briefly privacy issues and ethical issues as pertaining to recommenders, we do not return to them here. Instead, we focus here on the directions that the recommender research is taking with special emphasis on the user issues as that is where this study fits as far as recommenders are concerned.

Since the beginning of recommender systems in the days of Tapestry [Goldberg et al., 1992], the first recommender system developed [Terveen and Hill, 2001] by researchers at Xerox PARC, recommender research has heavily emphasized the development of better algorithms [Herlocker et al., 2004]. The “goodness” of algorithms has typically been measured by various accuracy metrics, such as predictive accuracy metrics (usually measured with mean absolute error), classification accuracy metrics, and rank accuracy metrics [Herlocker et al., 2004]. Additionally, the algorithms have more recently been evaluated by non-accuracy metrics, such as coverage, “a measure of the domain of items in the system over
which the system can form predictions or make recommendations,” and learning rate, how much learning data is necessary for the algorithm to start producing good predictions [Herlocker et al., 2004].

However, researchers are increasingly waking up to the fact that the emphasis on algorithms and accuracy metrics, while certainly important, has left the user outside of the equation when the user should in fact be in the center of it [McNee, Riedl, and Konstan, 2006]. Users do not simply want accurate recommendations; they want useful, actable recommendations [Herlocker et al., 2004].

Consequently, the researcher community is beginning to emphasize that using recommenders should constitute a pleasant user experience [McNee, Riedl, and Konstan, 2006]. The strength of recommenders should be supporting discovery [Hangartner, 2007] by recommending novel, serendipitous items. The emphasis on accuracy that tends to result in both predictable suggestions and lack of diversity [Herlocker et al., 2004] has been detrimental to the user experience [McNee, Riedl, and Konstan, 2006].

The study of algorithms continues [Zhang and Pu, 2007], as it should. The algorithms need to become more robust [Mehta, Hofmann, and Nejdl, 2007] (less sensitive to shilling and other attacks that attempt to bias the recommendations [Anand and Mobasher, 2007]), deal with cold-start situations [Nguyen, Denos, and Berrut, 2007], and serve under various use contexts, as in some use situations producing some good items is enough while in another situation it might be necessary to find all the relevant cases [Herlocker et al., 2004]. For instance, a lawyer needs to get all the relevant precedents while we only need a few good recommendations to select a video for renting [Herlocker et al., 2004].

Additionally, we need to refine the ways we collect user preference, whether it is done implicitly or explicitly. Personalization is data driven [Anand and Mobasher, 2007]: Without good user data on which to base the recommendations and personalization, the result is not going to provide a compelling user experience. Research efforts try to refine different aspects of data collecting, such as finding out which items to have users to rate explicitly to produce the widest benefit for the quality of recommendations [Rubens and Sugiyam, 2007] and how to evaluate the meaning of implicitly collected behavior data [Fox et al., 2005; Gadanho and Lhuillier, 2007].
However, next to algorithm research, there is a growing interest in the user side of the equation. Transparency, explaining why certain items are recommended, has attracted wide research interest [Sinha and Swearingen, 2002; Goy, Ardissono, and Petrone, 2007]. Cosley et al. [2003], on the other hand, studied how the interface affects the ratings that we make, providing interesting insight into how to design rating interfaces. Furthermore, several researchers have studied the roles of individual users in the making of the recommendations. Rashid, Karypis, and Riedl [2005] studied influence, “the effect of a user on the recommendations from a recommender system,” and Mohan, Keller, and Ramakrishnan [2006] studied what roles different users had inside the system, identifying such roles as scout, promoter, and connector.

Consequently, recommender research is becoming increasingly user-centric and user-factors are beginning to receive the attention they deserve. However, the research is still characterized by and large by attention to some individual aspect of the whole. There is little research on the whole, how users use the recommendations as a part of the whole use process in complex environments. While such projects as MovieLens provide us with important insight into recommender systems, they do not really allow us to study the use of recommenders in the context of complex information environments of today. The only in-depth study of this kind that we know of is Kalas, which is discussed in Section 2.2.4.

In this sense, our study is among the first steps towards understanding the use of recommenders in the context of complex information environments and towards understanding how users use these environments and cope with the plethora of possibilities afforded by such environments. In addition, we shed light on the item-finding strategies that emerge in such environments.

### 2.2.3 Customer reviews

While customer reviews are considered recommendations [Schafer, Konstan, and Riedl, 1999], they are clearly different from the algorithm-based recommenders discussed above, and thus merit a separate discussion. Online user reviews today constitute an increasingly important source of information to consumers and are both substituting and complementing other forms of word-of-mouth communication [Chevalier and Mayzlin, 2003; Chen, Dhanasobhon, and Smith, 2006; Kumar and Benbasat, 2006; Hu, Liu, and Zhang, 2007; Gretzel, Yoo, and Purifoy, 2007]. In fact, Forrester found in 2006 that over 80% of online shoppers used customer reviews
[according to Gretzel, Yoo, and Purifoy, 2007]. As the link between word-of-mouth and sales has repeatedly been shown to be significant [Hu, Liu, and Zhang, 2007], many managers believe that providing such community content is important for building loyalty [Chevalier and Mayzlin, 2003].

A famous anecdote highlights the power of reviews and reviewers: The phenomenal success of Da Vinci Code has been partially attributed to one Amazon reviewer, “Francis J. Mcinerney”, then a Top Ten Reviewer [Paumgarten, 2003].

Customer review format remains rather standard across sites [Ludford, 2007]: The text comments are typically accompanied with a rating, such as Amazon’s five-star scale star rating, and potentially a way for other users to mark the review as helpful (or useful) or not helpful. Furthermore, reviews are typically presented in a serial order with limited sorting capabilities [Ludford, 2007]. This was certainly the case in Amazon when our study was conducted although Amazon’s current version boasts many significant improvements.

This way of presentation is problematic because users have been found not to read all customer reviews, only ones on the first or perhaps second page [Brynjolfsson, Dick, and Smith, 2003]. Information overload is in works here, too, as some books in Amazon can have hundreds of reviews. Consequently, the presentation does not support human decision-making or common sense, domain specific decision-making practices [Ludford, 2007]. There is evidence that humans weigh negative information differently from positive information—negative information weighs in more heavily [Chevalier and Mayzlin, 2003; Ludford, 2007]—and so the order in which the reviews are shown matters. As only the first one or two pages are read, an uncharacteristic set can mislead users [Ludford, 2007]. However, the presentation order is typically based on recency, which leaves the representativeness of the reviews that users typically see more or less up to chance [Ludford, 2007].

While attempts have been made to remedy this situation with helpful ratings of the customer reviews, the time required to collect this data causes challenges. In addition, as users typically only read the topmost ones, they also tend to rate them, thus creating a vicious circle. A further problem with the helpful rating approach is that some information is time-sensitive. For instance, a hotel may have been renovated since the review was written, and the previously shabby lounge may now be luxurious. [Ludford, 2007]
Nevertheless, Chen, Dhanasobhon, and Smith [2006] found that reviews with high proportion of helpful votes in Amazon create additional sales, the effect being more evident on non-popular books. They collected data daily for 195 days on product sales levels and customer reviews on 535 books from Amazon.com’s web pages. However, they were unable to confirm that users actually considered the helpful votes in the item-selecting process, and could only show that books with reviews voted helpful sold better, not that the helpful votes affected the sales _per se._

The ratings by other users are often given to the user as an arithmetic average, as on Amazon’s list page and on top of the item page. Majority of the reviews, however, give polarized opinions, rating a book with 1-2 stars of 4-5 stars. Consequently, the distribution can be “U” or “J” shaped while the average may be 3 stars, showing a medium user experience and thus misleading the user. [Hu, Pavlou, and Zhang, 2006; Hu, Pavlou, and Zhang, 2007]

Furthermore, an arithmetic mean does not summarize information the way humans do, thus making it a flawed guide for human decision-making [Ludford, 2007]. Ludford shows that the last read item influences our perception heavily and that negative items influence us even more heavily. In addition, she conjectures that the personally meaningful features in the reviews weigh heavily in the assessments and that this might go down to a sentence or phrase level. Our study confirms these conjectures and supports her conclusion that an arithmetic average does not summarize the reviews as human users do.

All in all, however, the relationship between item ratings and sales is somewhat inconclusive, as some studies have found positive correlation between the two while others have concluded that higher ratings do no necessarily lead to higher sales [Chen, Dhanasobhon, and Smith, 2006]. This argues that the averages or even the ratings themselves cannot capture all the dimensions of reviews, and the contents of and other cues in the customer reviews color the perception of the rating and its importance [Chevalier and Mayzlin, 2003; Chen, Dhanasobhon, and Smith, 2006; Gretzel, Yoo, and Purifoy, 2007; Ludford, 2007]. Our study certainly points to this direction. Furthermore, customer reviews do not exist in a vacuum but interact with other factors available to users in the decision-making process [Ludford, 2007].

Ludford [2007] found that users considered shorter reviews less useful. In contrast, Chevalier and Mayzlin [2003] found that longer reviews tended to have a negative impact on sales. They conjectured that the reason was because longer
reviews tended to be less enthusiastic about the item. For instance, 4-star reviews in Amazon.com were 849 characters long on average while 5-star reviews were 796 characters long on average [Chevalier and Mayzlin, 2003]. Thus, it appears that the contents of the reviews determine their impact [Chevalier and Mayzlin, 2003; Ludford, 2007].

In spite of common conjectures that reviewer reputation is important [Hu, Liu, and Zhang, 2007], Chen, Dhanasobhon, and Smith [2006] found that the reviewer reputation in fact is not a factor in the purchase decision. Considering how widely reputation systems are studied in this context, this finding is important. Our study indicates that the reviewer is evaluated on the spot, based on the information available on the review, such as language and content, rather than by trying to find out more about the reviewer. Further research is necessary here, as this is one of the central questions as far as future development is concerned.

One problematic aspect of customer reviews is that it is very easy to “free ride” on them [Chevalier and Mayzlin, 2003]. Google can take us directly to reviews on a site and we can read the reviews there only to return to complete the purchasing on another site, as one of our participants reported doing repeatedly: He used Amazon’s “Customer Reviews” but often bought books from another online bookstore after this. Furthermore, during our study, some participants went to other sites for reviews when not enough information was found on Amazon.

In spite of all the problems, however, Chevalier and Mayzlin [2003] report that customers behave as if customer reviews did help them find better-fitting books for themselves. Gretzel, Yoo, and Purifoy’s [2007] survey of TripAdvisor11 users showed that over 80% of the respondents felt that such reviews increased their confidence in their decisions, made it easier to reach a decision, and helped them plan trips more efficiently.

Clearly, customer reviews have a great potential for helping users select the right item(s) from the plethora of choices. However, this potential cannot be fully realized when the information overload affects the reviews as well as the items themselves. Too large set of customer reviews cannot be effectively used if we do not find a way to personalize and summarize the reviews to users. [Ludford, 2007]

---

11 http://www.tripadvisor.com/
One area of intense research effort is also how to get users to contribute reviews [Cosley et al., 2005; Gretzel, Yoo, and Purifoy, 2007]. Gretzel, Yoo, and Purifoy [2007] conclude that the motivations are to a large extent intrinsic and that rewards do not seem to have much impact on the willingness to contribute. Ozakca and Lim [2006] found that the motivations vary greatly. Often, a review was written if the item was exceptionally good or bad, thus giving rise to “U” and “J” distributions. The reasons not to contribute ranged from lack of time to self-professed laziness.

In spite of intensive research interest on reviews, only few studies have focused on how customer reviews are used in decision-making [Ludford, 2007]. One that is similar to our study as far as customer reviews are concerned is a study by Ozakca and Lim [2006], who observed and interviewed six students using customer reviews and ratings in their prototype site. They found that ratings affected more user selection when the users did not have prior information, such as recommendations by friends. However, contents of the reviews and genre of the item (for instance in movies and books) affected the equation greatly, and high rating (or score) was no guarantee of getting selected. However, a low score with no prior knowledge did clearly reduce an item’s chances of getting selected. Consequently, Ozakca and Lim [2006] conclude that rating scores and textual reviews are complementary. Users want to see the reasons for the rating, and not just the rating. In addition, the item itself did affect the equation.

Moreover, Ozakca and Lim [2006] found that users did not find reviews by professional writers more valuable than end-user reviews. Users felt that both were useful, albeit differently so.

Our findings complement and deepen those by Ozakca and Lim [2006]. Our similar methods appear to have produced similar results. As in many other areas of research, many other studies used other methods than user observation, ranging from omnipresent surveys [Gretzel, Yoo, and Purifoy, 2007] and existing data based methods (using data available on the Amazon site) [Chevalier and Mayzlin, 2003] to portfolio methodology (comparing “the purchase of an item from Amazon with the help of online review vs. that of a consumer buying /selling [sic] a stock with the help of analyst forecast revision”) used by Hu, Liu, and Zhang [2007]. Some differences in results are likely to be explained by different methods used.
2.2.4 Complex information environments

Many user-related studies use students, often computer science students, to play the part of users [Sillence and Briggs, 2004], and many studies that boast big numbers of participants are questionnaire studies where the say-do problem cannot be ruled out [Kobsa, 2007]. The same applies to studies on trust in the context of e-commerce, which additionally are often plagued by “scales that are neither theoretically derived nor rigorously validated” [Grabner-Kräutera and Kaluscha, 2003].

In many ways, such research approaches are understandable, as recruiting people from real user groups is time consuming and motivating actual users to continue using a system can also be a problem in longitudinal research [Svensson, Höök, and Cöster, 2005]. Furthermore, building complex information environments purely for research purposes is a slow and expensive undertaking. Money does dictate research by setting limits on what can be done [Grabner-Kräutera and Kaluscha, 2003]. However, the fact that such practices are common and understandable does not mean that they do not affect the results.

There are not that many studies that have studied complex information environments as organic, integral wholes. Most studies use laboratory settings and focus on parts. The study of wholes has been neglected, although the complex whole is what users face and have to cope with [Juvina and van Oostendorp, 2004; Freyne et al., 2007].

Consequently, Kalas stands head and shoulders above others in studying actual user behavior in complex information environments. The discussion of Kalas here is based on [Svensson, Höök, and Cöster, 2005] unless otherwise indicated.

The accomplishments of Kalas researchers are impressive. First, Kalas, a social navigation system for food recipes that was built purely for research purposes, is “one of the most complete social navigation systems ever built” [Riedl and Dourish, 2005]. Secondly, they had 302 genuine users use the system for six months, out of whom 73 filled in a questionnaire and four were interviewed in-depth over telephone. The interviews, however, were not to collect additional data but “to contextualize and better understand the findings in the postquestionnaire and conclusions drawn from log data” [Svensson, Höök, and Cöster, 2005]. The use was organic use, as there were no tasks. The study was conducted in collaboration with a Swedish online cooking portal, hemma.net, which provided the researchers with over 3000 recipes.
The 302 users, recruited through hemma.net, were 30-50 years old, mostly women living in smaller cities. There are some indications that women seek more engagement while men appear more utilitarian, and thus we cannot safely generalize across genders [Cyr et al., 2007]. In contrast to Kalas, we had all-male participants, and this might go some way to explain some small differences in results, especially when it comes to perceived sociability of the environments.

Kalas offered users many different features. The algorithmic recommender system recommended recipes to the users, and the users were able to enter text comments about the recipes as well as evaluate them by voting “thumb up” or “thumb down.” Furthermore, the interface enabled users to see where other users currently were—real-time presence—and enabled chatting. Additionally, Kalas offered users many ways to see traces of other users, thus providing a rich social texture. [Svensson, Höök, and Cöster, 2005; Riedl and Dourish, 2005]

Nevertheless, even the Kalas experiment did suffer from the small number of users. There were not enough users logged in simultaneously to create a sense of real-time co-presence, and consequently chances for chatting were few. Only 12 users chatted or attempted to chat. Furthermore, from the statistical perspective, the recommender system had too few votes—4.77 per user on average—to produce reliable predictions. This underlies the inherent difficulties of studying complex information environments.

The Kalas researcher team found that explicit searches and recommendations had complimentary roles, as none of the 20 most popular recipes the users chose for using from the recommendations and from explicit searches were the same. Hangartner [2007] considered that searches are for finding items that we know—or perhaps of which we know something—while recommendations are for discovery. While we do not have enough information to say that this is one of the reasons why searching and recommendations in Kalas had complimentary roles, it is an interesting idea that we discuss in further detail in our discussion of our results of search and recommender use.

In Kalas, the decision of printing or downloading a recipe depended on many factors, such as ingredients, cooking instructions, and picture of the dish, in addition to the social texture consisting of the author of the recipe, a figure for the number of downloads, and user comments. Kalas users indicated clearly that reviews—user comments—were important to them in spite of their low number in the system.
However, the importance of the social texture to the decision of choosing a recipe was determined from the questionnaire replies. This leaves the conclusions open to the say-do problem, as the importance is determined only by what people said.

The Kalas research team speculated that those users who did not reflect on the meaning of the thumbs-up symbol in recommendations might have been influenced by them subconsciously. Out of 73 respondents to the questionnaire, less than half had understood the basis of the recommendations, and so the researchers speculated that one reason for this might be the fact that the recommender system used in Kalas did not require users to first rate items before receiving recommendations, as in MovieLens, for instance. Forty (out of 73 respondents) liked the recommender feature, and so recommendations were seen to encourage and inspire people to study the information environment.

While users liked the idea of being able to comment on the recipes, in fact only 11 comments were entered. Thus, motivating users to contribute directly proved difficult even though the concept itself was well-received. However, the research method prevented the researchers from making any further findings concerning the use of reviews.

In many ways, our study method makes our study complementary to Kalas, as they had exactly the kind of quantitative data—postquestionnaire data (“a mixture of Likert-scale questions and statements to agree/disagree with” [Svensson, Höök, and Cöster, 2005]) and log data—that we did not have. Instead, we had qualitative data from observations with verbal protocol and in-depth interviews that allowed us to look at the other side of the coin. While both the Kalas research team and we used in-depth interviews, the Kalas team only used them to gain insight into the quantitative data while we contrasted the data from them with observation to gain deeper insight into the whole process, thus explicitly to get data. Additionally, using purely log data denies any insight as to why the participants behaved in a certain way while verbal protocol allowed us to hear at least parts of what the participants were thinking.

Thus, Kalas can tell us what happened in the information environment, what features were used and what information was accessed (log data), and how the users perceived the environments (questionnaire) while we can add insight into how the environment was used and why it was used that way.

While the Kalas study was both in-depth and longitudinal, ours was in-depth but not longitudinal. Consequently, while the Kalas research team did not need to use
tasks, we were forced to give the users tasks to make sure that we get relevant data in the limited time that we had for each session. In addition, our method of applied ethnography limited us to six participants, which pales next to the 302 users of Kalas.

As to the future of Web 2.0 and complex information environments it has brought along, Rick Hangartner [2007], Chief Scientist of MyStrands\textsuperscript{12}, stated in a guest column in Msearchgroove.com that in the future, search engines will help us find what we know we are looking for while discovery helps us to find the rest. He thinks that eventually the recommender industry will grow more pervasive and sophisticated than the search industry. Searching will continue existing next to recommenders but it will incorporate more features from recommenders in future. This view appears rather prescient in light of our findings about the current use of recommenders.

Hangartner [2007] consequently thinks that we need increasingly refined ways to understand how people both explicitly and implicitly signal their individual needs that they do not necessarily fully understand themselves. Interestingly, the current drive to improve the implicit gathering of preferences [Fox et al., 2005] is close to an economic theory of “revealed preferences”, according to which people reveal their preferences with their choices as consumers [Harford, 2006]. Our subjective values are revealed in what we actually do [Harford, 2006]. While this is not the focus of our study, we contribute to it by increasing the understanding, the whys, of user strategies and behavior.

Overall, to our best knowledge, our study was the first in-depth study of the actual use of complex information environments that used applied ethnography, and thus used qualitative data to understand the whats, hows, and whys of user behavior in such environments.

\subsection*{2.2.5 Social presence}

Lack of consumer trust has long been seen as a major obstacle to the growth of e-commerce [Gefen, 2000; Hassaneina and Head, 2007], and understandably much research effort has been focused on how to increase trust in e-commerce. Social presence is widely considered one of the antecedents of trust [Gefen and Straub, 2004;\textsuperscript{12}]

\begin{footnotesize}
\textsuperscript{12} http://www.mystrands.com/
\end{footnotesize}
Hassaneina and Head, 2007], and it is also seen as increasing enjoyment and loyalty, thus leading to “more favorable consumer attitudes” [Hassaneina and Head, 2007].

This is hardly surprising since humans thrive on social interaction [Hassaneina and Head, 2007] and have an innate need for connectedness [Rettie, 2003]. Thus, that we approach our environment socially and build trust through social interactions should come as no surprise.

In fact, research has shown that forming social relationships is an important goal of Internet users and that providing support for virtual communities that facilitate connections between the site’s customers creates value for the online store. For instance, one study found that the customers who used the community features accounted for one-third of the visitors but generated two-thirds of the sales. [Kumar and Benbasat, 2006]

Many researchers see the challenge for e-commerce to be that online stores have typically low social presence, which stifles its growth by preventing trust formation and lowering enjoyment [Kumar and Benbasat, 2006; Hassaneina and Head, 2007]. In the past, the focus has been on the transactional relationship between websites and their customers while the current research is increasingly pointing to the importance of social relationship in the online environment [Kumar and Benbasat, 2006].

However, the concept of social presence is not well defined in the literature and various, occasionally conflicting definitions exist. For instance, the concept has been used to refer to as both “a property of the medium in a mediated communication” and as “the perceptions, behavior or attitudes of the participants in a mediated interaction.” [Rettie, 2003]

In the Kalas study, social presence was defined as a perception of “not being alone in the space” [Svensson, Höök, and Cöster, 2005]. The study showed that most users interpreted the signs indicating social presence, the social texture of the interface, correctly, and that such texture aided users in their choices and gave them confidence in their choices. Another good definition of social presence is “the extent to which a medium allows users to experience others as being psychologically present” [Hassaneina and Head, 2007; Cyr et al., 2007]. Furthermore, we are talking about perception of presence, be it synchronous or asynchronous.

In addition to the above definitions, our definition of social presence follows closely Dourish and Bellotti’s [1992] definition of awareness as “an understanding of
the activities of others, which provides a context for your own activity.” Rather than emphasize “warmth” [Hassaneina and Head, 2007; Kumar and Benbasat, 2006], “intimacy” [Rettie, 2003; Kumar and Benbasat, 2006] and “immediacy” [Rettie, 2003], we define social presence as being aware of other people acting in the information environment, both synchronously and asynchronously, by seeing signs that they have left in the environment, and consequently feeling that others exist in the same space and are present by the marks that they have left in it. Some of these marks are useful for our task-context, perhaps as recommendations of items of predicted interest or as knowledge of what is “hot” in the community, as in Technorati’s “Where’s the Fire? What’s Hot, and Why,” when looking for interesting blogs. Tagging and other newer approaches widen the possibilities, and simultaneously add to the complexity of the environments.

In a sense, our concept of social presence is comparable to walking through a city street late in the evening. We see cars that people have parked for the night and see lights in some windows. We see that grass has been cut or footprints on the snowy sidewalk and tire marks on the road. We are conscious of other people being around because we see signs of them everywhere even if we do not see any actual human beings at the very moment.

Raento [2007] suggests that the mere knowledge of others using the system can produce feelings of not being alone even if they do not engage us in any meaningful way. This could lead to the emergence of space where the system becomes a locale that is co-habited [Raento, 2007] or, with social texture and personalization, even a place where people interact [Kumar and Benbasat, 2006].

Another complication that trust researches face as far as social presence is concerned is that while interactions typically take place between humans, in an online store the user is interacting with a website [Gefen and Straub, 2004], and not with the personnel of the shop, which adds human touch and socially rich face-to-face interactions in brick-and-mortar stores [Cyr et al., 2007]. However, shopping is seen as social [Kumar and Benbasat, 2006], and consequently, online shoppers are seen as craving for “socially rich experiences” [Cyr et al., 2007]. Overall, adding social presence is seen as a way to overcome the challenge of online stores being impersonal.

---

13 http://www.technorati.com/
and socially poor environments by adding a feeling of warmth and sociability [Cyr et al., 2007; Hassaneina and Head, 2007].

Somewhat in contrast to this, Höök et al. [2000] found in a study made in preparation to Kalas that social signs appeal to many users but that some users are alienated by them. This might be connected to privacy questions as the more socially rich the environment gets, the more anonymity we lose [Höök et al., 2000].

Not surprisingly, there is growing research interest in how to embed social presence in the interfaces [Gefen and Straub, 2004; Hassaneina and Head, 2007; Cyr et al., 2007]. Svensson, Höök, and Cöster [2005] talk of social texture, which is the sum of all the cues of other users.

Many papers discuss increasing social presence by adding interactive elements for the customer to communicate with the shop representatives directly, for instance by email, or with other customers, for instance with chat [Hassaneina and Head, 2007] as in Kalas. However, in the context of Amazon and similar complex information environments that are not social by purpose, we are more interested in the findings that indicate that perceptions of social presence can be produced with socially-rich text, forums, and human video [Kumar and Benbasat, 2006; Hassaneina and Head, 2007]. Amazon and other similar information environments have customer reviews, discussion forums, and Amazon occasionally has video-clips of the authors introducing their books 14.

Interestingly, Kumar and Benbasat [2006] showed that recommender systems and customer reviews increase the perception of social presence in addition to increasing the perception of usefulness. They suggest that a good recommendation system generates an impression of “being present and engaged in one-on-one dialogue with the customer.” This creates “a sense of personal connection” with the website, resulting in a perception of higher social presence. Consequently, effective use of recommenders can also increase customer loyalty. [Kumar and Benbasat, 2006]

The question of how removed from the actual users these signs of life can be still to produce a feeling of others being socially present is one that deserves further research, as the answer determines how to design interfaces that give the right social texture to generate an appropriate perception of social presence.

Amazon has a rich social texture. The titles of features, such as “Customers Who…” refer directly to other users, and the other users are present in star ratings and customer reviews. In addition, we have access to reviewer profiles, which have potential to be socially rich with photographs and text [Hassaneina and Head, 2007], although typically they seem to contain few social aspects, and there are discussion forums, “Listmania!” lists, and “So You’d Like to…” guides by users. Tags and blogs are among the newer additions to this wealth of social cues. What is missing, however, is direct communication between users, such as chat, although Amazon does offer means of keeping track of “Amazon Friends and Interesting People.” However, none of our participants had used these or were even aware of them, so their perception of social presence was based on the social texture of the interface alone.

The central question concerning social presence for us was how this social texture affected the user’s perception about the inherent sociability of the environment and if this perception affected their behavior. Was the environment perceived as social with other people present in it, or was the environment perceived as inherently solitary, and the social texture simply as features to help in the task and not as social signs of other humans?

If the environment is perceived as social and other users as present, even if not necessarily concurrently, we can consider complex information environments as inherently social and expect to see some evidence of the environment affecting the user behavior. There is evidence that awareness of others affects user behavior. For instance, Cosley et al. [2005] found that oversight (meaning moderation) resulted in better-quality contributions and less anti-social behavior in MovieLens even if the participants were not told of the oversight explicitly. However, the impact was greater if users were told of the oversight. Our results shed further light on the question of the perception of social presence and its impact on behavior.
3 Method and participants

3.1 Method

We used applied ethnography, a combination of observation with verbal protocol and interviewing at the participants’ homes with them using their own computers, as our method to get an authentic view of the real use.

While observation shows what users do and how they do it, it does not reveal their motivations and other reasons behind the actions. We might be able to induce some whys from observation but we need to ascertain them somehow. Both verbal protocol and interviews can provide this insight. Interviewing does not disturb the flow of actions taking place but it takes place after the fact. Verbal protocol, in spite of its limitations and potentially behavior-altering influence on the situation, is still our only way to get inside the participant’s head during the action without the clouding of reflection that interviewing introduces. Thus, verbal protocol provided us with online insight while interviewing provided reflective insight into the actions of the participant.

Combining interviewing with observation was also necessary to avoid the say-do problem. Say-do problem refers to the human tendency to describe what they do differently from what they actually do [Jordan and Dalal, 2006; Kobsa, 2007], and thus what the participants say has to be contrasted with observation.

During the observation-interview sessions, one per participant with each lasting 2-3 hours, the participants were first instructed on the verbal protocol. Then they were given four tasks and asked questions before, during, and after each task. Care was taken not to direct participants’ attention with the questions. After the tasks, a semi-structured interview was conducted. Typically, the task section and the interviewing section lasted about as long. Finally, the participants filled in an online demographic questionnaire.

The sessions were videotaped with the video camera pointed at the computer screen to provide the context for thinking aloud (verbal protocol). The video camera was also used to record the interviews. The participants were explained the use of the video tape and signed a permission for videotaping before starting to record. The sessions were conducted between March 15 and March 27, 2007.
The tasks were given to the participants on a web site made for the study with each task on its own page. However, only the first two tasks are within the scope of this thesis.

Task 1: “Buy a book (or books) from Amazon. Do not buy a book that you have already decided to buy. Instead, you should find a book that you have not decided to buy beforehand.” The participants were further instructed to continue the process until the book was in the shopping cart and to complete the buying after the session to avoid videotaping their credit card information. On average, the participants used 23 minutes on Task 1.

Each participant was given 15 € towards purchasing the book(s) in Task 1 to make sure that they selected a book they really wanted. This is significant because research suggests that people use “affect or other simple heuristics to guide their decisions” when the task does not involve them, the task is trivial, or they are not motivated, while in high-involvement situations, when people have something to lose or are simply deeply engaged, people use “cognitive analytical processing” [Sillence and Briggs, 2004].

In many studies, people are asked to do a task but nothing is done to involve them deeply with the task. Consequently, the results might not reflect the actual use [Sillence and Briggs, 2004]. What is also important to note is that in our study, the users were not paid per se or given a chance to win something by participating, which might motivate them to take part in the study but not engage them any deeper in the tasks. Instead, what they received depended on how they did the task, involving them deeper in the task itself. Consequently, our participants were “genuine” [Sillence and Briggs, 2004] in that they were actual users of Amazon and the use was motivated by a genuine reason.

The participants were instructed to use the Amazon site that they most typically used. Three participants used .co.uk, two used .com, and one used both .com and .co.uk. The choice of site was given to preserve normal conditions although there are differences between the two interfaces.

Task 1 was our main task. Task 2 was designed to support it by focusing the participants on selecting items from a list and deciding which items to buy, as some participants could—and did—complete Task 1 with only minor time spent on these aspects that were of great interest to our study.
Task 2: “You have bought a good digital camera and now you would like to buy a photography guide from Amazon. Which one of the books on the list would you buy?” The task page provided a link to the list page that was constructed to look like a list page from Amazon.co.uk (Figure 5). The page included books with high star rating, low star rating, no star rating, and one book with “Search Inside” function available. A mock-up page was used to make sure that all these different conditions were present for all the participants. The links on the page led to actual item pages in the .co.uk site. On average, 11 minutes were used on Task 2.

Figure 5. The list page used in Task 2.

Task 3 asked the participants to find a book that they had read and read some “Customer Reviews” in Amazon to see if the reviews gave the same impression of the book as they had, and Task 4 asked the participants to decide between iPod and Creative Zen Vision, both having 30 GB capacity, with the help of ePinions.com reviews.
As mentioned, these two tasks are not analyzed for this thesis. Task 3 only provided redundant data. Its purpose was to capture the participant reflections on customer reviews in case we did not get them in Task 1 and Task 2. However, as the participants used customer reviews extensively in the first two tasks and Task 3 was a step further away from organic use, its data was not analyzed.

The purpose of Task 4 was to make it possible for us to contrast the customer review use for different types of items. However, the strong views that some participants held for and some against anything by Macintosh resulted in the decisions being made more based on brand than anything said in the reviews.

<table>
<thead>
<tr>
<th>Quantitative</th>
<th>Qualitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single truth</td>
<td>Multiple realities</td>
</tr>
</tbody>
</table>

**Methodology**

<table>
<thead>
<tr>
<th>Deductive (testing a theory)</th>
<th>Inductive (formulating a theory)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis-driven</td>
<td>Exploratory</td>
</tr>
<tr>
<td>Reliable</td>
<td>Dependable</td>
</tr>
<tr>
<td>Reproducible</td>
<td>Auditable</td>
</tr>
</tbody>
</table>

**Data types**

| Numerical data                | Usually takes the form of words or images |

**Findings**

| Statistically significant     | Valuable                         |
| Generalizable                 | Idiographic/transferable         |
| External validity             | Internal validity                |

**Criticism against**

| Data outside of context not real | Data is too subjective |
| Experiments by their nature change data | Hard to apply to HCI problems |

*Table 1. Key differences between qualitative and quantitative approaches [Tremaine, 2007].*

All sessions were transcribed and then contrasted for analyzing. No analysis software was used. Because the study method produced qualitative data, the goal of the analysis was to describe the observed behavior, and to find patterns and formulate theories based on them. The study did not start with any preconceived hypotheses and thus did not attempt to prove any hypotheses right or wrong. Table 1 summarizes the differences between qualitative and quantitative research approaches as presented by Tremaine [2007].
While the quantitative approach has been more of a norm in computer science, the qualitative approach enables us to answer research questions that we otherwise could not answer [Tremaine, 2007]. Even though some researchers still take extreme stances on the question, the emerging consensus is that the two approaches are complementary rather than conflicting [Tremaine, 2007]. The research question and other factors determine which approach is better in specific circumstances. In this case, the qualitative approach was the only possible one as we needed to observe the participants doing what they normally do in their normal surroundings to see what strategies had emerged in complex information environments.

3.2 Participants

The participants consisted of six Finnish males, aged between 33 and 44 (on average, 38.5 years). All participants were in working life, had at least polytechnic-level education, and were experienced Internet users, using computers on average 6.3 hours a day and Internet on average 3.9 hours a day (the numbers include both work and leisure use).

A book purchase from Amazon was required for recruitment to make sure that all were actual users of Amazon. On average, the participants had purchased 10 books, the smallest number purchased being 2 and the largest 30, from Amazon prior to the study. In addition, three participants had also bought other items from Amazon. Table 2 summarizes the items that the participants had bought from Amazon prior to the study as reported by the participants. In the table as well as in the whole thesis, P1 refers to Participant 1, P2 to Participant 2, etc. The order of the observation-interview sessions and the participant number is not necessarily the same.

<table>
<thead>
<tr>
<th>Items bought from Amazon altogether</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Book</td>
<td>5</td>
<td>4</td>
<td>10</td>
<td>2</td>
<td>30</td>
<td>10</td>
<td>10.2</td>
</tr>
<tr>
<td>CD</td>
<td>4</td>
<td></td>
<td>2</td>
<td>5</td>
<td></td>
<td></td>
<td>1.8</td>
</tr>
<tr>
<td>DVD/Video</td>
<td>1</td>
<td></td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
<td>1.5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>10</td>
<td>4</td>
<td>10</td>
<td>2</td>
<td>35</td>
<td>20</td>
<td>13.5</td>
</tr>
</tbody>
</table>

Table 2. Items bought from Amazon per participant prior to the study.

Thus, our participants were actual users of Amazon. In contrast, many studies use students acting as consumers [Grabner-Kräutera and Kaluscha, 2003; Sillence and Briggs, 2004], which raises questions of external validity [Gefen and Straub, 2004]. However, while our participants were all “genuine” [Sillence and Briggs, 2004], they
were all male, and men and women are known to have at least some differences as e-commerce customers [Cyr et al., 2007], and the number of our participants was low. Consequently, our results can be seen as having external validity only when most or all participants concurred and only as far as males are concerned. Furthermore, cultural issues prevent us from generalizing the results too widely, and our conclusions need to be retested in other cultures.

The common reason for the participants to use Amazon was the large selection of books. Half the participants also mentioned lower prices as a reason while two participants actually maintained that the books in Amazon were not any cheaper than in other online or brick-and-mortar bookstores.

The participants estimated that they visited Amazon on average 34.7 times a year, ranging from 4 times to 104 times a year, and had used Amazon on average for four and half years, ranging from 2 to 7 years. Table 3 shows the visiting frequency per participants and how many years the participants had used Amazon.

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visits per year</td>
<td>52</td>
<td>104</td>
<td>12</td>
<td>4</td>
<td>24</td>
<td>12</td>
<td>34.7</td>
</tr>
<tr>
<td>Years used Amazon</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>4.7</td>
</tr>
</tbody>
</table>

Table 3. Frequency of visiting Amazon and years of using Amazon per participant.

In addition to Amazon, all participants had used other online stores, too. The participants estimated that on average they bought 14.3 items online annually. Table 4 shows the breakdown of the items the participants had bought online for personal use prior to the study.

<table>
<thead>
<tr>
<th>Items/participant</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Books</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>CD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVD/Video</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Games</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Computer parts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Travel services (tickets, hotels etc.)</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (concert tickets, home appliances etc.)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Items bought online per participant.

Five participants had bought items online both at home and at work while one had shopped online only at home. However, the purchases made at work were typically work-related, and thus all participants typically shopped online for personal
purchases at home. Consequently, the observation-interview sessions were conducted at the participants’ homes.

Four participants used the Windows XP operating system while the other two used Macintosh. Two participants used laptop—one XP and one Mac—while others used desktop computers. During the study, the participants were instructed to use the browser they typically use. Four users used the Firefox browser, one used Safari, and one used Internet Explorer.
4 Results and discussion

4.1 Finding items of interest

4.1.1 Overview

The ways people locate items of interest evolve with the information environments where they look for the items. The environment naturally dictates the item-finding process to an extent by determining what tools and features are available. Today’s complex environments provide users with multiple approaches to choose from. We first look at the way each participant approached item-finding in Task 1 before discussing how searches and recommendations are increasingly complementing each other in the item-finding process, providing together both finding and discovering.

The process of finding items of interest might start with a recommendation email from Amazon (P3), with an Internet search on some topic that eventually leads users to Amazon (P1, P2, P4), or with users hearing of a good book from friends (P2, P4, P5, P6). Table 5 summarizes the typical reasons that had brought the participants to Amazon prior to the study.

How the pre-arrival story goes affects how users arrive at Amazon. P3 might click a link in a recommendation email and arrive directly to the item page while P1, having borrowed an interesting book from library, might come to Amazon’s home page and make a keyword search with the book’s title. How much is known about the book affects how users search for it. If enough is known, query search is a natural approach to find the particular item [Hangartner, 2007]. On the other hand, if not enough is known or the item found turns out not to be interesting for the user, recommenders kick in and help in the discovery process [Hangartner, 2007]. For instance, users seed recommendations by making keyword searches to see what is recommended to them on the item page of a book on the topic area that they are interested. Furthermore, users might opt to go directly to recommenders for discovery, as P2 and P3 did.

During the item-finding process, we saw that the participants used recommenders in two ways, strategically and opportunistically. Strategic use refers to intentional use while opportunistic use refers to using recommendations organically as they are offered without prior intention.
During the item-finding process, we saw that the participants used recommenders in two ways, strategically and opportunistically. Strategic use refers to intentional use while opportunistic use refers to using recommendations organically as they are offered without prior intention.

<table>
<thead>
<tr>
<th>Recommendations</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personalized recommendation emails from Amazon</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Friends</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

| Internet | | | | |
|----------|---|---|---|
| Searching for books or information on a topic on the Internet | X | X | X | 3 |
| Reviews on the Internet | | X | X | 2 |
| Discussion forums | X | X | X | 3 |

| Physical books | | | | |
|----------------|---|---|---|
| Seen in library | X | | | 1 |
| Seen in bookstore | X | X | | 2 |

| Other | | | | |
|-------|---|---|---|
| Magazines (reviews etc.) | | X | X | 2 |

Table 5. How the participants heard about books and were motivated to come to Amazon.

Furthermore, users appear to divide into two groups as far as the overall item-finding process is concerned. Some look for the best item while others are happy with an item that fulfills their need. We call these two approaches Best item strategy and Good enough item strategy.

The participants did not limit their searches for information on the books to Amazon. Instead, they used other online sources to help them in decision-making process especially if Amazon did not offer them enough information. Furthermore, the way the participants approached Amazon’s list pages and keyword searching bore some hallmarks of their Google use, as discussed in Section 4.1.6.

We first take a look at the item-finding process in Task 1 and then discuss the ways the participants used to find items of interest. Then we look in more detail opportunistic and strategic uses of recommenders before discussing Best item strategy.
and Good enough item strategy. Finally, we look at how non-Amazon sources affect the equation.

4.1.2 Item-finding process in Task 1

“I’ll ask them [Amazon] what they have for me.” – P2

Task 1 asked the participants to find a book to buy that they had not decided to buy beforehand. P2 and P3 started with personalized recommendations. P3 was automatically signed in and he first glanced at the recommendations on the home page. However, he was not really interested in the general recommendations but wanted to see personalized recommendations for books. He spent over two minutes looking for them before noticing the link: “All right, here is what I was kind of looking for, here are the types of topics I [wanted].”

P3 checked out the recommended books by moving the mouse cursor over the cover images to get more information on them. He was especially interested in the star ratings of the books. Having found one interesting book with almost full five stars—The Complete Idiot’s Guide to Music Composition—he went to its item page. He first read the two topmost reviews in “Customer Reviews,” both of which were only three lines long\(^{15}\), then scrolled through other reviews on the page in a glancing manner, and finally read one more review that had 16 lines of text. After using the “Search Inside” feature, he decided to buy the book.

The recommendation was serendipitous. P3: “This book is for making music and the topic is the theory of music. I haven’t really gotten into music theory anywhere, so I am really, really interested in the topic. I think that the price level and the topic match each other really nicely for me.”

P2 also started with personalized recommendations, remarking: “I’ll ask them what they have for me.” However, as he had not visited Amazon previously with his brand new laptop, he first had to figure out how to sign in: “This is not mine, that’s for sure. There’s the sign in [link]! Why is it so darn small?”

After signing in, he felt more at home—“All right, now it starts to look familiar”—and started to look at the recommendations. Failing to find anything

---

\(^{15}\) The number of lines depends on the size of the browser window, and so the numbers here only give a general idea of the length of the review.
interesting that he did not already own and without attempting to improve recommendations by rating the items he had, P2 made two searches with authors’ names. First he searched with “Thomson[,] Geoff” (a self-defense specialist) as keywords and went through three list pages of his books and DVDs without finding anything new and without going to any item pages. Then he searched with “Linklater[, Kristen]” as keyword and eventually located a book he had heard of and was interested in: *Freeing the Natural Voice*. Even though he had been recommended the book by friends, he nevertheless wanted to read “Customer Reviews.” However, the book did not have any. P2: “There’s no review. … it doesn’t matter because I’ve heard that it is a good one. I’ll take this.”

Saying that he liked to “hesitate and ponder an awfully long time” before deciding on an item, P2 described his typical item-finding strategy in Amazon: “First I usually see the recommendations and then books I know or authors that I’m familiar with. I search by their names and see if anything interesting pops up.” Thus, recommendations play extensive role in his item-finding strategy and provide him with discovery.

P1, P5, and P4 all started with keyword searching. P1 wanted a book on digital video making. The first keyword combination, “video production,” did not at first seem to return anything interesting: “The search seems to give me mainly books related to some programs.” Eventually he found two potentially interesting ones on the second list page and opened the two that he found most interesting to their own tabs to store them there while he continued searching. Glancing through the item page of one of the books in the tabs, he dropped it saying: “So there’s only one review and that’s by whom? By the author [laughs] himself apparently? I can’t tell where this information is from. Well [sighs], it doesn’t mean that just because it hasn’t gotten any reviews from people that it is bad.”

He made another keyword search with “video filming” as keywords, checked that the category was still books, found another interesting book, and placed it in its own tab. He looked for the “Search Inside” feature from the item pages slightly exasperatedly: “I am trying to find that feature that some books in Amazon have—or is it that all the books have it—or is it only in dot-com—where you see what the book really looks like [opens a bigger picture of a book’s cover while trying to find the “Search Inside”] and not only the cover.” In fact, while neither book had “Search
Inside,” one of the books he had quickly checked had had it. He also checked one book’s “Customers Who…” recommendations but did not notice anything interesting.

After completing searching phase, P1 compared the two books in the tabs by moving between the tabs, and then picked the first book that he had placed in a tab. After placing the book in the shopping cart, he got interested in the targeted impulse items (“Customer’s who bought … also bought”) that were shown next to the shopping cart (Figure 6): “Well, let’s take a look at what others bought in addition to this one. Most likely I would buy another book, you know, just to get the delivery expenses down.”

![Figure 6. Impulse recommendations next to the shopping cart.](image)

P1, who felt that the title was important, got interested in a book called *The Guerrilla Film Makers Handbook*, went to see its item page, and read the synopsis and the one customer review available that had given the book five stars: “This looks good. But I kind of would like to see if I can find—I mean the book being more expensive and all—if there’s a review of it somewhere else than in Amazon as well.” He also

---

16 In fact, “shopping basket” as he was in Amazon.co.uk. However, for the sake of simplicity, we use Amazon.com’s names for features as the names are close enough not to cause misunderstandings.
noted that the book was offered in “Perfect Partner” together with the book that he already had in the shopping cart.

P1 made a Google search with the book’s title and “review” as keywords, located one review, read it, and decided to get the second book as well. Altogether, he spent £40 even though he had mentioned that he would like to get a book on the topic for about £10. Thus, he altogether used more of his own money than was given to him (€15).

P4 started by making a search with “phone tapping government.” On the list page, he set “Browse by category” to “Books” before starting to look at the items. The books ranged from self-healing book to books on George Bush’s America. He found a book whose name interested him and went to the item page. There he located two reviews and read them carefully: “Well, they say that this is a good book, better than … fiction because this is true. Based on that, I presume that it is nonfiction [scrolls back to the top of the page] and that’s what I’m… Now I want, I need more information. Based on this, I don’t, for instance, know who this guy is.”

To find more information, he went to see the other books written by the author. Satisfied with them, he decided to find more information: “This would seem potential, so normally I would here, when I don’t find enough information here, I could make a quick search with Google.” He found the author’s homepage by searching with the author’s name but could not find if the book covered wiretapping or not. P4: “You know, the author has been cleared as far as I’m concerned. He seems like a good, reliable character that has had a lot of exposure on the topic area. Now I’m going to make a search in Google where I connect his name to wiretapping, tapping phones, which is what I originally was interested in. Does this guy have any knowledge on phone tapping stuff?” However, the search produced no results.

P4 returned to the list page and refined his search by replacing “government” with “EU” but got only one book that was off-topic. Giving up on the first topic, he started to search for a book on tankhas. The problem that he immediately experienced was that there is no standard way to spell tankha17. He tried two spelling options and looked at the few books on the list page, which contained many books only marginally connected to the search keyword. When taking a quick look at one of them

---

17 Wikipedia.org, for instance, gives four different spellings for the word: Thangka, Tangka, Thanka, and Tanka (http://en.wikipedia.org/wiki/Thangka)
he the discovered “Search Inside” feature—“But this ‘Search Inside’ is good. I haven’t noticed it before at all”—which he continued to use extensively after that.

On the second page he found Departed Angels: The Lost Paintings by Jack Kerouac and Ed Adler and went to the item page. “Searching Inside” the book he clicked “Surprise me” and was given the sign-in page: “Well, this is not the kind of surprise I wanted [Laughs]! I clicked ‘Surprise me’ and now it’s asking me to register or something. I won’t bother. That was no happy surprise.”

Although he dropped the book, the theme stayed with him and he made a keyword search by the name of a book he had seen in the list page earlier on, When I Was Cool: My Life at the Jack Kerouac School by Sam Kashner. He used “Search Inside” on the book and also looked at the other books by Kashner: “He’s written a lot. Here’s Sinatraland. He’s tracing the soul of America in many ways. He seems to be a specialist on that, so I would say that it must be interesting. The kind of author who knows his stuff, very productive, and has made a book about these people. I could buy it on impulse.”

He was already about to place the book in the shopping cart when he suddenly said: “Hey, wait a moment. … I’ll just peak at the customer reviews quickly. You know, to see what they’ve said about it. It’s not that it’s so expensive but it does take time to read it. That’s more expensive.” He read carefully the first five reviews and then glanced at the rest that were on the item page. The customer reviews were praising the book but something caught P4’s eyes: “Here it says that it is a bit HC 18, that is focuses very heavily on the beatnik-culture. But I might be interested in that. Probably. It is perhaps, yes, maybe… I’ll check that it is not some huge behemoth of a book because with these coordinates I’m not up to reading one of them. A snack-sized one would be good.”

Trying to find how many pages the book had he noticed the “Customers Who…” feature: “Oh, hey, hey, hey! Now I’ll still…Yeah, now I found a really good one! I mean true enough. I was kinda left feeling a bit vexed about Kerouac. I mean Kerouac is for me kinda like, I mean I notice that I’m chasing after him here. This ‘Customers Who Bought This Item Also Bought’ is throwing at me this Windblown World: The Journals of Jack Kerouac 1947-1954. … Well, this showed itself to be useful, you know, something like this can pop up out of nowhere at you. If these two

18. “HC” means “hard core.”
books are side by side, there’s no doubt which one I’d take. I don’t really even need to read any more on this one. I know Kerouac’s life by and large. I know his style of writing.” He placed the book in the shopping cart in the “Used and new” section as the book was not sold by Amazon directly. Thus, also this recommendation was serendipitous, and it allowed discovery.

P5 had the clearest strategy of all the participants. He made keyword searches, went through the list pages briskly, and opened any potentially interesting book in its own tab. P5 made several keyword searches, quickly refining the search if the results did not appear interesting: “I see that this keyword is returning all kinds of stuff, also stuff that I’m not interested in. Yeah, this keyword didn’t find me anything interesting so I’ll try the next one.” He also had a problem with keywords as he was interested in Chinese health exercises known as \( \text{\textasciitilde chi kung} \), which can be Romanized either as “chi kung” or “qigong” or even “ch’i kung,” and in Chinese philosophy based on \( \text{\textasciitilde dao} \) which can be Romanized either as “tao” or “dao.”

After a few searches and over a dozen books in the tabs, he turned his attention to the items on the tabs, closing the ones that did not interest him and leaving the interesting ones. P5: “Now that I’ve found these I’ll go through the same elimination process. I take a closer look at them. You know, to see if any appeared interesting on the same basis. If there’s nothing on it, I just drop it. I do want to know something about the books that I buy. But if not, I don’t want to buy a pig in a poke. [Dropping some books] There’s so little info on these ones that I’m just dropping them.”

He used “Reviews,” “Customer Reviews,” and “Search Inside” to decide which ones to close and which ones not to close. If not enough information was available, he took a look at other books by the author and also made Google searches to find more information if the book somehow interested him. P5: “I could look at this a bit closer but there’s no information here. Now I’m going to search if I could find some more exact information about it.” P5 was using Amazon.co.uk, and so some of his searches took him to Amazon.com where there were more reviews. However, he also used several other sites, including smaller specialist bookstores that had more information about the books.

When the books were quickly whetted down to a handful, he started to compare them. He checked out the book offered in “Perfect Partner” with one of the
books he had in the tabs but he already had it. Worried that he might have one of the books as a part of some large collection, he tried to find more information about that aspect. Failing to find anything conclusive, he nevertheless felt satisfied with his effort: “Well, I can order this one, no problem. It seems good from any viewpoint, the kind that I can invest some money on.”

P6 commented immediately on the amount of items that Amazon boasted: “…there’re too many things here, confusing abundance, there’re awfully many things on the screen as it is, and I’m left to figure out where to start. I’m trying to decide which way to take…” Eventually, P6 started to look for a book by categories (“Browse Genres”), something that he would not normally do, as it turned out. He persisted with the approach for quite a while without finding anything interesting before giving up and going for a keyword search.

After failing to find a book with some keywords, “stretching,” “PNF” (a specific style of stretching) and again “stretching,” P6 started to search with the name of an author he knew. The trouble was that he was not able to remember the exact spelling, and so he made a few abortive attempts before trying to locate the author by searching with words in the title of one of his books. He found the author’s name from the list page and went to the item page. There he found “Customers Who…” recommendations: “Now that I know the author’s name, let’s take one book [book page] and then we’ll look for… Just a moment… Where is it? This is the list I was looking for.” None of the books on the list interested him, however, and so he clicked the “Explore similar items” link for more recommendations. He took a look at the page and returned to the list page via the item page.

Now P6 made a search with “body flow” as keywords. He found one interesting book but there was no “Review”, no “Search Inside”, or “Customer Reviews.” P6: “No good.” P6 then tried to find a book by an author with a Russian name. Again, the spelling was problematic. P6 found nothing on Amazon.co.uk, but, undaunted, he switched to Amazon.com and now managed to locate the author and the book that he had in mind.

After checking that the book cost $19.95 and had 125 pages, P6 went down to the “Customer Reviews.” He read the first one and eyed the others: “Five stars, five stars, four stars, five stars…” Then he read two more reviews in a glancing manner: “Everybody’s praising it.” P6 decided to buy the book from Amazon.com and commented on the finding process: “I can’t really say what the mental connection that
“brought me to this book was.” Table 6 summarizes the books bought by each participant in Task 1.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Book title</th>
<th>Method of finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Making Short Films: The Complete Guide from Script to Screen</td>
<td>Keyword search</td>
</tr>
<tr>
<td>P1</td>
<td>The Guerrilla Film Makers Handbook</td>
<td>Impulse recommendation</td>
</tr>
<tr>
<td>P2</td>
<td>Freeing the Natural Voice</td>
<td>Keyword search</td>
</tr>
<tr>
<td>P3</td>
<td>The Complete Idiot's Guide to Music Composition</td>
<td>“Recommended for You”</td>
</tr>
<tr>
<td>P5</td>
<td>The Book of Balance and Harmony: A Taoist Handbook</td>
<td>Keyword search</td>
</tr>
<tr>
<td>P6</td>
<td>Let Every Breath... Secrets of the Russian Breath Masters</td>
<td>Keyword search</td>
</tr>
</tbody>
</table>

Table 6. Books bought by the participants in Task 1.

4.1.3 Searching vs. recommendations vs. categories

“…I mean it happened to me out of the blue because I don’t know where this Kerouac thing came from. It came some weird route. The system kind of drove me to it. I kept getting closer and closer all the time and when I eventually was about to take the other book that was more at general level, it pushed this Kerouac’s memoirs at me [laughs] and I couldn’t resist it or ignore it. If I were in a bookstore, how the hell would I end up with something by Kerouac? I’d be there looking at some painting books and the link between Kerouac and tankha-paintings would be hard to draw, it just wouldn’t happen, and in that sense I’d be there, probably looking at some impressionistic painting guides [laughs], and think that maybe this is not quite what I wanted.” – P4

In Task 1, the participants bought altogether seven books. Four books were found with keyword searches while three were found through suggestion recommendations. With “suggestion recommendation” we refer to recommenders that offer suggestions of item(s). Figure 7 summarizes the overall process of how the books were found in Task 1.
However, even though four books were found by keyword searching, in practice all participants used recommendations in the item-finding process. Table 7 summarizes the recommender use in Task 1.

<table>
<thead>
<tr>
<th>Task 1</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bought a book offered by algorithmic recommender</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Bought a book found by keyword searching</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>4</td>
</tr>
<tr>
<td>Used keyword search</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>5</td>
</tr>
<tr>
<td>Used categories for searching</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Used “Perfect Partner”</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Used “personalized recommendations” (at the beginning of the process)</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Used “Customers who bought/viewed this item…”</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Used “Explore similar items”</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 7. Recommendation use in Task 1 by the participants.

Interestingly, all three books found by recommendations could be characterized as serendipitous. Participants found items that they would not have found otherwise and that were exactly what they wanted. The recommenders were clearly providing discovery in each case.

It is also fascinating to note that two participants used personalized recommendations as the starting point. Furthermore, several comments by participants underlined that they were actively looking for recommendations, especially in the situation where the book on the item page did not really interest them. The
participants clearly perceived recommenders as natural part of the information environment instead of treating them as novelties.

Only P3 had partially erroneous idea concerning a recommender, namely the “Customers Who Viewed This Item Also Viewed” feature. He thought that the authors of the books on the list had bought the book on the item page, and saw the feature as an endorsement by the authors of the books in addition to it being a list of recommendations.

Consequently, recommenders have become an integral part of complex information environments also in users’ minds, and play a significant role in item-finding strategies. Not to have them would handicap users and make discovery less likely, even though keyword searching can also produce interesting results. Our findings are in line with the studies that suggest that recommender systems are necessary and useful in finding items in the era of information overload.

Meanwhile, the role of keyword searching, once the mainstay of item-finding, has diminished somewhat due to the advent of recommenders. However, it is the natural approach or at least a starting point when the topic is known but not much more. This is especially true when searching for an item on a new topic or not being signed in. Recommendations need “seeding” before they can produce results, and if the past history is not known (new users or users not signed in—four participants in our study were not automatically signed in) or the topic area is such that the user has not searched for or bought items in that area before, recommenders cannot successfully suggest items. Therefore, users need to make searches and view at least one item in the topic area before recommenders can start helping in the process.

One of the major problems with searching is naturally to come up with correct keywords. With incorrect keywords, for instance because of synonymy, many words meaning one thing [Golder and Huberman, 2006], items of interest are not found, and we can never know what we did not find. For instance, P4 used as keywords “phone tapping government.” That search produces 22 results in Amazon.com while a search with “wiretapping government” produces 215 results (March 18, 2008). However, P4 did not come up with “wiretapping,” and so he concluded wrongly: “Perhaps a book with that stuff in the way that I want it has not been written.” Finding the right keywords can be further complicated when the system is not in the native language of the user, as with our participants.
The opposite problem, polysemy, one word meaning more than one thing [Golder and Huberman, 2006], causes us to get irrelevant results but at least does not leave relevant items unfound. Nevertheless, P4 was wryly amused when his search for “tanka guide” returned *ABC Practical Guide to Dog Training* by Steven Appelbaum (the book contained a reference to another book published by Wakan Tanka Press) when what he wanted was information on Tibetan paintings.

Furthermore, names—many participants used author’s name as keyword—and concepts with foreign or complicated, non-common words caused spelling problems. Recommenders can help when at least one spelling that produces an item related to the concept is found, because with words like “tanka” with four alternative spellings, it is unlikely that the user can come up with all of them. Additionally, if the user tries to use the word together with other words—P4 searched with “tanka painting” and “tanka guide”—the variations multiply.

However, participants had strategies to deal with such situations. For instance, when P6 failed to remember the spelling of an author’s name, he instead searched for a book he knew to be by the author, as he knew how to spell the words in the book’s title. Finding a book by the author helped him to access relevant recommendations.

This underlines the finding that recommenders can complement searches and inspire new searches, just like searches seed recommendations. In the light of this study, searching and recommenders do not compete with each other but instead complement each other in many ways.

However, it is Amazon’s ability to make recommendations based on just one item viewed that makes this possible. If recommendations were simply based on previous purchases and did not react to the item at hand, users could not integrate into the item-finding process as they now do in Amazon. For instance, users looking for a book in a new topic area would not get suitable recommendations to their current interest. Consequently, to allow recommenders and searches to complement each other, the recommenders have to be responsive to the current task context. This is what makes discovery possible.

Consequently, our findings are in line with Hangartner [2007] who concludes that searching is not disappearing because of recommenders but can be enhanced with recommenders, and that recommender industry will continue to grow in sophistication and importance, becoming eventually “more pervasive than the search industry and search technology as we know it.” While Hangartner’s insistence that “while search
engines help you find things you know you are looking for, discovery helps you find the rest” somewhat oversimplifies the situation, by and large it does describe the underlying tendencies in the complex information environments.

Searching by category, on the other hand, appears not to be efficient or popular. Only P6 tried it, and even he admitted that doing so was not his typical approach. The only way categories were used effectively was to limit searches to books. In the current version of the “Customers Who…” recommendations, users can adjust recommendations by categories (see Figure 1 on page 7). The size of the category name communicates the size of the category, that is, the number of recommendations to be found in it: The bigger the title font, the more recommendations behind it.

Our results also underline the importance of studying recommenders—and indeed any feature of complex information systems—as parts of the whole in addition to studying them in isolation. Recommenders cannot be separated from the whole, as both their use and usefulness depends on the use context and how well they are integrated to the whole to support user strategies.

### 4.1.4 Opportunistic and strategic use of recommenders

“And I think that this feature is good, this ‘those who have bought this book have also bought that book.’ I have found some books by that. For instance, I think that when I was looking for a book on these mercenaries, it gave me a good list. I found something that had something to do with it, and then I could search through it, and it worked very quickly, because I can do kind of a cross-search, search for books on mercenaries. Then when I read about some of them, some that I might be interested in, and then I take one and then I go to this ‘who bought this also read these,’ and it shows books with similar themes.” – P4

Recommendations were used in two ways, strategically and opportunistically. Strategic use means that the recommender is intentionally a part of the current item-finding strategy. The strategy might be accessing “personalized recommendations,” as P2 and P3 did, or searching for a particular book to see “Customers Who…” recommendations, as P6 did by finding a book by an author to see what other books were recommended on the item page. He had no interest in buying that particular book, but he wanted to see similar books. Intentionality shows in two ways in this
strategy: P6 both intended to go to the recommender to see what was recommended and in his conscious attempt to influence the type of books to be recommended.

Opportunistic use refers to users stumbling upon recommendations and using them there and then. It lacks the intention that characterizes the strategic use. Opportunistic use is possible only if recommender features are displayed in the natural context as part of the interface. On the other hand, recommendations that require us to access them intentionally, such as personalized recommendations (“Recommended for You”) can only be used strategically. However, recommenders that are displayed as a part of the interface, such as “Customers Who…,” can and are used strategically and opportunistically, as we witnessed in our study.

Consequently, the two approaches do not correspond directly to how the recommendations are displayed, and thus cannot be traced directly back to how Schafer, Konstan, and Riedl [2001] categorized recommenders by the way recommendations are delivered. Nevertheless, the way of delivering does affect the available approach to some extent.

The secret to helping users use recommenders opportunistically is to deliver the suggestions when users are pre-disposed to attending to them. For instance, delivering “Perfect Partner” recommendation to P5 when he was vacillating between two books in Task 2 that recommended the two books together at what appeared to be a slight discount (it was not) helped P5 to decide to take both. In the same way, delivering recommendations to P1 when he had put one item in the cart caught him at the moment he was not doing anything else, and so he was pre-disposed to check the suggestions out.

“Customers Who…” recommendations work in a similar manner. If the user is not sure about the book on the item page, he or she is likely to be interested in alternatives to it. Thus, designing recommenders for a complex information environment includes designing their position in the process that they are meant to support.

Opportunistic and strategic uses of recommenders are both ephemeral, and users cannot be categorized by them. While P3 did simply look at the personalized recommendation in Task 1 and found a book, thus using recommenders only strategically, in any longer item-finding process users are likely to move from strategic use to opportunistic use and back again. The user might start a strategic use process only to be sidetracked from it and start using another recommender
opportunistically or the user might exhaust a strategic approach without finding items of interest and move on to opportunistic use. Also, opportunistic use might give rise to strategic use. How smooth the transitions are depends on how well the environment is designed to support discovery and user strategies that emerge from that environment.

4.1.5 Best item strategy vs. Good enough item strategy

“The price was the biggest reason plus in a way I saw what I was looking for. It was like, a-ha, this is something that will help me get started.” – P2

The participants divided into two groups as far as their decision-making strategy was concerned. We call these strategies “Best item strategy” and “Good enough item strategy.” The strategies emerged from the way the participants decided on items, and it was related to their browsing approach.

All but P5 used the same browser window when moving back and forth between the list pages and the item pages. Thus, when they saw a potential item on the list page, they opened the item page in the same window, and afterwards returned to the list page with back button navigation.

However, when P1 found an item that he could consider for buying, unlike others he opened it to a separate tab and then returned to the list page to see if there were other interesting items. P5, in contrast to others, opened all interesting items directly to tabs as he went through the list pages. The end result was the same for P1 and P5, however. The final phase of finding an item for both of them consisted of going through the items in the tabs. P5 dropped the items that were not really interesting and then started to compare the remaining items to decide which one(s) he would buy. P1 moved faster to the comparison phase because he had pre-selected the items in the tabs more carefully by visiting the item page before opening the item in a tab. However, eventually both ended up comparing items to determine which one was the best for them.

P2, P3, P4, and P6, on the other hand, moved to the item page and then decided if the item was the right one for them there and then. If it was, they placed it in the shopping cart, and if it was not, they moved back to the list page and found the next potential item that they then evaluated on its own merits without directly comparing it to other products. There was no evidence of the shopping cart being used for storing items without intent to buy it. However, since Task 1 ended with the item
being placed in the shopping cart, we cannot be sure of different ways of the participants would have used the shopping cart.

Thus, P1 and P5 were concerned with finding the best item (from the ones that they had found) while P2, P3, P4, and P6 were concerned with finding a good enough item to satisfy their current need. Interestingly, the Best item strategy led to P1 buying two books in Task 1 and P5 to indicate that he would have bought two items in Task 2. The participants using the Good enough item strategy only bought one item each in both tasks.

Although the participants rather consistently used one strategy on both Task 1 and Task 2 with only P3 showing some tendency towards comparing items on Task 2, we cannot draw the conclusion that people would use only one strategy consistently with our small sample. One possibility is that the strategy choice depends on the importance placed on the item. With important items, we might be more inclined to make a more careful choice while with less important things we only need any item that answers to the need.

On the other hand, the fact that the strategy was connected to the browser-using strategy does argue for consistency. Furthermore, when the participants left Amazon to make Google searches, they appeared to behave in the same way. P5 opened potential sites to tabs to attend to them later while others moved back and forth between Google list page and offered sites.

Interestingly, the phenomenon was discussed by Simon [1955] already back in 1955 when he rejected the idea of all-knowing “economic man” and replaced it with an “organism of limited knowledge and ability.” Simon suggested that the actual decision-making process is sequential to reduce the cognitive burden. First, the number of possible items is reduced to a manageable level, and then the remaining items are evaluated at more detailed level. Simon, a chess fan, likened this to quickly dropping the moves that are clearly non-productive and focusing on the potential alternatives. However, if a move that brings the desired result is found, such as a sure way to check-mate the opponent with five moves, the alternatives are no longer considered. It does not really matter if there was a way to check-mate the opponent with four moves.

Thus, Simon’s “behavioral model of rational choice” has elements of both Best item and Good enough item strategies. Keyword searching or recommendations take us to a reduced set of items. The dividing question is if we continue to reduce the
number of items with detailed evaluation until the best item remains or if we satisfice to a good enough item earlier in the process. Furthermore, it is a question of how detailed information we use for reducing the number of items. The Good enough strategy involves selecting which items to see in detail on the list page while the Best item strategy leads to seeing more items at item page level, thus with more details on which to base the decision. Moreover, the Best item strategy is characterized by direct comparison of items.

Although all real-world decisions involve satisficing to some degree as all options cannot be known or evaluated in detail, the question here is one of degree: Do we try to make sure that we get the best item, given the constraints, or are we happy to find something that works for us in the particular context?

Thus, we witnessed that there are two strategies, Best item and Good enough item, and that the choice of strategy showed evidence of being constant among the participants. How clearly people divide into two categories and how consistent the behavior is need to be determined with further research.

4.1.6 Non-Amazon sources and their influence: Google and others

“Internet was a way to get information on the topic, in this case private armies, you know, these private companies that offer military services to the US government. And there’s a lot of information on this kind of topic on the Internet and slowly you start to find books, and that’s how you end up there [in Amazon]. This way Amazon is a natural option then you decide to go take a closer look at the book.” – P4

The rest of the Internet influenced the item-finding process at least in three ways. First, the impetus to come to Amazon often came from the Internet, a cornucopia of information on a plethora of topics. Maybe P4 was looking for information online and ended up buying a book in Amazon, perhaps P5 read an online book review and came to Amazon, or P3 might have noticed an interesting book in a discussion forum and come to Amazon for a close look. In fact, only P2 did not mention any Internet sources as a motivation to come to Amazon (see Table 5 on page 39).

Second, the participants used the Internet to find more information about potentially interesting items. This invariably meant “googling,” as the participants frequently equaled Internet searching with using Google: All used Google as the preferred search engine. While only three participants made Google searches on items
during the sessions, all mentioned making them normally. Table 8 summarizes the Google searching done in Task 1 and 2, showing both the frequency and the keyword-forming approaches.

<table>
<thead>
<tr>
<th>Use of Google</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Made Google search in Task 1 (times)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Used author name as keyword</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Used book title as keyword</td>
<td></td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Used both book title and author name as keyword</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Used also other keywords (with book title or author name)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Made Google search in Task 2 (times)</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Used also other keywords (with book title)</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Searches altogether</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 8. Use of Google during Tasks 1 and 2.

Normally, having users leave the site is considered bad for business. The Internet is full of competitors, and frequently these searches led to competitor’s site. Our task setting limited the users to buying books from Amazon, but P5 did bookmark a competitor’s site for future reference during the study sessions: “There’s plenty [of things] on this Chinese culture site. This book looks interesting. In fact, the whole site looks interesting. I’ve never been here before but there’s plenty of interesting stuff. I’ll bookmark this. I may find something interesting from here later on.”

The reason for the participants to leave Amazon was lack of information. P5: “I could look at this a bit closer but there’s no information here. Now I’m going to search if I could find some more exact information about it.” Likewise P4, who wished for more “authoritative” reviews by editors or specialists instead of customer reviews, did not leave Amazon to look for them if Amazon provided enough “Customer Reviews” and “Reviews,” and perhaps “Search Inside” because these gave him enough information to form a decision. P4: “…when I don’t find enough information here, I … make a quick search with Google.” Thus, the key to keeping customers on the site is to provide them with enough information on which to base their decisions.

The last part of Google’s influence was very interesting. It appeared that Google had colored at least to some extent the participants’ expectations about keyword searching and list pages. The keyword use for searching in Amazon looked
like keyword use we would also expect in Google: “tankha guide,” “tankha painting,” “phone tapping government,” “Daoism,” “body flow” etc. However, keyword searching is keyword searching, and we would not have noticed the mental connection if some participants had not referred to Google while searching in Amazon. For instance, P4: “…I tried to search inside the book directly just like with Google…” and P5: “Well, I find them quite often by searching with keywords in Amazon, searching with keywords in Google, these are the main ones.”

Google’s dominance as a search engine seems to be molding the way we see other information environments. The participants who made Google searches at least appeared to approach list pages in Amazon as they would in Google. P1, explaining why he did not typically go beyond the first list page even if there were more results on other pages: “Maybe it is because when you search with Google, what you want is on the first page.” P4, explaining why he had not missed tools for organizing the search results: “I mean in this situation when you have at most fifty books, I didn’t somehow consider it very important. Maybe it comes, well, fifty is beginning to be a borderline case, but it is kinda like in Google. If you get, say, I mean the pain threshold is somewhere around fifty, more or less. If you get more hits than that, then you start to think about it.”

The evidence that we found about Google affecting the way users perceive other information environments is circumstantial. As Google certainly appears dominant enough to set framework expectations for users, we feel that the anecdotal evidence is strong enough and the implications serious enough to require further research.

4.2 Star ratings and Helpful ratings

4.2.1 Overview

“I don’t think that one Joe Blow’s say-so matters. You could say that the strength of Joe Blow reviews lies in the mass. And that’s it. If 200 Joe Blows give four stars on average, then that’s kind of reliable.” – P4

Above we have discussed the item-finding process at a strategic level. Here we move to the nitty-gritty and discuss the impact of recommendations for a particular item. Amazon’s recommendations for a particular item are not personalized in that they do
not try to predict how the user in question would rate the item. Rather, they summarize how other users have rated the item and leave users to draw their own conclusions. On the list page, we have average of star ratings that the users who wrote customer reviews gave the items, and on the item page we have the helpful ratings for individual customer reviews.

We first look at how these affected the participant behavior before turning our attention to the use and impact of “Customer Reviews” in the next section. Customer reviews are recommendations of one user to all the other users, and thus are not personalized to individual users, either.

4.2.2 Star ratings and other factors on the list page

“I must’ve done it automatically because of looking at those stars there. This one’s gotten five stars while the other one’s gotten three and a half. But when you go to the page, you see that there are only two reviews. So in that sense it’s humbug but that’s the reason why I went there.” – P4

Star rating was only one of the many pieces of information that were available on the list page for selecting items for scrutiny. Table 9 summarizes to which pieces of information the participants attended in Task 1. Interestingly, only half the participants admitted being influenced by star ratings when selecting books for scrutiny in the list page. (While only P3 attended to the star ratings on the list page in Task 1, P2 and P4 attended to them in Task 2.) Observations supported this, nevertheless. For instance, P1 went to an item page and his first reaction was: “Oh, average customer review one star?” P6 was likewise occasionally surprised by the average star rating when arriving on the item pages while P5 opened books with complete indifference to their star rating.

To the other half, however, star rating was important factor in selecting which items to view at item page level. For P3, the star ratings were the starting point: “…of course I first start to look at those stars, what others have given in these reviews.” P2 also used the star ratings to differentiate the items on the list: “Then I noticed that most of them seemed to have equal number of stars so it was difficult to distinguish them.”

P4 was even amused at the power of the star ratings, especially since he did not see them relevant: “Well, it is, I mean I notice that I am looking at it [stars]. It is
that somehow, I find it interesting myself, … they do affect anyway. I mean here I have three books in front of me and two of them don’t have any stars and one has, and so I immediately feel like clicking this one with five stars. … I mean there could just be one guy who’s written something and there he’s written that a brilliant book, hallelujah, and that the review is absolutely pointless for me but it matters because it is because of the stars I go to see the book.”

The last quote by P4 in fact captures the major point of frustration among the participants who did use the star ratings to select which books to look at more carefully. All wished to know how many reviewers were behind the average rating because the importance of five stars from one reviewer depended entirely on the review’s content while five stars on average from dozens of reviewers already indicated more about the quality of the book. P2: “If there are five stars and it’s been read [reviewed] by 20 people and it’s still that [five stars], it already tells that there’s been all kind of people in that group. That’s a good thing. But if there’s only one, then the book can be anything at all from my viewpoint. The more reviews behind the stars, then maybe the likelihood of the stars being useful increases.”

While Amazon has already corrected the immediate problem by adding the number of reviewers behind the stars on the list page, the deeper point here is the need for transparency in recommendations. As discussed earlier, many researchers have underlined the importance of transparency to help the user and here we have a clear case of users needing the information behind the recommendation to evaluate its relevance.
While arguably a prediction of how individual users would rate the items themselves would be more useful here, the sparsity problems and computational requirements of making such predictions for each item on the list page are likely to prevent us from having them.

However, the question has been asked if showing the average is the most efficient way to assist the user here, and alternative ways are being researched [Ludford, 2007]. The fact that some of the participants in this study were interested in the distribution of the ratings argues that there might be a better way.

The most important factors on the list page for the participants in general were the book title and cover picture. The importance of the title in non-fiction books has long been known by publishers: “the title is the book you buy” [Cole, 2003 pp. 19-20]. The title could get the participants to look at the item page without any other factors affecting the decision—P5: “This book has a very interesting title. It caught my attention immediately. There’s ‘creative thinking’ here in the title. It immediately gives it its own perspective”—while a poor title prevented the participants from even considering the book. P6: “That name… ‘Step-by-step’ is a bit; I’m not into those ones.” P1 nicely summed the point up: “You can even make the title of a billet-chopping machine servicing manual interesting.”

The cover picture was also very important to the users, again something known to the people in the publishing business: “the cover sells the book” [Cole, 2003 p. 21]. A bad one could lead to rejecting the book. P2: “If the cover is absolutely senseless like in that one, I just jump over it.” In contrast, a good cover worked, in words of P6, as a “waking” factor.

P4 was the only one not to place much importance of the cover: “…I was just thinking that they are so small, the pictures of the cover, that in practice they don’t matter here.” Also P5 considered the covers too small even though he did look at them. Amazon has also remedied this by having made the cover pictures significantly larger since the study was conducted.

One piece of information that was significant enough to cause the participants to reject a book was not shown on the list page: Number of pages. A book could be too long for some needs—P4: “I’ll check that it is not some huge behemoth of a book because with these coordinates I’m not up to reading one of them. A snack-sized one would be good”—or too short for other needs: P3: “One hundred and sixty pages. This doesn’t… I drop this just because it has so few pages [laughs], I mean for this
topic...” Furthermore, number of pages could be a reason in itself to make the book appear interesting. P3: “Oh! Here when I see the number of pages [408], I’m of course interested.”

Identifying the information that can alone lead to the item being rejected out of hand or make it appear very interesting alone is crucial to making list pages useful to the users. Furthermore, the way the information is presented determines if users can truly benefit from it.

4.2.3 Helpful ratings on Customer Reviews

“It’s one of those pieces of information that I take with a grain of salt. It’s maybe a little, well, I wouldn’t put much weight on it.” – P6

Users who are signed in can rate individual “Customer Reviews” as “Helpful” or “Not helpful,” and the number of those who found the review helpful is contrasted with the number of all raters above the review to assist users in selecting which “Customer Reviews” to read (Figure 8).

However, unlike with the star rating on the list page, none of the participants placed any emphasis on the helpful ratings of customer reviews when selecting reviews for reading. Three of them, P2, P3, and P4, had not even noticed the feature before the study session. P2 and P3 only realized that there was such a feature when asked about it while P4 noticed it during Task 3.

4 of 5 people found the following review helpful:

Jack of all Trades, 10 Sep 2007

olve all my reviews

Perhaps this book is aimed at compact camera users, or should be as it is very much a beginners book. But as a DSLR newbie I found it too time consuming to read as it’s rather basic and repetitive, talks about presets which I’ve no interest in and is rather dull reading for someone with prior knowledge. Could be good for someone who is completely new to photography.

I bought this as it was suggested as a good accompaniment to the Scott Kelby book. The Kelby book is incredibly useful, fun to read and succinct, exactly what I wanted. I don’t think this book was a good suggestion for pairing together.

Figure 8. Helpful rating is located above the review title and star rating.

In fact, P4’s experience in Task 3 was the only indication that the helpful rating was taken at all into consideration, and showed an example of how a user noticing the feature for the first time approaches it.

“Here’s this C. Ebbing has written one where 18 out of 19 have seen it or considered it useful. I don’t get it. What does it mean? Out of what 19?
Apparently I could go here and click—I mean if I could be bothered... I don't. I mean I think that these kinds of things... Well, here's another one where 95 out of 95 have considered it useful. Let's read this one and see what it says... I mean just because of the volume. Here so many people have found it useful. Almost one hundred, so I take my place in the line and see what it says.” – P4

P4’s comments again indicate that as with star rating averages, if helpful rating has any impact at all, its impact depends on the mass of users. However, the large number of Helpful votes only extended to arousing P4’s interest but not to affecting his view of the review: He found the review unhelpful.

The other three had noticed the feature before but did not consider it relevant for selecting which reviews to read. P1 saw the number more as an indication that people were interested in the book or in the topic rather than as any kind of indication that the book was good or bad. He did not entirely even connect the helpful rating to individual reviews in that sense. However, he was the only participant who actually had voted some reviews as helpful prior to the study session. He had marked some reviews “Helpful” that had convinced him that the item was the one that he wanted.

In contrast, P5 and P6 had a clear idea on the meaning of the rating but found the rating irrelevant. P5 questioned the whole premise of the rating in scathing terms: “I think the whole concept is useless [laughs] because I can’t know if the review was useful or not before I buy the book and read it.”

Interestingly, the premise of helpful ratings was not unambiguous. P1 considered a review helpful if it helped one to decide whether or not to buy the book while P5 thought that one had to read the book before one could know how good a review was, that is, whether it was helpful or not. P4, who tried to mark a review as helpful, saw the situation the same way as P5 did.

The wording in the feature—“Was this review helpful to you? Yes / No” and “... of ... people found the following review helpful”—is ambiguous, and the resulting different understandings again underline the need for transparency in all aspects. If the question had been worded “Did this review help you to make up your mind? Yes / No” or “Did this review match you view of the book? Yes / No,” the concept behind the rating would have been clearer. Every word matters in the interface because that is what users have at their disposal when building mental models of the features. Now it is likely that the rating mixes ratings of people who
commented on whether or not the review helped them to make up their minds about buying the book and people who expressed agreement or disagreement with the review’s content.

4.3 Reviews, Search inside and Customer Reviews

4.3.1 Overview

While we are mainly interested in how “Customer Reviews” were used, they cannot be taken out of their context. The evaluation of a book that ultimately leads to buying or dropping of it takes different factors into consideration, and to discuss “Customer Reviews” without mentioning the other factors would be unfair. The decision is like a jig-saw puzzle: All pieces matter although some pieces are more equal than others in forming the picture.

Although the item page contains numerous features, the relevant ones for evaluating books were “Search Inside,” “Customer Reviews,” “Reviews,” and “Product Details” for the participants. The one feature that was used often in addition to these was clicking the author’s name to see other books by him of her. While this was often done in order to find other books by the author, it was also done to get an understanding about the kinds of topics the author had covered to put the current item into this context.

The other features were either never used—and the never used ones constituted the large majority—or they were used once or twice, almost as if by accident. For instance, P2 used “Customer Images” in Task 2 when the feature had replaced “Search Inside” (same location and look) without noticing that it was not the feature that he was looking for.

While most looked at “Product Details,” typically the only relevant information it offered was the number of pages in the book. As the number of pages could be a decisive factor, causing the book to be dropped without regard for other considerations, this piece of information should be made available higher on the item page and it should also be featured on the list page. Now P4 had difficulties finding the information, and he tried to find it from “Search Inside”: “Eh, where’s the number of pages said here? It’s not here. [Returning to the product page] Is it here, then?” P5 alone was not interested in the number of pages and the four others found the information easily on the item page.
The two main things that the participants wanted to know about a book were: What is the content of the book and is this book the right book for me? Even the number of pages was used for understanding the scope of the book. P6: “Next, let’s see—147 pages. Not a very thick one, is it? Can’t be that thorough, either, then, can it?”

Of “Search Inside,” “Customer Reviews,” and “Reviews,” “Search Inside” was clearly the most important feature for the participants. P2: “I want to be able to open it, Table of Contents, and a few pages. In a way, that’s what’s the most important, the best thing for me.” “Customer Reviews” were also considered very important, and they could be enough alone for deciding if there was no “Search Inside.” On the other hand, “Reviews” were often read but none of the participants emphasized their importance the way they emphasized the importance of the other two features.

We first take a look at “Search Inside” and “Reviews” use before looking in more detail at “Customer Reviews” use.

4.3.2 Search inside

“The others [without ‘Search Inside’] are as if you were a blind man looking at a book.” – P3

“Search Inside” was seen as a way to browse the book and see it as it really was. P5: “It really is like browsing a book in a bookstore. If I am in a bookstore, I might browse tens of books even if they are only marginally related [to what I want], and I like to do that here, too.” The participants emphasized seeing the style of writing, the pictures and how the book was printed, the layout, and the organization of the content. These factors were related to using the book, the quality of the book, and if the book was fun and easy to use. As the participants were looking for non-fiction books, the organization of the content was important for the ease of using the book as well as for understanding its scope.

However, the most loved feature of “Search Inside” was Table of Contents. P2: “You know, what in the book itself helps the best is if I can see for instance Table of Contents. That’s a lot already.” Table of Contents cut to the heart of the matter, what the book covered, that is, its content. The style and general look and feel of the book were important but the understanding of the content was the most important aspect
that the participants felt “Search Inside” offered. Some other online bookstores, for instance, ILEX online bookstore19, show the Table of Contents for most books, and based on the wishes of the participants in this study, this is certainly a good alternative to having “Search Inside.”

Table 10 summarizes how often “Search Inside” was accessed and how it was accessed and closed. Most participants used “Search Inside” simply as a way to browse the book, moving from page to page in the given order. Only P4, who only discovered the feature during this study, used it to actually search for keywords inside the book. While others considered “Search Inside” a reasonable alternative to having the physical book in their hands, P4 considered it in some ways better than browsing the actual book: “Actually, it’s even better than browsing the book in a bookstore. You can search—I mean as long as it worked like it seemed to work pretty well—you can search with a keyword, what was important to you. And then it tells you that the thing that’s important to you is only mentioned on one page. I didn’t need to browse it any further. I knew that it wasn’t my book.”

<table>
<thead>
<tr>
<th>“Search Inside” usage in Task 1 and 2</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Used “Search Inside” in Task 1</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.8</td>
</tr>
<tr>
<td>Used “Search Inside” in Task 2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Used “Search Inside” altogether</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>2.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>“Search Inside” accessing method</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessed by clicking link below the picture</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Accessed by clicking cover picture</td>
<td></td>
<td></td>
<td>2</td>
<td>3</td>
<td>2</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>Accessed by mid-page’s “Inside This Book”</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Accessed by pop-up layer interface</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>“Search Inside” closing method</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Back button navigation</td>
<td>2</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>Using the link back to the item page</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Closed the tab</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

Table 10. How Search Inside was used in Task 1 and 2.

Thus, “Search Inside” was used to get the look and feel of the book and understand its content. However, even with “Search Inside” available, the participants uniformly looked at “Customer Reviews” for further understanding of the content and

19 http://www.ilex-press.com/
how the subject matter was approached in addition to trying to calibrate if the book was the one for them.

4.3.3 Reviews

“There’s one review [‘Reviews’] and that’s written by whom? By the author [laughs] himself? I don’t know where this info is taken from.” – P1

The participants saw “Reviews” as a potential source of information about the contents of the book although they were also used to see if the book was the one for them, especially if the expertise level was correct. However, “Customer Reviews” were more often used for the latter.

Occasionally, some participants said that they wanted to read the “Customer Reviews” and then proceed to read “Reviews” instead. The two were not always clearly distinguished from each other. Apparently Amazon has reached the same conclusion as the feature has been renamed to “Editorial Reviews” in Amazon.com and “Product Description” in Amazon.co.uk. This is certainly a step in the right direction as the features have to be distinguishable from each other to allow users to make up their minds about how to approach their task in the environment.

The confusion was further increased by the fact that the “Reviews” section contained a rag-tag of different types of information, such as “Synopsis,” “Book Description,” and various reviews from other sources titled variously (for instance, “photo-i Website July 2006”). Sometimes the sources were clearly mentioned while other times the participants were left wondering about the source of the review. Unclear source did lead to reliability problems, and P4, who tended to place more trust on reviews by professionals, ended up accepting one review only after checking the website where the original review appeared. A direct link would have assisted checking up the reference and would have been in line with the Stanford’s Web Credibility project’s recommendations [Stanford Persuasive Technology Lab, 2004].

Overall, “Reviews” were a secondary source that typically required support from “Customer reviews” or “Search Inside,” and no book was bought based solely on the information in “Reviews.” If no other information was available, none of the participants would have bought a book solely based on “Reviews.” On the other hand, books were occasionally rejected based on them. P2: “/The first sentence on ‘Reviews’ is: ‘This is the nearest we've come to a digital imaging bible on outdoor
photography.’ ] Yeah, this concentrated on outdoor photography. I wouldn’t take this one [returns to the list page].” Thus, “Reviews” provided supporting information, not decisive information for purchase decisions although they could be crucial to decisions not to purchase.

4.3.4 Customer reviews

“Like an old maxim goes, opinions are like a certain part of human anatomy, everybody’s got one. The Internet is full of information and opinions. You can’t take it at face value. You first have to figure out who is the person who says something. And that’s why I somewhere earlier, when someone had given two stars, I didn’t really care because I immediately saw that he didn’t know what he was talking about. … So I knew the guy was stupid and so his two stars were irrelevant. That’s what’s important: You have to figure out who’s this guy who’s reviewing.” – P5

All participants stated that “Customer Reviews” played an important role in deciding whether or not to buy a book but that they were typically not the only factor. Table 11 summarizes the effects of “Customer Reviews” on purchase decisions. “Customer Reviews” tended to have a more pronounced preventive than encouraging effect, as five participants reported having decided not to buy one or more books solely based on “Customer Reviews.” (P6 said that he probably had not decided not to buy a book based on “Customer Reviews” alone, but he did drop a book in Task 2 based on one customer review’s say-so.) In contrast, only three participants reported having decided to buy one or more books solely based on “Customer Reviews.” (With only two books bought, P4 had not done either.)

<table>
<thead>
<tr>
<th>Importance of “Customer Reviews”</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do customer reviews play a part in your buying decisions?</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Have you decided not to buy a book due to negative customer reviews?</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Have you bought a book based on positive reviews?</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 11. Importance of “Customer Reviews.”

Observation underlined that “Customer Reviews” could have a pronounced negative effect if even just one of them contained some significant piece of information that indicated that the item was not the right one for the participant. In fact, participants often dropped items based on just one sentence or phrase in a
customer review when it indicated that the item was not right for their level of expertise or otherwise unsuitable for them. P6: “Now I’ve got to read why this one has only liked worth one star. [Reads aloud from the review] ‘It is more like a book marketing Adobe to beginners.’ I believe that [returns to the list page].” Likewise, P5 dropped a book because of just one phrase: “…that first phrase when they talked about moving from film to digital photography. I knew instantly that I did not belong to the book’s target group.”

P4 also dropped a book he was about to buy in Task 1 because of one piece of information in a customer review: “Then at the end he said that he’d recommend it only to die-hard fans. …it was a good review for me because I got a feeling that perhaps I’m not a die-hard fan even though the book seemed really nice when I was browsing it [with Search Inside]…”

In contrast, while the decision not to buy an item could be sealed with one sentence or phrase in one customer review, the purchasing decision was typically a sum of many factors. The decision to buy a book was never based on one sentence or phrase during the study. However, based on observation and interviews, it is perceivable that one thorough customer review that “talks” to the user can alone be a deciding factor. P5: “And one can really be enough. Of course, if there are more, then the decision is very easy.”

Overall, “Customer Reviews” served the participants in two ways. First, they were seen as sources of information about the contents of the book. For this, the stars and the positive or negative recommendation inherent to the customer review were irrelevant. The participants wanted to know what the book contained, and “Customer Reviews” were seen as a source for this information together with “Search Inside” and “Reviews.”

Second, “Customer Reviews” were used to evaluate if the book was the right one for the participant, given his own level of expertise and type of need. The participants reflected the reviewer’s needs and level of expertise against their own to see if the review contents and the rating were relevant to them. P1: “A person who knew more about the topic says that there’s not enough information. A beginner who liked it felt that it was suitable. Well, I’m not a stark beginner and so I’m looking for something more thorough.”

A negative review was typically not seen as a deterrent if the reviewer’s needs or level of expertise in the field were different from the participant’s. Likewise, a
positive review did not count for much if the participant did not see the relevance of what the reviewer said in relation to himself. P2 said that depending on who the reviewer is, “you could mentally turn the rating upside down or not.”

In fact, a negative review can seal the decision to buy if the needs of the reviewer and the user are diagonally opposed. P1 told how he had been looking for a digital photography book for his workplace. He had searched Amazon together with one of his colleagues for a suitable book. They had found what seemed like a good match and read the customer reviews, and decided to buy the book based on a very negative customer review. Showing the review, he said: “Well, somebody here is moaning that it’s written in an American style and that it talks too much about memory cards and all other stuff and that’s no good. But we were looking for exactly the kind of book that would talk about everything related to single lens reflex cameras, tripods, flashes, memory cards, bags, and how they’re used and why and where. So, it was exactly what we were looking for even if that person didn’t like those aspects.”

In the same way, a positive review by somebody whose level of expertise differed clearly from that of the participant could constitute a reason not to buy the book. P4: “And in that sense, if somebody who was a great guru of photography was praising a book sky high, I would get a feeling that it’s too advanced for me.”

Mirroring the reviewer’s needs against one’s own involved understanding the person behind the review in order to understand the meaning of the review. P4: “…that I can somehow imagine the person who had written it in my mind and I somehow understand that okay, this is what it means.” Thus, the process appears not only intellectual but also seems to involve emotions. As we did not measure the participants’ emotional states, more research is necessary here. However, evaluating the reviews certainly appears to be at least partially a social process.

The star rating of the review became almost irrelevant when the review text was available. The participants wanted to see the reasons behind the rating and then decide if the review was relevant to them at all. P1: “…if he gives four stars then what is the reason why he gives four stars.” Consequently, the participants uniformly liked the reviews to be somewhat longer, as they all felt that the shorter ones cannot give proper reasoning for their conclusions.

While four participants claimed to typically read about five reviews when available, observations did not support this. The common start-off strategy seemed to be to pick one to three reviews for more careful reading after a general glance. No
participant went to the second page of reviews, and some were not even aware that
there would have been more reviews available. P1: “I didn’t even notice that there
were so many reviews [laughs].” This is in contrast to the list page, which could be
scanned more quickly, and consequently it was not rare for the participants to view 2–3
pages.

While the participants claimed to favor longer reviews that were written in a
matter-of-fact style—P1: “The ones where the language is clear and the reasoning is
shown clearly, those ones I read through”—they typically first took an orienting look
at the reviews before starting to read them, just as they tended to quickly scroll the list
page through for a quick impression before looking at the books in detail. This gisting
(the term is borrowed from [Rivadeneira et al., 2007]) or gestalt forming look gave
the participants an impression of how uniformly the ratings were distributed in
addition to showing them what the general tendency was. While it is difficult to
determine how big a role this impression played in the final decision based on our
data, it seems likely that it did have a priming function, giving the participants a
positive or negative first impression. Furthermore, the perception of the distribution
could direct which reviews were read. P5 gave a good description of what happened
during the gisting phase and how it affected the subsequent choices:

“First I take an overview. If there’s, say, 10 reviews, you immediately see if
the opinions are greatly dispersed or is it very uniform. If it’s uniform and
everybody has given five stars or if it’s uniform and everybody has given two
stars, then you can trust it, especially if there are lots of people: Ten, twenty,
sometimes even one hundred. You know, if a consensus has emerged. Then it
could also be like, for instance, that the majority has given good ratings and
one or two have given something totally different from the mainstream. Then I
have to read carefully and go through this process. I try to see which group I
belong to. Am I with the resistance or part of the bigger mass, and in this
situation, the separation is important.” – P5

Thus, while the star ratings lost their relevance when the review text was
available, they nevertheless affected in part which reviews were read. P1, P5, and P6
compared the star ratings of the reviews to see if the reviewers had reached consensus.
P3, P5, and P6 tended to check out the extreme views that were identified by high or
low stars and did not find middle-of-the-road reviews useful. P4 considered negative reviews the most useful because they pointed out the potential problems. For P1 and P2, the positiveness or negativeness of the reviews did not really affect their choice. P2 tended to seek for reviews by reviewers he knew personally or whose books he had read, but if he did not find any, he tended to read one or two reviews from the top, as did P3 in case there were no extreme ones.

In contrast, as discussed, helpful votes had negligible effect on which reviews were read. P2, P3, and P4 had not noticed the feature prior to the study, and the others did not give it much consideration in selecting the reviews for reading. This leads us to question if the correlation between reviews with the high proportion of helpful votes and additional sales observed by Chen, Dhanasobhon, and Smith [2006] was in fact due to the review qualities rather than the helpful ratings.

The title of the review was clearly an important factor in selecting reviews for reading. They were scanned for information on the content and viewpoint of the review. The importance of the review title mirrors the importance of the book title in selecting books for a closer look.

Another selecting strategy appeared to be scanning for certain keywords in “Customer Reviews” and in “Reviews.” For instance, in Task 2, participants looked for words like “beginner” and “professional” from both types of reviews. Some wanted an elementary book while others wanted a more advanced book, and these words indicated for whom the book was written, or the level of expertise of the reviewer, both of which were important information for the participants. Determining how far this keyword scanning goes in determining which reviews are read requires further research, but the implications are interesting, as this might enable us to guide users to the reviews that are relevant to them for instance by using keyword clouds where the central words of the reviews could be presented like tags in a tag cloud, and users could click a keyword to access all customer reviews where the word is used.

All in all, the process of selecting customer reviews for reading appeared only partially conscious, as P2 said: “It comes from somewhere deep from some random number generator, that’s where it pretty often comes from... Yeah, you could say it’s some kind of a feeling, and it’s rather hard to rationalize why the ones that were picked were picked.” The conscious factors were there, but the data from this study only partially illuminated the whole process and further research on this is necessary.
When establishing the relevance of a customer review, the participants paid close attention to the language. Two participants expressed clear dislike for emotional tone. Bad English, misspellings or grammar problems, and improper tone were also common turn-offs. Thus, the language used did clearly affect the assigned relevance and reliability of the review. Furthermore, the language used could persuade the participant to move on. P3: “If there’s some blathering that doesn’t seem to have a head or tail, I move on to the next one.”

While five participants said that the name the reviewer goes by was important to them, none of them had noticed Amazon’s “Real Name™” badge for verifying the identity of the reviewer. They did not warm to such pseudonyms as “serialkiller” or “Shit-4-Brains.” However, the effect of a questionable pseudonym was more to prevent reading the review than affecting the reliability of the review once it was read. P2: “Of course they tend to affect it somewhat, but it affects more my inclination to read the review than what I think after I have read it.” Once a review was read, its content determined its reliability and relevance.

Only two participants had tried to look at other reviews written by the reviewers or reviewer profiles to get further information on them. No participant had ever learned to recognize and trust a reviewer by reading the reviews or by other means Amazon provides for getting to know more about the reviewers. Overall, the content and style of the review appeared to be much more relevant factors in determining the relevance and reliability of the review than author-related factors.

The number of book reviews by customers in Amazon can run in tens and even hundreds or thousands and span even hundreds of pages. For example, Harry Potter and the Sorcerer's Stone (Book 1) had 5,404 customer reviews spanning 541 pages (May 20, 2008). As this study indicates, users can have strategies for selecting which reviews to read, and so the service should provide tools for using these strategies effectively. Furthermore, such tools could also help others to formulate such strategies.

At the time of the study, there were no tools for organizing “Customer Reviews.” They were simply given in the order of recency, a fact of which the participants were not even aware. Three guessed it correctly when asked but admitted not knowing for sure.

In effect, Amazon has greatly improved its “Customer Reviews” feature, showing the distribution of ratings clearly—and thus giving rise to user strategies that
utilize this information—and giving the reviews in two columns, one ordered by recency and one by helpful votes (Figure 9). Our study underlines the need for such tools. However, the tools provided by Amazon are not personalized, and so there is a need for research on how to personalize the review use experience, how to recommend the most relevant reviews or bits of reviews to the user in question.

Customer Reviews

<table>
<thead>
<tr>
<th>Rating</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 star</td>
<td>(19)</td>
</tr>
<tr>
<td>4 star</td>
<td>(2)</td>
</tr>
<tr>
<td>3 star</td>
<td>(2)</td>
</tr>
<tr>
<td>2 star</td>
<td>(1)</td>
</tr>
<tr>
<td>1 star</td>
<td>(1)</td>
</tr>
</tbody>
</table>

Average Customer Review: 4.5 stars (23 customer reviews)

Most Helpful Customer Reviews:

41 of 43 people found the following review helpful:

🌟🌟🌟🌟 The Importance of Living, July 5, 2005
By Michael N. McIntyre (Eagle, NE United States) - See all my reviews

This review is from: The Importance of Living (Dover Books)
I bought a copy of this book (the original 1937 edition) in a secondhand book store in the 1970s for the princely sum of $1.00. Through countless moves since then I have somehow managed to retain this book, which is surprising since I have lost or given away so many books, almost all of which cost me much more money than this one did. I can honestly say it was the most profitable dollar I ever spent. This is a wonderful book -- rambling at times, it is true -- but it contains many gems. Yutang is a superb writer and his quote of Chuangtze (as he spells the name of the famous Chinese philospher) is classic: "Split forth intelligence." This, along with William Sturk's famous dictum "limit needless words," is a phrase every writer should live by.

Most Recent Customer Reviews:

🌟🌟🌟🌟 Lin Yutang and the New Age Left: There is a Noble Art To Doing Nothing
I first heard of Lin Yutang, I am almost reluctant to say, many years after my exposure to Alan Watts and the "New Left", long before the "New Age" came to be a force... Read more
Published 2 months ago by Chieh Ching

🌟🌟🌟🌟 What was true about Chinese and Americans no longer hold, but the brilliance of this book is for all human kind!
What was true about Chinese and Americans no longer hold, but the brilliance of this book is for all human kind! Read more
Published on October 21, 2005 by Chieh Ching

🌟🌟🌟🌟 A delightful book to savour
Lin-Yutang (1895-1976), a Chinese humanist and humanist, was steeped in the ancient wisdom of his motherland. Lin-yutang was also a cosmopolitan. Read more
Published on September 24, 2009 by Hakuju

🌟🌟🌟🌟 Welcome back to your Childishhood!
First off, allow me to say that I'm shocked this book hasn't had more reviews. This book

Figure 9. Current “Customer Reviews” feature interface in Amazon.com.

4.3.5 Customer reviews vs. expert reviews

“It depends on the topic and matter. If I needed to get some computer-related thing for work or something like that, then expert review would probably be somewhat more important. If I were looking for a book to read in my summer holiday, maybe user reviews would affect the decision more. That's the way. It's perfectly natural that there's a difference and that it depends on the situation. I'd say that both are good but they have their own place that depends on the situation and context.” – P6

On surface, P1, P3, and P4 said that they are more inclined to trust expert reviews than user/customer reviews. However, their behavior in Amazon did not confirm this, as “Customer Reviews” were clearly more important to them than “Reviews.” The
problem might partially be due to the fact that “Reviews” contained various materials, and the participants did not get a feeling of reading an expert review. P1: “So here’s just one review [‘Reviews’] and that’s written by whom? By the author [laughs] himself? I don’t know where this information is taken from.” As mentioned, Amazon’s new naming policy can help to clarify the situation: “Reviews” are renamed “Editorial Reviews” in Amazon.com and “Product Description” in Amazon.co.uk.

P2 did not trust an expert to be an expert just because of somebody’s say-so: “First you’d have to know who the editor is. That’s important. If he’s good, then it’s ok. On the other hand, how many darn editors are such experts that they know everything about everything? They don’t exist. I mean if he’s reviewing very wide spectrum of topics. Perhaps we could say that if the editor is very knowledgeable in his field, then maybe the feeling is that it’s more reliable than customer reviews. But it could be the reverse; the guy could be a total idiot.”

P5 and P6 echoed P2 in questioning the definition of an expert, especially in the context of the Internet. P5: “Experts? Weeeell… The line between an expert and a user is rather unclear on the web. …especially if we are talking about technical devices, computers and digital and stuff, the fact is that the hobbyists are often more knowledgeable than the so-called professionals. They know more and so you can find on forums and such much more exact information than from a more official source.” P6: “…they’ve been advanced hobbyists, they have known things very well, and in that sense they’ve been experts.”

Overall, both were seen to have their own place and to provide important information from two different perspectives. Neither could replace the other nor provide a complete picture alone. P6: “A user perhaps pays attention to different types of things than a real expert. That way you end up with somewhat different type of feedback, which is important: You get a little different viewpoint and that way you get a more complete understanding. … I think they both have their places.”
4.4 User-generated content: Motivations, perceptions, and contributions

4.4.1 Contributing to user-generated content

“I can’t really imagine myself writing a review of Dharma Bums on the Internet. And probably in the same way I would find it difficult to click that button to say if I agree or not. On the other hand, I think that the system is good in principle; it’s just my way of seeing it. I somehow find it strange for me. Probably it’s just that I’m not so familiar with it.” – P4

Only P1 had contributed to Amazon by marking some reviews “Helpful.” P4, who had not done so before, tried to mark one review “Helpful” during the session but gave up when he was required to sign in.

The reasons for not contributing varied. P1 did not like writing in general, but in principle had nothing against contributing to the community. P2 had once written a message to a discussion forum and had gotten so many negative comments that he had decided to stay out of all web writing, including even voting a review helpful type of features. P3 felt that his English was not good enough for him to stand behind his writing, but otherwise he was comfortable with the idea of contributing. P4 felt uncomfortable with writing reviews, conjecturing that perhaps he just was not familiar enough with doing it. P6 had written some reviews to forums related to his work but never to Amazon. P5 was most vitriolic about making contributions. He said that he was “lurker” and that he could see no point in contributing: “…in Amazon I feel that I am a buyer, and to write a review would just take my time so why would I do it? I don’t feel that I would benefit from it in any way.”

The inclination to contribute might be related to the perception of social presence. P3, P4, and P6 perceived Amazon as having social presence, and P4 tried to contribute during the study, P6 had contributed in another context, and P3 excused himself because of language, not because of having anything against contributing. P1, who had contributed prior to the study, did not perceive Amazon as social environment but saw a possibility of seeing it in those terms if he used it more.
In contrast, P5 and P2 saw online shopping as a distinctly unsocial process, and P2 professed preferring to stay away from big communities in any case: “...I'd rather stay away from these huge communities. I don’t like them much at all.”

The connection between perceived social presence and propensity to contribute seems to be there, and so we conjecture that if users feel the presence of others, they are more likely to contribute. While establishing this firmly requires further research, it is an interesting possibility because that would give us actionable knowledge on how to create an atmosphere inductive to contributing.

Consequently, we conjecture that increasing perceived social presence leads to increases in both the quality and the quantity of the user contributions. Moreover, Cosley et al. [2005] have already shown that oversight increases contribution quality and decreases anti-social behavior even when users are not told of the oversight explicitly.

Considering that Amazon.com has much more customer reviews than Amazon.co.uk—some participants even went to Amazon.com to read reviews even though they shopped in Amazon.co.uk—it is possible that there is also a cultural factor at work. The fact that being able to write a review in Amazon.com requires a purchase while in Amazon.co.uk only signing in is required might also reflect this.

4.4.2 Perceptions of people who contribute and their motivations

“...if a failed author becomes a critic, does it then mean that a failed critic becomes a critic in Amazon?” – P4

All participants used “Customer Reviews” widely to assist them in purchase decision making. Consequently, it might come as something of a surprise that on the whole the participants did not perceive the reviewers themselves in a very positive light.

Although P1 occasionally questioned the motives of the review writers during the tasks, on the whole he had a positive view of their motives. He considered them extraverted people who liked to write—unlike him—and wanted to assist others in decision-making. P2 thought that while there were people with altruistic motives among the writers, the majority was still people who simply had to say something about everything, whether they knew anything about it or not. P3 remembered cases where the author of the book had written about his or her own book or the reviews were otherwise fakes. P3: “You can well ask if you can trust these all.”
P4 thought that the review writers included many kinds of people. Some pushed some products or trashed competing products for personal benefits while some were neurotically interested in some area or topic and spent all of their free time writing online about it in various arenas. He thought that some were also motivated by the current atmosphere of everybody trying to be somebody, and that they wrote for the same motives that made some others compete in American Idol. Everybody had to have an opinion and Amazon was just another arena for having an opinion. Finally, he did also think that some of the writers were motivated by selfless willingness to help and contribute: “Then some of the writers have a citizen-society idea. They really just want to share what they know, and that’s good, really good. And perhaps that’s the thing that I was talking about earlier; you look for the ones that seem to talk to you. Perhaps it’s like that if I need to define it somehow. The writer gives you the feeling that his motivation, why he does it, is that he knows something and wants to share that information with others. And in that sense it’s a selfless motive.”

P5 admitted also that some people were motivated by altruistic goals, but felt that many were people to whom “their own voice is the most beautiful voice in the world and the text they spew out is the most beautiful text in the world.”

P6 had somewhat more positive outlook. He considered the review writers typically to be people who read a lot, were interested in the topic, and felt that they had something to say about it.

Considering how negative perceptions the participants had about the review writers, it is interesting that “Customer Reviews” were nevertheless so widely used to assist in decision-making. It is likely that the perception during the actual use and reflection on the matter diverge here. The comments above were reflections while the actuality was that all participants clearly used “Customer Reviews” and were confident in their ability to pick out the ones that were relevant to them. The importance of picking out the good ones—reliable and relevant to the participant—was emphasized repeatedly in the interviews. Interestingly, the overall quality of “Customer Reviews” was considered good in Amazon.

4.5 Social aspects of complex information environments

“It of course does have a feeling of a community, what with all the reviews and what not. You get a very clear feeling of a community, but it is difficult to define it, to define
what it means in practical terms. It most likely affects clicking these buttons and that kind of things.” – P4

We were interested to see if the participants would perceive a complex online shopping environment as social because of the inherent social texture consisting of recommenders and customer reviews in addition to other signs of other users being present in the environment. We were looking at two aspects. One, if the participants perceived other shoppers as socially present in the environment, and two, if their behavior was affected by that perceived presence.

In effect, the participants neatly divided into two groups as far as perceiving the environment as social was concerned. P1, P2, and P5 felt that they were alone in the online shop and that the social texture did not make the environment any more social. P1: “It is more like a convey belt than a social environment. …I don’t really see it as a social environment and the reviews by anonymous people don’t help to make it any more humane.”

While P2 and P5 found no social aspects whatsoever in Amazon, P1 did relent his position a bit. He remembered having looked at the other reviews of some reviewers once or twice and having had a feeling that “he’s interested in the same things as me.” He thought that if he bought more books, spent more time in Amazon, and consequently looked more at other reviews by reviewers and used other similar features, he might begin to perceive the environment as more social.

For P3, P4, and P6, on the other hand, the social texture made the environment inherently social. P3 felt that the recommendations “enlivened” the environment and that without personalization and personalized recommendations it would appear “dead.” Likewise, P6 felt that the presence of the community was very positive: “It’s like that, you know, ok, yes, others had felt the same thing about it, about this book, and oh, ok, he thought like that, I don’t agree but it’s good to know that people can see it like that, too. It goes like that; the community emerges out of it.”

P4 perceived the environment as even more social than the other two and he also felt clearly that the social aspects affected his behavior in the environment. P4 explained how the social aspect had affected his behavior when he found a review that 95 people out of 95 had found “Helpful”: 

78
“I felt that it wasn’t helpful, but like I said, I didn’t click the button because then I’d be a killjoy. That’s where the sociability kicks in. Then there was one where three had read it, I mean, had evaluated it and all agreed that it was not helpful. So I somehow thought that I’ll rebel against it and be the first to think that it is helpful. Then I’d actually do something positive [laughs]. That I didn’t click the button or that I would have clicked the button, the motivation didn’t have anything directly to do with the book or even the review but all to do with the social context and how I perceived that social situation. …the critical mass of Joe Blows, then the social dimension kicks in and those who disagree no longer have the face to disagree [laughs] and do it [vote against the majority] when the critical mass has been reached.” – P4

Cosley et al. [2003] found that what is shown on the recommender interface affects the evaluation and that oversight improves the quality of user contributions and decreases anti-social behavior [2005]. Now we have evidence that part of that effect can be due to social factors. We hypothesize that the more users perceive other users as being present in the environment, the more they take them into consideration in their behavior. However, this effect requires further research so that its full implications and extent is understood.

However, even P4 stopped short of feeling that he was part of a community: “…but perhaps I don’t really feel like I’m a member of a community in the full sense of the word.” For P4, too, at the end of the day Amazon was a bookstore where he went to buy books for himself.

While here we discuss the perception of social presence, all the participants did use social skills and social cues—albeit in socially poor circumstances—to assess the needs, levels of expertise, and even personalities of the “Customer Reviews” writers to mirror them against their own to assign relevance and reliability to the reviews. Furthermore, arguably decisions influenced by “Customer Reviews” to buy or not to buy a book or to look at a book in more detail at least partially because it had five stars, were all actions that were influenced by the social texture. However, we feel that what makes an environment social or not is the perception. If a user perceives that other users are present because of the recommendations and reviews, then the environment is social for that user, and if a user perceives recommendations are part
of the convey-belt shopping environment, then the environment is not social for that user.

Consequently, the social texture in Amazon was not alone enough to make all participants perceive the environment as inherently social. How easily people perceive an environment as social is probably related to their personality and personal definition of sociability. Some people need only a slightest of hint to perceive an environment as social while others require synchronous conversations with video image before they perceive an environment as social. For instance, P5 did not even see going to a brick-and-mortar bookstore as social activity: “...I don't go to a bookstore to be social.”

In addition, some people are distinctly uncomfortable with communicating online with unknown people. For instance, P2 stated that he would not want to communicate with anybody directly in Amazon: “Not before if it some day is unavoidable and you have to do it not to become a recluse. But maybe, well, I’ll have to see about it then.”

Consequently, making complex information environments involves a balancing act of offering social features to the people who want the social aspects and see benefit in them, and allowing those who want to make their work in relative solitude that right. Having ways to opt out of social aspects is important in the light of this study. People know what they are comfortable with, and features need to be transparent to allow people to evaluate them, and the features should offer a way to opt out for the people who feel uncomfortable with them.

4.6 Dealing with growing complexity and its implications

“I remember I was looking for something here in Amazon last autumn—I didn’t buy anything that time—but I was left with an impression that these pages, eh, I mean I get the same impression now, too, as I look at this. You know, there’re too many things, confusing abundance, there’re awfully many things on the screen as it is, and I’m left to figure out where to start. I’m now trying to decide which way to take...” – P6

New features can offer new, innovative ways to shift through the mass and find items of interest for the intrepid user who starts to use them. At the same time, however, the wealth of features offered increases the inherent complexity of the information
environment. The price for increased complexity ranges from features being ignored and users sticking to what they know to disorientation and users feeling themselves outsiders in an environment meant only for the initiated all the way to users moving to other environments where they feel more at home. We witnessed all these in this study. However, we also witnessed how a feature that was new to the participant was tried and became instantly an important part of the item-evaluating process.

The participants dealt with the complexity and overwhelming number of features by using a limited number of features and ignoring the majority. P3: “I was thinking that quite often the things below here, I never really pay much attention to them, [I look at] just a few pages [screens] but it continues down there, and these I never really look at. ...All the stuff below was wasted. I didn’t really notice it well enough to pay attention to it and perhaps I didn’t even realize that some stuff was left unseen and how much stuff there was after all.” He felt overwhelmed by an ever-increasing number of features: “...this is of course very difficult because there are always new ones coming in.”

When the interface did not offer what was expected, participants felt disoriented. P4 decided to use Amazon.co.uk in Task 1, but after seeing the home page he quickly moved to Amazon.com: “Ok, this looks more familiar. This is what I was expecting. This is a bookstore. Here there’re books on the cover, right here on the home page. The other one had some electric devices, albums and all, everything under the sun.”

Likewise, P2 felt disoriented because he had a brand new computer with which he had not visited Amazon, and so he was not automatically signed in: “Interesting, this is... Well... How did this happen?” He could not find personalized recommendations from which he usually started his item-finding process. After understanding what had happened he still had difficulties figuring out how to sign in: “This is not mine, that’s for sure. There’s the sign in [link]! Why is it so darn small?”

The same disorientation could be seen in how some participants remembered some features but did not remember how to access them. For instance, P1 spent a long time trying to figure out how to get to “Search Inside” on the item page: “I am trying to find that feature that some books in Amazon have—or is it that all the books have it—or is it only in dot-com—where you see what the book really looks like [opens a bigger picture of the book’s cover while trying to find the ‘Search Inside’ on a book that does not have it] and not only the cover.” Finally he managed to access the
feature from the middle of the item page where the feature is called “Inside This Book” (Figure 10) for one book. He never found out how to tell if a book had the “Search Inside” feature on the list page or on top of the item page.

In the same way, P6 spent some time trying to locate the “Customers Who…” recommendation that he wanted to use strategically: P6: “Where is it? [Scrolls the item page and finally locates ‘Customers Who…’] This is the list I was looking for.”

Most participants did not visit Amazon that often, and consequently the interface was in a sense re-learned at every visit. Most felt that their level of experience in Amazon was inadequate to what was required to being fluent in it. P3: “So maybe, apparently the consumers who use this more can find all these.” P4 concurred: “I haven’t been here for about three months and in a way many of these things that there are now, I rediscover them again, like oh, yes, there was something like this here.” Even P6, whose work was related to computer systems, felt like not having adequate skills: “…the interface could do with some development at least as far as users like me who don’t use it all the time are concerned. Occasionally you have to stop and think like, wait a moment, how did you go there and so on.” As mentioned, P1 also felt that he might perceive the environment as more social if he visited the site more often and used the features more.

- Understand every aspect of your digital camera, from menu settings to the technology that makes it work.
- Refresh your compositional skills and hone your technique to take advantage of the speed and flexibility of digital technology.
- Get to grips with exciting modern camera features such as predictive autofocus, multi-pattern metering and programmed exposure.
- Unleash your creativity with the in-depth coverage of traditional photographic skills, including shutter speed, aperture and other camera controls.
- Set up your digital darkroom using the practical guide to computers and image-manipulation software.

With its combination of clarity and detail, photographers of all skill levels will quickly discover The Book of Digital Photography to be an indespensible companion.

See all Reviews

Inside This Book (Learn More)
Browse Sample Pages
Front Cover | Copyright | Table of Contents | Excerpt | Index
Search inside this book:

Customers viewing this page may be interested in these Sponsored Links (what is this?)

Compare Book Prices
www.BookShire.co.uk • The Digital Photography Book Find the Lowest Price.
Digital Photography Course
www.indean.co.uk • Take a digital photography course for beginners – sign up in a flash.

Figure 10. Access to “Search Inside”—called “Inside This Book” here—in the middle of the item page.

All in all, only P5 had no qualms about his abilities to deal with the environment. He had a clear process with which to find items of interest, but his approach also ignored most features and focused on what he knew from before. His method of dealing with the complexity was, therefore, also that of ignoring the
extraneous and focusing on what he knew. That meant that when he accidentally closed a tab, he re-found the book with a keyword search instead of using “Your Recent History” at the bottom of the page. He did not know about that feature because he never went to that part of page, as none of the features he used were there.

Furthermore, P5 also made some wrong moves that produced unexpected results from which he was however quick to recover. For instance, while trying to access “Search Inside,” which the book in question did not have, he accessed “Publisher: learn how customers can search inside this book” that was located almost on the same place and in the same style as “Search Inside” would have been (Figure 11). The same location and style, crowned with the actual words that he was looking for, combined to cheat him.

![Digital Photographer’s Handbook (Hardcover)](image)

**Figure 11.** The “Publisher: learn how customers can search inside this book” feature is located at almost the same position as “Search Inside” would be and looks very similar in style.

Other participants also occasionally had difficulties differentiating features from one another. P2 used the “See larger image and other views” feature (instead of “Search Inside”) that was likewise located in the same area where “Search Inside” would have been without being able to tell that the feature he had accessed was not the one he had planned to access (Figure 12): “Good. How does this work now? Where’s the Table of Contents? What?” The feature was similar enough to “Search Inside” not to allow him to realize that he was using another feature that did not offer Table of Contents.

In addition, occasionally a participant read “Reviews” after stating that he is going to read “Customer reviews.” Moreover, the participants did not appear to differentiate between different “Customers Who...” features, either. They were
looking for “Customers Who”…lists and used them with total disregard as to whether they were “Customers Who Bought…” or “Customers Who Viewed…” lists. It was questionable if the difference was relevant to them in the first place.

Not surprisingly, the feature-rich and ever-changing environment offered unwanted adventures. P4: “I don’t know, but that’s what it is. You go there and browse and try to find some way to move ahead. It’s simply based on trial and error, nothing else.” P4 described the feeling that resulted from not being familiar with the environment: “…perhaps it also affects it that from time to time you feel yourself an outsider because you don’t know these things so well. You feel like a hang-around in some community where the initiated go and do all sorts of things…”

Figure 12. “See larger image and other views” feature opened.

However, it was the same P4, who had used Amazon the least of the participants, that showed how flexibly and quickly the strategies are formulated and adapted to the current situation. When he accidentally noticed that moving the mouse cursor over the cover-picture of a book with “Search Inside” opened the “Inside This Book” layer (Figure 13), he quickly integrated using it to his item-information browsing approach. In fact, he was the only one to use keywords to search inside
books, as the others typically used “Search Inside” in a more traditional way with which they were familiar, that is, by opening the feature and looking at the offered pages one by one. P4: “But this was good, this ‘Search Inside.’ I didn’t notice this the first time, and I haven’t noticed this before, either.” After discovering the feature, he used it more than any other participant, as shown in Table 10 (see page 65).

Consequently, designers of complex information environments have to deal with the fact that users tend to re-learn the environments each time they visit them. Expecting a certain level of environment-expertise even from a returning visitor can lead to feelings of being outsider or to users switching to another site, as P4 switched from Amazon.co.uk to Amazon.com (keeping in mind that the task setting limited him to Amazon sites). The inherent complexity of the environment brought on partially by the number of features and, especially in Amazon’s case, the fast development and constant updating typical to Web 2.0 services, a state that Tim O’Reilly [2005] called “the perpetual beta,” makes it also hard for less frequent users to stay on top of the features in the information environment.

If we look at the features in the item page that the participants used and did not use, we notice interesting factors. All used pretty much the same set of features: “Search Inside,” “Customer Reviews,” “Reviews,” “Customers Who…,” “Perfect Partner,” and “Product Details.” Furthermore, with only minor exceptions, such as P4, P5 and P6 using the author’s name on the item page to access the author’s other books,
P3 once checking the other reviews of one reviewer, and P6 once checking “Explore similar items,” they did not use any other features offered, such as “Amapedia,” “Listmania!,” “Customer Discussions,” “Look for similar items by subject,” or “Your Recent History.” Thus, the participants by and large used the same set of features and ignored the same set of features.

This rather stable set of features appeared to provide the participants with what they needed. However, using this small set of features could also signify other things in addition to need-based selection. One central question is if growing complexity at some point leads to blocking out some features in order to concentrate on the core features that one is familiar with. If so, the reliance on a relatively small set of features could be seen as a coping mechanism in a complex information environment.

The features that the participants used have been part of Amazon for a long time, and so the participants typically had met them first when the item page was not as fully loaded with features as it is today. Thus, familiarity could have had an effect on these features having been selected. Even if some participants could not remember how exactly to access the features, they were consciously looking for them. Were the features to be attended selected based on their usefulness to the task at hand or was familiarity used as a criterion in that the participants ignored unfamiliar features and used familiar features in order to reduce complexity? Observation showed that unfamiliar features were typically not studied during the use. The participants simply ignored them as noise. This would argue for complexity reduction to be at least a partial reason.

By and large, the features that the participants used appeared towards the top of the page while the features not used appeared towards the bottom of the page. The participants were not necessarily even aware of the unused features that appeared at the bottom beyond the general knowledge that there still was something there. Again, were the features at the bottom ignored because they were not considered useful or were they not attended because the page was too complex, and attending only to the topmost features was a way of reducing complexity? Our data implies strongly that the ignored features were not consciously seen as unnecessary but that they were ignored in order to reduce complexity as they simply were never attended to.

Nevertheless, we also saw that features that communicate their meaning effectively, are positioned properly, and are truly useful can be adapted to the item-finding process very quickly and effectively.
This study cannot conclusively answer the question of why the participants used the set of features that they used and did not use the others. However, we conjecture that familiarity with the features and their position towards the top of the item page played a role in the selection. Furthermore, we conjecture that new or otherwise unfamiliar features can be blocked out of attention as a method of reducing complexity, especially if located in the areas to which users do not typically attend.

Overall, we can say with certainty that more is not inherently better. If the number of features grows very big, some features begin to be ignored. If each user finds the features that are meaningful to him or her personally and ignore the rest, forming in a sense mentally their own page, there is no problem. However, the participants by and large used the same features and ignored the same features, and so the process on the item page was not individual by nature. Instead, there were numerous features that simply were not useful to any participant. Complexity had been increased without benefiting any of the participants.

It is also clear from the data that features need to be clearly distinguishable. Users must know which feature they are accessing and using. Not only does this reduce complexity of the environment by keeping users on the map, so to say, but it also reduces feelings of confusion and frustration. Amazon’s new naming policy of “Reviews,” discussed in Section 4.3.3., is likely to lead to improved understanding of the environment and the meaning of the feature. On the other hand, calling “Search Inside” “Inside This Book” on mid-page (see Figure 10 on page 82) does not help users to form a proper mental model.

In addition, making features clearly distinguishable can reduce the number of features. For instance, the participants did not differentiate between “Customers Who Bought…” and “Customers Who Viewed…” recommendations. If the difference is not meaningful, then the two features should be combined into one.

However, whether users differentiate between such similar features in actual use is something that can only be found out by observing the actual use, as interviews and focus groups easily produce false positives in such cases [BusinessWeek, 2005; Barlow-Busch, 2007]. People easily find distinctions when intellectually thinking about something while in actual practice they do not make the same distinction. Many features might seem interesting and useful to users if they are asked about them but only observing the actual use can tell if the features have become part of the actual process.
Complex information environments are inherently complex because they need to support many strategies and approaches, and so the number of features in them is inherently high. As excess of features is detrimental to user performance, we need to be on guard against incorporating features that do not serve users.

4.7 Study limitations

Researching complex information environments and measuring the impact of IT artifacts, such as recommender systems, with any research methodology is challenging [Minocha et al. 2006; Kumar and Benbasat, 2006]. Furthermore, some limitations are inherent to any study method as no study method can describe reality as it is, considering all its facets. Moreover, in any study, there is an inherent tension between internal validity and external validity [MacKenzie, 2007]. The more we attempt to control sources of variation, the less the study reflects real-world reality and vice versa [MacKenzie, 2007].

As we used applied ethnography, we made no attempt to control sources of variation, and attempted to have as naturalistic setting as possible for the observation-interview sessions. Consequently, we purposely had not internal validity and the question of limitations in our study centers on the question of transferability of the results [Tremaine, 2007].

As discussed, our participants were genuine users, not students acting as customers, and they had a genuine motivation to find items in which they were really interested. Consequently, the limitations in our study arise from three factors. One, we had only six participants and all were male. We explicitly decided to recruit only males as the small numbers made it impossible for us to reliably compare the behavior of the two genders. Two, we used books as item. Three, we used tasks to seed activity for observation.

While we can try to excuse the low number of participants with explanations of how much work went into each session and consequent work on material, the fact remains that with six participants, we can only formulate ideas and observations that need further confirmation. However, without this work to build research questions and hypotheses for testing, the subsequent work would be challenging, as we would most likely end up asking wrong questions and test many wrong hypotheses. As discussed, the research on the use of complex information environments has focused on parts in
isolation rather than on the whole, and we felt that insight into the whole was sorely missed.

Moreover, while our participants were all genuine users, they only represent a small group of male Amazon users in Finland. Thus, we have to be cautious about generalizing the results too far.

The second limitation was that we used books as items. Different items may be found and selected based on different criteria [Ozakca and Lim, 2006; Kumar and Benbasat, 2006]. In fact, the observations here can safely be extended only to non-fiction books, as the criteria for selecting fiction and non-fiction books tends to differ [Cole, 2003 p. 20].

Furthermore, online environment allows us to experience products to different degrees. We cannot hold a book in hand or try the cloths for size but we can preview music or movies pretty much the way we would experience the final product. Consequently, each item has its own domain and related search and selecting criteria.

The third limitation was the use of tasks. Because of it, the situations that we observed were not 100% naturalistic [Minocha et al., 2006] even though they were close to it otherwise. The central question here is how much the task setting changed the participants’ behavior.

According to the participants, the requirement of Task 1 that they were to find a book they had not decided to buy beforehand confused them somewhat at the offset. P1 said that in many cases, he already knew quite well which books he was going to buy, but that when he needed a book for a particular topic, he did use Amazon in this way: “That is how I search for books in Amazon with keywords…” Both P2 and P3 said that they typically started with personalized recommendations, as they did here. P5 normally used his favorite authors as a starting point, as did P2 in addition to recommendations. P2, P4, and P6 said that they typically already knew something about the book they came to buy, which made this way of searching somewhat different from normal. However, all said that after getting on with the task they ended up using Amazon pretty much the way they normally do.

Thus, the main difference for all participants was that they usually had some kind of a starting point in mind when arriving at Amazon, and now they did not have it. P6’s approach was affected the most because he spent a lot of time trying to locate an item of interest by using categories, an approach he would not typically use. Finally, however, he also reverted to his normal pattern of searching.
Consequently, we feel confident that the task setting did not change the observed behavior to any large extent and that we can use the material for drawing conclusions about actual use.
5 Conclusions

“Now that I’m looking around and adventuring around the wonderful world of Amazon, they’ve got all kinds of cross-searching possibilities and god knows what else, and they do feel useful.” – P4

Research on complex information environments has largely focused on different aspects of such environments, for instance recommender algorithms, rather than on understanding and studying such environments as organic wholes. Furthermore, much research has been conducted with surveys or using students as consumer surrogates. While important advances have been made, we have only begun the process of understanding how actual users approach complex information environments as wholes and how the pieces are actually used in the context of the whole, and not in isolation prevalent in typical laboratory experiments.

We conducted an applied ethnography study, observation with verbal protocol combined with semi-structured interviews, with six participants, all genuine users, to understand how users work in and deal with complex information environments, and what kind of item-finding strategies have emerged from that interaction. We used as environment Amazon online store as Amazon has consistently been a forerunner in the industry and early adapter of various approaches to assisting users in item-finding and selecting. Consequently, our data shows how genuine users actually find items of interest in today’s complex information environments.

We were able to show the growing importance of recommender systems to item-finding, and how they are being used strategically, for instance in conjunction with keyword searches, in addition to being used opportunistically. Recommenders affect the process both by directing users to interesting items and by helping them to select items. While recommenders have grown in importance and made discovery easier, searching also has a solid place in the equation in the foreseeable future. On the other hand, categories are no longer being used to locate items but rather to refine searches or recommendations.

Furthermore, we saw how customer reviews play today an important role in decision making. While customer reviews tend to have a more pronounced negative impact, they also play an important part in the positive decisions. However, as the
number of reviews keeps growing and information overload is already beginning to affect the use of customer reviews, we also recognize the need to personalize customer review experience for each user.

As users already appear to scan for keywords when selecting customer reviews for reading, we hypothesize that finding such keywords and presenting them to users, for instance as a keyword cloud (cf. tag cloud), would help users in finding the reviews that are relevant to them.

We witnessed how the participants divided into two groups as far as item-finding strategy was concerned. Some searched for an item that satisfied their need, thus satisficing to a good enough item, while others strove to find the best item, comparing items until they felt that they had found the best item. Interestingly, the participants using the Best item strategy ended up buying more books. We do not know if the strategy selection depends on the item in question, the importance of the item to the user, or if it is an inherent user characteristic, and thus further research is necessary. Based on our data, we hypothesize that users have a tendency for one type of strategy but that the item and its importance to the user can also affect the choice of strategy.

We repeatedly observed that the participants did not limit themselves to Amazon but used the whole Internet as an information source. If Amazon did not offer enough information for the participants to make up their minds, they “googled” the items and authors for further information.

In fact, the constant use of Google appeared to have created expectations that bled over to other environments and affected the user behavior in them. Based on our data, we hypothesize that users have ready use patterns that they adapt to the environment at hand. These patterns carry certain expectations that color the user behavior, and not fulfilling these expectations can lead to users looking for other environments that confirm to their expectations and ready use patterns.

The perception of social presence in complex information environments was not uniform. Some participants perceived Amazon as inherently social and their behavior was affected by the perceived presence of the community while others felt that they were “alone” in the environment and that the existing social texture, such as recommendations and customer reviews, did not make it any more social. Furthermore, some participants had clearly negative attitude towards making
information environments overtly social. Both views should be respected, and transparency combined with means for opting out is recommended.

The propensity to see information environments as social might be connected to personality, and thus making the social texture more explicit than it already is in Amazon might have a negligible impact on the users who do not already perceive the environment as social. This needs to be studied further as social presence is currently an area of keen research interest in e-commerce.

Interestingly, the propensity to contribute appears to be linked to the perceived social presence. The participants that perceived other users as present had more positive views towards contributing than did those who felt that they were “alone” in Amazon. In addition, perceiving an environment as social affected the behavior of at least one participant who consciously was aware of considering his actions in relation to the community-at-large. Consequently, we hypothesize that we can increase both the quality and quantity of the user contributions by increasing the perceived social presence. Furthermore, we hypothesize that the more users perceive other users as being socially present in the environment, the more they take them into consideration in their behavior.

Information environments have already reached the level of complexity where users are mentally blocking out features to reduce the complexity to a comfortable level. The participants used a constant set of features, ignoring many other features entirely. Furthermore, all used by and large the same features and ignored the same features. Consequently, we need to understand further why certain features are used, and use this knowledge to keep the information environments lean and mean and the growth of complexity at bay.

Our study and results underline the importance of studying the whole in actual use context in addition to studying the parts in isolation. Studying the parts improves the whole and studying the whole improves the parts. Neither can be ignored if we want to build information environments that truly serve their users.


Acknowledgements

This work was supported by the Finnish Funding Agency for Technology and Innovation (project 40279/05) and by NORDUNet (project PriMa, Privacy in the Making). I am also indebted to the University of Tampere for various working opportunities that have been invaluable to this work.
References


[Minocha et al. 2006] Shailey Minocha, Marian Petre, Ekaterini Tzanidou, Geke van Dijk, Dave Roberts, Nick Gassman, Nicola Millard, Barry Day, and David Travis, Evaluating E-Commerce Environments: Approaches To Cross-


[Rashid et al., 2002] Al Mamunur Rashid, Istvan Albert, Dan Cosley, Shyong K. Lam, Sean M. McNee, Joseph A. Konstan, and John Riedl, Getting to Know You:


