Consumer Decision Patterns Through Eye Gaze Analysis

Sylvain Castagnos
EPFL - HCI Group
IC IIF Station 14
CH-1015 Lausanne
sylvain.castagnos@epfl.ch

Pearl Pu
EPFL - HCI Group
IC IIF Station 14
CH-1015 Lausanne
pearl.pu@epfl.ch

ABSTRACT

Eye tracking technology is a powerful tool used in the HCI field to measure the attention of users. This has of course numerous favorable financial consequences, since it helps developers of e-commerce websites to redesign their interfaces and layouts in accordance with users' habits and expectations. Most of the time, gaze data are consequently used to make an application more user-centered. In this paper, we propose a different approach consisting of analyzing consumers' behaviors and decision aid agents' utility when introduced to an online perfume shop. The literature about purchase decision making is enriched by many studies focusing on the interactions between customers and sellers. However, few studies have been performed to examine consumers' online strategies through interactions with intelligent systems. The understanding of the sub-processes involved in the online purchase decision making process will help us to modify and improve the conception of decision aid agents. The results of our study show that each entity in common e-commerce websites has a specific role. In particular, a recommender system helps users to build the basket set and facilitates exploration. This tool is as influential as the description of products, or the list view of available items in a catalog.

Author Keywords

Eye tracking system, e-commerce, decision theory, recommender systems, usage patterns

ACM Classification Keywords

H.1.2 Models and Principles: User/Machine Systems – Human Information Processing; H.5.2 Information Interfaces and Presentation: User Interfaces – Evaluation/Methodology

INTRODUCTION

We examined the usage and impact of two decision aid agents and two information entities on consumers as they search for an ideal product and construct the consideration set in common e-commerce websites via the use of an eye tracking system. We use both eye gaze data and action logs to understand how each of these four elements helps users in selecting products to put in their basket of final selections. The two decision aids include the more traditional multi-criteria filtering tool (MCF) and a recommender system (RS). The two information entities comprise the lexicographically ordered list view (LV) and a product description box (D). Our experiment shows that each tool has a distinct role. Users mainly employ the MCF tool, combined with the list view LV, to localize a sub-set of valuable alternatives for consideration. They identify and fix a set of criteria using MCF through several search cycles, and they consult the description box for additional information about products. Finally, they use the recommender RS to help them build the basket set. RS seems to provide additional and efficient exploration opportunities and appears to increase users' confidence in the purchase decision. The cumulative usage of the recommender is shown to be as frequent as the usage of the list view or the description box, and up to two times the usage of the MCF tool.

The rest of the paper is organized as follows. We first review the related work in terms of usage of eye tracking technology in the HCI field. We will explain how we can benefit from eye gaze data to analyze user behaviors when exposed to a realistic e-commerce application, and to better understand the sub-processes involved in purchase decision making. We then describe the experiment setup in detail and present the participants' background information. The paper continues by reporting the results of the study, and discussing them. Finally, we present our conclusions and ideas for future work.

RELATED WORK

Eye tracking is a technique whereby an individual's eye movements are measured to identify both where a person is looking at any given time and the sequence in which their eyes are shifting from one location to another [12]. It helps HCI researchers understand the factors that may impact upon the usability of system interfaces. Eye-movement recordings provide a dynamic trace of where a person's attention is being directed in relation to a visual display. Measuring other aspects of eye movements, such as fixations (moments when the eyes are relatively stationary, taking in or "encoding" information), can also reveal the amount of processing being applied to objects at the point-of-regard. In practice, the process of inferring useful information from eye-movement recordings involves the HCI researcher defining "areas of interest" over certain parts of a display or interface under eval-

Submitted for review to International Workshop on Eye Gaze in Intelligent Human Machine Interaction (IUI 2010).

uation, and analyzing the eye movements which fall within such areas. Thus, the visibility, meaningfulness and placement of specific interface elements can be objectively evaluated and the resulting findings can be used to improve the design of the interface [7]. For example, the Nielsen Norman group has recently used eye tracking technology to evaluate web usability. As a result, a set of guidelines for writing for the web has been generated, including the concept that the first two paragraphs must state the most important information since users will not read the text thoroughly in a word-by-word manner [11].

Mainstream psychological research has also benefited from studying eye movements as they can provide an insight into problem solving, reasoning, mental imagery, and search strategies [2, 8, 14, 15]. The commercial sector is also showing increased interest in the use of eye tracking technology in areas such as market research, for example, to determine what advertisement designs attract the greatest attention [9] and to determine if Internet users look at banner advertising on websites [1].

In this paper, we propose an original way of applying eye tracking technology to study the human-computer interactions leading to a decision within an e-commerce website, as a continuation of our previous works [3, 4]. This in-depth study will help us design new filtering strategies related to attention economy [6] in recommender systems. According to previous research, there are six fundamental stages within the purchase decision making process [10]. First, users become aware of a new need. Then, they have to determine from whom to purchase the desired item. Everything customers experience on a website feeds into the building of a rapport between the buyer and the seller. Users search elements that promote trust, such as the ease of navigation or the relevance of answers and recommendations. In the meantime, they evaluate product alternatives to make a choice. At this stage, called *product brokering*, interactions help recommender systems to iteratively characterize their needs and present options. Stages 4 and 5 consist in negotiation and purchasing. The seller must provide security and confidence in order to close the sale. Finally, users' satisfaction in relation to the overall buying experience can be measured in a sixth stage if post-purchase product service is involved.

The purpose of this work is to analyze the usage patterns during the product brokering stage. In particular, we aim at understanding the role of tools provided in common e-commerce websites.

The following section describes the setup of the experiment.

EXPERIMENT

The Material

The experiment consisted of an in-depth real-user evaluation with an eye tracker on a perfume e-commerce website. The eye tracker used in our experiment was a `Tobii 1750`. This device consists of a computer screen with a camera installed on the top. The system's software is then capable of capturing the user's point of gaze. Except for a short calibration phase, the setup allows users to look at the screen in a natural way without the need for a head mount.

The perfume website was setup specifically for this user-study, reproducing the layout of a real e-commerce website. The database consisted of more than 3,500 perfumes, crawled from popular perfume websites, and contained all popular brands and perfumes that are available in regular perfume shops. We chose the perfume domain, as it is a slightly above-norm field in terms of complexity. A more common domain would have been less engaging, possibly resulting in some "shortcut" behaviors. On the contrary, perfumes are complex public taste products with a lot of characteristics involving a non trivial decision (brand, quantity, price, category, fragrance, design of the bottle, fit to a specific personality). Moreover, most people are generally considered non-experts and have stable preferences with respect to perfumes. Such constraints mean that the recommendations and the overall quality of an online perfume website must be very persuasive – since users cannot smell the perfumes and form their own opinion about new perfumes – and thus constitute a good framework for study.

The design of our website has been deliberately chosen to reproduce a generic recommender system enabled e-commerce website. In this case, two decision aid agents are usually offered: a multi-criteria filtering tool (MCF) and a recommender agent (RS). The latter is most likely to function when users look up the details of a product. Thus, we split the site into two main windows, as shown in Figures 1-a and 1-b. The first window – called "search page" – was divided into two parts. At the top of the page, a multi-criteria search tool (MCF) was conceived, that included brands, price ranges, quantity ranges and types of perfume (*eau de parfum*, *eau de toilette*, aftershave). Below this, a double column, lexicographically ordered item-list of perfumes was available. This part displayed the perfumes respecting the selected criteria of the upper multi-criteria search tool. In this double choice-list, each perfume was laid out with a picture of the bottle, its exact name, the brand, price and quantity. Additionally, the first two lines of its description were shown.

In addition to the MCF tool, we proposed a classical re-ordering tool allowing users to sort the list of results by brand, price (low to high, or high to low), and popularity. The search page also included a possibility to set the number of results displayed on one page and the currency. By default, the results were displayed in US dollars, sorted by popularity, sixteen at a time.

The second main window that users encountered was for the detail of any perfume (hereafter "detail page"). In complement to presenting the same information as in the list view (see above), specific data is provided here including: a full description, a big-sized picture, a best-selling rate, average user ratings, gender, source website, and the possibility to rate. The page had an "Add to shopping cart" button. At the same time, on the right of this detail, a column of recommendations was proposed. These were pro-

Figure 1. Snapshots of the main interfaces (inc. AOIs): the search page on the left (a), the detail page in the middle (b) and an example of heat map on the right (c)

posed in five classified boxes, all displaying their classification label. The seven possible labels were:"More popular and cheaper", "More popular but more expensive", "Same brand and cheaper", "Same brand but more expensive", "Just as popular and cheaper", "Same price range and just as popular", and "People who like this also like". Although seven types of recommendations were available, we chose to only display five at any moment in random order to reduce users' habituation to screen position. Each box contained up to six items, which were horizontally scrollable without reloading the page. The whole set of recommendations were displayed under the label "you may also like", in addition to each recommendation category's label.

Since it has been determined that the quality of recommendations has an impact on click-through, the recommendations were generated from Editorial Picked Critiques (EPC), adapted from the preference based critiquing method with users' needs for popularity and editorial information [5]. The EPC algorithm has been proven to be as accurate but 2.42 times more preferred than general critique-based recommenders [13]. We classified recommendations in the categories mentioned above. The "People who like this also like" group relied on a standard collaborative filtering algorithm. Other recommendation categories were related to products' features. They were computed by adjusting values of some variables and keeping other variables constant. Our recommender system (RS) provided categories with different levels of diversity as explained in [4].

In the next subsection, we introduce the indicators of usage that we employed.

Indicators of Usage

In order to measure the concrete impact of recommenders on users' behavior, we chose to combine two kinds of data: eye gaze and action logs.

We applied the eye tracking technology to catch users' usage patterns, reveal the exploration strategy, and better understand decision making sub-processes. Indeed, what a person is looking at is assumed to indicate his/her cognitive processes [8]. The practice of inferring useful information from eye movement recordings involves defining Areas Of Interest (AOIs) over certain parts of a display or interface under evaluation and analyzing the eye movements which fall within such areas. We used two metrics to reach this last goal: the *fixations* and the *reading heat maps* [3]. Fixations designate moments where a user looks a particular area for a fair amount of time. Reading heat maps provide an overall view of activity on a page (see Figure 1-c above). To create the heat map, data from each person looking at the page is combined to show what percentages of people viewed each part of the page. The colors reference the proportion of participants whose eyes fixated on certain elements of the page. The red areas indicate where the larger percentages of users looked most.

Apart from the usual gaze plots and heat maps which can be collected with an eye tracker, we decided to rely on a large palette of *Areas Of Interest* (hereafter AOIs). Exporting the time spent on each AOI is important as it is an objective data about users' actions on the website. We defined 27 generic types of AOIs. First we created four main AOIs: one for the *multi-criteria* search tool, one for the returned list of choices (*list-view*), one for the detailed information about the desired perfume on the detail page, which we labelled *description*, and finally one for the whole *recommender* element. Then we defined many other smaller AOIs which include the recommendation categories, the photo, title, price, quantity & category, rating, descriptive comment and rating for the detail of any perfume. Likewise we defined AOIs in the multi-criteria search for each category brand, price, quantity, category. A selection of all these are detailed in Figure 1-a and 1-b.

In addition to the eye gaze data, we collected implicit indicators under the form of a collection of access logs stored on the server in a CLF format.[1] These files notably contain information about uniform resource locators (URLs) of files corresponding to the request of users, and the time that the server finished processing this request. Such data reveals the time that users spent on each page, and the set of consulted pages. We also used a javascript to instantly track all users' clicks and collect them in the access log files. The clicks highlight deliberate interactions between the users and the

[1]http://www.w3c.org/Daemon/User/Config/Logging.html

system, or the users' explicit interests (adoption).

Experiment Procedure

The study was design as an in-depth one hour lab-study. At all times, participants could ask questions of the available assistant conducting the study. The general online evaluation procedure consisted of the following steps:

Step 1. The participant is welcomed by the assistant. He is briefly introduced to the topic of the experiment. He is informed that the perfume e-commerce website he will test contains over 3,500 most commonly used and sold perfumes in the world. He is also told about the incentive for completing the study.

Step 2. The user is asked a detailed set of questions, to collect his background information (age, sex, etc.). The responses to this step are detailed in the following subsection of the paper.

Step 3. Before starting the experiment, the eye tracker is calibrated to the user's eyesight. The experiment can now start and the tracking session is launched by the assistant, who encourages the user to explore the system before fully launching into the first task.

Step 4. The user then has two separate tasks to complete. One goal is to select up to three perfumes that he has never heard of or used before, but that he would be prepared to buy for himself. He is asked to put them in the basket, and informed that he may select more than three and delete some at the end. In the rest of this paper, this recording will be called "Session S" (Self). The other goal consists of searching for one perfume he would like to offer to someone, preferably of the opposite sex, in order to reduce potential bias of product habituation. This will be called "Session G" (Gift). These two sessions are recorded in two different files. In order to reduce another bias linked to fatigue, we alternate the order of sessions. Half of the users complete the task of Session S before Session G, the others start with Session G and end with Session S.

Step 5. To conclude the study, fourteen preference questions are asked in order to explicitly assess user's overall perception of the system after the experiment. This allows us to match explicit and implicit data to confirm our hypotheses.

The preference questions in this study are statements to which a user can indicate his level of agreement on a five-point Likert scale, ranging from -2 to $+2$, where -2 means "strongly disagree", $+2$ is "strongly agree", and 0 is neutral. The post-stage questions are listed in Table 1. The questions were asked in random order to eliminate ordering bias.

Table 1. Post-Stage Assessment Questionnaire

ID	Statement
P1.	The items under "You may also like" are attractive.
P2.	The items under "You may also like" are educational.
P3.	The items under "You may also like" appeared to be a good deal.
P4.	The items under "You may also like" appeared to be marketing material.
P5.	The items under "You may also like" influenced my selection.
P6.	The items under "You may also like" will influence my future selection.
P7.	The names of the categories are useful and adequate.
P8.	I am satisfied with the overall quality of the interface.
P9.	I found the interface easy to use.
P10.	I would buy the perfumes recommended to me, given the opportunity.
P11.	If this were a real website, I would use it in the future to find perfumes.
P12.	I believe that the recommender algorithm is efficient.
P13.	The recommended perfumes were diverse.
P14.	The recommended perfumes were novel.

Participants' Background

The user study was carried out over a period of three weeks. Immediate incentives, chocolate or wine, were offered directly after the study. More importantly, users who had completed the study took part in a draw for a CHF 100.- voucher to buy one of the perfumes he/she had added to the basket. By proposing this high value incentive, we ensured that users behaved candidly throughout the selection process. A total of eighteen volunteers were recruited as participants. They were from three different continents, with different professions (student, worker, Ph.D. student) and educational backgrounds (high school, graduate school). Table 2 shows some of their demographic characteristics.

All of the participants expressed that *fragrance* was an important feature, necessary in order to describe a perfume as shown in Graph 2. Other important aspects include price, brand, quantity and design. Of the background questions, two assessed users' experience levels with regard to Internet (online media, information retrieval, Internet communication, online communities, online entertainment and e-commerce) and online shopping. Both revealed that all users' had strong web experience, although online shopping experience remained limited to classical items such as books, music, travel and electronic items, much less so for food, drinks, groceries or clothes.

The second part of the background questions surveyed users' predisposition towards perfumes. Six of the eighteen participants considered themselves to be knowledgeable about perfumes. In the rest of the paper, we will call them "perfume experts". Six participants said they bought perfumes about once a year, nine a few times a year and one nearly monthly. When questioned about how they discovered new perfumes, 61% said that they preferred to just test perfumes alone in a shop. Other answers included relying on friends' advice or

Table 2. Demographic characteristics of participants

Gender	Female 9 (50%)	Male 9 (50%)	Total 18 (100%)
Age	17-25 4 (22%)	26-40 13 (72%)	41-55 1 (6%)
Education	High school, Graduate school		
Profession	Student, Ph.D. student, Worker		
Nationality	American, Chinese, French, German, Russian, Serbian, Swiss		
Online Shopping Experience	travel	94% a few times a year	
	books	56% a few times a year	
	music	22% monthly	
	electronics	39% a few times a year	
	food	88% never	
	groceries	28% a few times a year	

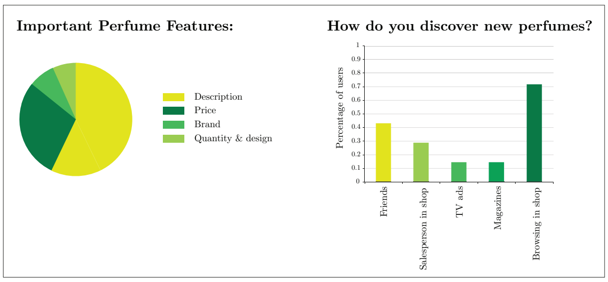

Figure 2. Users' knowledge about perfumes

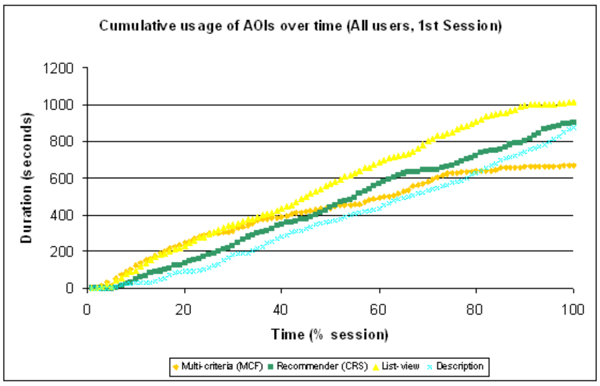

Figure 3. Evolution of main AOIs over time during the first session.

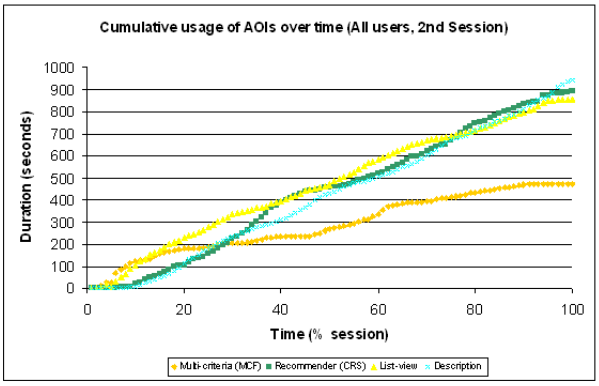

Figure 4. Evolution of main AOIs over time during the second session.

the seller in the shop, as shown in the second part of Graph 2. Most users also told us that they were prepared to reveal information, such as previously liked and disliked perfumes and price, in order to get recommendations (from a seller). Interestingly, all users were prepared to reveal information about smells that they like, but told us it was difficult to describe smells and hence relied on other aspects.

RESULTS AND DISCUSSION

Averaging 1,450 fixations per user, we recorded 20,306 fixation points throughout the study. We defined 3,648 AOIs, sorted into 261 different web pages and corresponding to 27 variables. These pages were of two sorts as explained above: the search pages and the detail pages. We then computed the total fixation durations and the number of times that each user looks at the different AOIs over time. We paid particular attention to durations for four variables: the multi-criteria box `MCF`, the lexicographic ordered list `LV` (list-view), the description of perfumes `D` and the recommender system `RS`.

Cumulative usages of these four AOIs over time for the overall set of users are made explicit in Figures 3 to 7. Figure 3 displays the fixation durations during the first session when users were introduced to our website for the very first time. During this first session, half of the users were looking for a perfume for a gift (Session G) while the others were asked to select up to three perfumes for themselves (Session S). On this figure, we can see that the global time spent looking at the recommender is equivalent to the time looking at the description box, and 89% of the time is spent looking at the list view. This score is very impressive since the recommender RS was only displaying five recommendations at the same time, while the list view was composed of a large subset of items (between 16 and 100 perfumes on each search page). We also noticed that the MCF usage is 26% lower than the usage of RS. We checked the statistical significance of these results by computing the Pearson correlation coefficients and p-values between duration vectors. Each of these vectors corresponds to the fixation durations of each user on one of the four main AOIs. MCF is certainly correlated to LV ($r = 0.699$, $p = 0.001$), and LV with D ($r = 0.604$, $p = 0.006$). More interestingly, RS strongly correlates with each of the three other AOIs: RS and MCF($r = 0.617$, $p = 0.005$), RS and LV ($r = 0.644$, $p = 0.003$), RS and D ($r = 0.705$, $p = 0.001$). This confirms that our data are statistically significant.

Figure 4 shows the fixation durations during the second session. The results highlight a factor of habituation toward the recommender, which constitutes a positive indicator of satisfaction relative to other available entities. Indeed, within this session, the usage of RS is 88% higher than MCF (p-value $p = 0.076$), 5% higher than LV ($p = 0.33$), and 6% lower than D ($p = 0.002$).

The results for the first and second sessions thus confirm the significant role of the recommender within the exploration

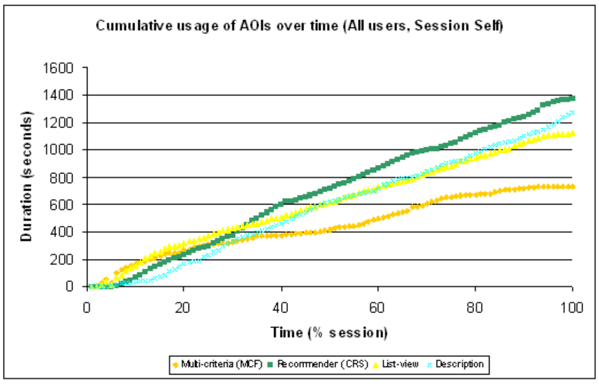

Figure 5. Evolution of main AOIs over time during the session S.

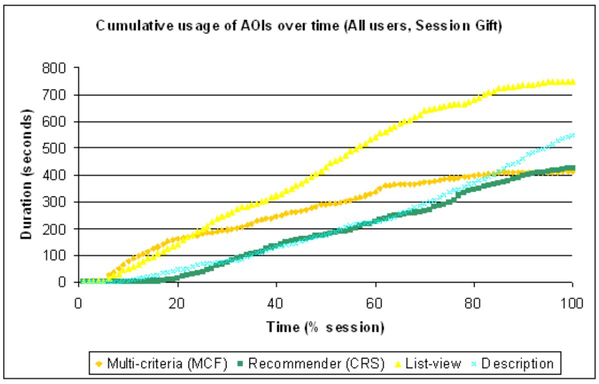

Figure 6. Evolution of main AOIs over time during the session G.

strategy. We also tried to separate the cumulative usages between session S where users had to select up to three perfumes for themselves (see Figure 5) and session G where users had to choose one perfume for a gift (see Figure 6). Users rely much more on the recommender when they are looking for perfumes for themselves. In [3], we have shown that users are more open-minded toward recommendations because of their need for diversity in order to reach a decision. In Figure 5, RS is the entity the most looked at, after the first 30% of the session duration has passed. The recommender is used almost twice as much as the MCF tool. This result is significant with a p-value $p = 0.037$.

However, users' behaviors seem very different when they are searching a perfume for a gift (Session G). Figure 6 shows that the cumulative usages of MCF and RS are equivalent at the end of the session. This result is somewhat significant with a p-value $p = 0.21$. Users spent globally two times longer looking at the list view than at the recommender, but this result is not individually significant ($p = 0.898$).

Finally in Figure 7 we present the cumulative usages of main AOIs for the overall set of users with all sessions taken together. This validates that RS, LV and D have similar importance within the exploration strategy.

We then aim at understanding the role of each of these main

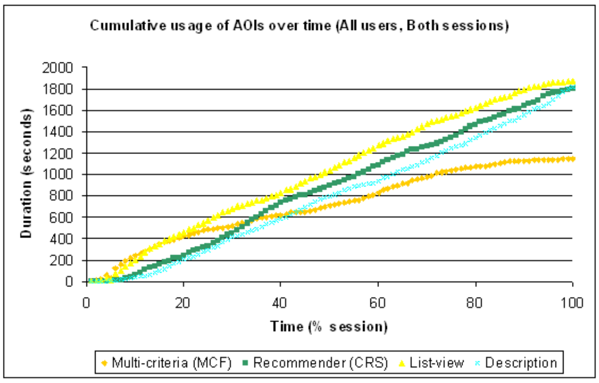

Figure 7. Evolution of main AOIs over time, all sessions taken together.

AOIs within the purchase decision making process. In order to do so, we defined sub-AOIs to obtain more detail about what catches users' attention. According to gaze data, 31% of the time spent looking at the description box D is dedicated to reading the descriptive comment, 23% to looking at the photography, 16% to paying attention to the brand name, 6% to checking the price, 5% to showing interest for the popularity (rating and best selling rate), 3% to reading the category of the perfume, and 2% to looking at the quantity.

Within the MCF tool of the search page, users spent 38% of the time choosing a brand, 25% to selecting a price range, 19% to fixing the quantity range, and 16% to considering the category criteria. Thus, the two main criteria in the MCF tool are only at the third and fourth position in the description box D. This tends to indicate that MCF and D are complementary.

As a continuation of these observations, we examined the impact of the different categories in our recommender systems. We computed the time spent looking at each of these categories individually and for the overall set of users. Figure 8 shows that each category has a significant importance when we cumulate all users' gaze data.

However, there are three dominating categories, and customers seem more attracted to more popular items (30% of fixations). Yet gaze data within the description box D reveals that popularity information only falls into the fifth position. Moreover, the usage of popularity sub-AOIs in the description box D is only correlated with recommender categories that do not deal with popularity, but with brand: "Same brand and cheaper" ($r = 0.585$, $p = 0.008$), "Same brand but more expensive" ($r = 0.469$, $p = 0.043$). The popularity in D is even negatively correlated with the category "More popular but more expensive" ($r = -.226$, $p = 0.352$). However, paying attention to the brand name in the description box is correlated with the usage of popularity-based recommendations (and not correlated with others): "More popular and cheaper" ($r = 0.475$, $p = 0.04$), "Just as popular and cheaper" ($r = 0.5$, $p = 0.029$), "Same price range and just as popular" ($r = 0.457$, $p = 0.049$), "People who like this also like" ($r = 0.606$, $p = 0.006$), and "More popular but more expensive" ($r = 0.512$, $p = 0.025$). These observa-

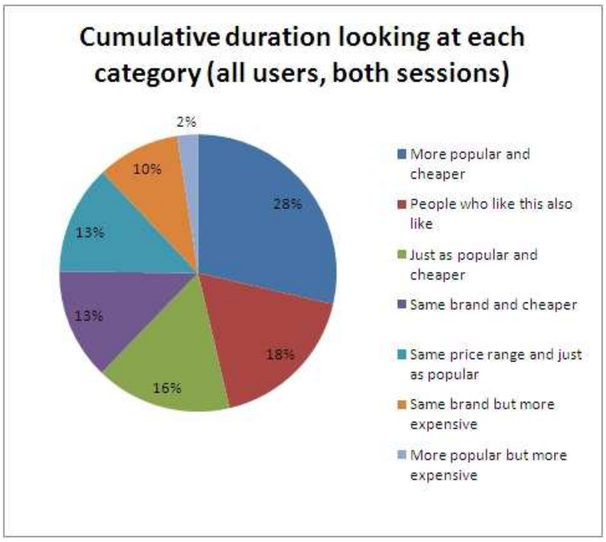

Figure 8. Importance of recommendation categories for the overall set of users

Table 3. Number of products added to the basket that came from RS in comparison with MCF

	1st Session	2nd Session	Both
Added after MCF selection without RS	11	17	28
Added after interacting with RS	23	19	42

tions prove that the description box and the recommender have different but complementary usages.

Finally, according to action logs, we evaluated the proportion of products added to the basket thanks to RS, in comparison with those which came from MCF (see Table 3). Of the 70 products globally added to the basket, 28 perfumes (40%) came from the MCF tool without any interaction with RS. 42 products (60%) were added following an interaction with the recommender. We used action logs to compute this, since most of the perfumes coming from the MCF tool were added after taking a look at the different recommendations on the detail page, but without clicking on them. Previous gaze data only revealed that customers used the recommender. This additional information about the origin of basket items shows that the recommender had an influence on users' decisions. The recommender helped increase the confidence of users during the basket construction and offered valuable exploration opportunities.

After the experiment, we asked users to fill out the assessment questionnaire displayed in Table 1. The responses to this post-study survey are summarized in Figure 9.

In order to measure and ensure the veracity of decisions to add products in the basket, we asked users if they would buy the chosen perfumes given the opportunity – keeping in mind that one of the participants was going to win a CHF 100.- voucher to buy one the perfumes he/she had added to the basket – or at least go in a perfume shop to smell them and learn more about them (P10 in Table 1). This data supports the idea that participants took the decision seriously.

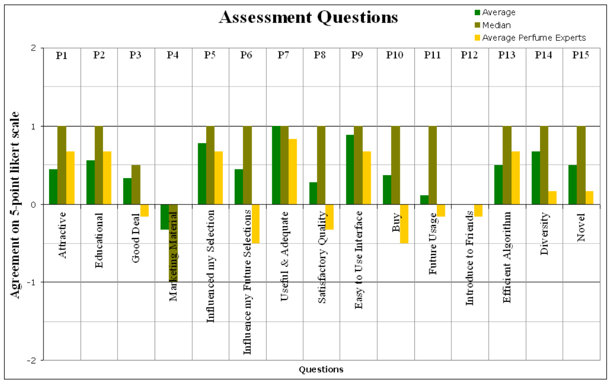

Figure 9. Answers of the Assessment Questionnaire

The main conclusion of our study is that the recommender and each RS category are useful and adequate, which is confirmed by users' answers to question P7 (cf. Figure 9). Answers about attractiveness (P1) and ease of use (P9) reinforce the confidence in categories' utility. The most important category for a given user changes from one session to another in 71% of the cases, proving that there is no habituation bias of usage for a category. The preference-assessment questionnaire also shows that participants were aware of the influence of recommender categories in their selection (P5), and even found recommendations educational (P2). They do not see recommendations as marketing material (P4), which tends to prove that they positively felt this influence and are satisfied with the recommendations (P3, P12). The correlation analysis revealed that users who searched for perfumes they had previously liked, seeking to find novelty from the recommender, also found that the recommendations appeared to be a good deal (P3). This correlation is strong and statistically significant: $r = .849$ ($p = 0.016$). The same users also showed a link with P4, "I believe that the recommender algorithm is efficient" ($r = .806$, $p = 0.029$). However, they do consider themselves to be systematically influenced by the recommender system (P6). Moreover, both the number of recommendation categories used (in terms of clicks) and the number of influential recommendations correlate strongly with the intention to use in the future (P11).

CONCLUSION AND PERSPECTIVES

This paper presented the workings of two decision aid agents and two information entities that are used in common e-commerce websites. A multi-criteria filtering tool (MCF) and a recommender system (RS) were the decision aid agents used. The information entities, under the form of a lexicographically ordered list view (LV) and a description box (D), completed the interactions with consumers. We examined the impact of these four entities over the product brokering task using both eye tracking data and action logs: how and when the agents are most useful to users, their influence on the basket construction, and their added values in comparison with the two information entities.

In their first encounter with the product brokering task, users' reliance on the recommender agent seemed to be the strongest. Action logs reveal that they consulted it more frequently and twice as many basket items come from RS than from its MCF counterpart (23 from RS vs. 11 from MCF). After users learned more about the domain knowledge, their reliance on both agents becomes somewhat comparable as can be seen from the comparison of their first and second encounter of the tasks (19 from RS vs. 17 from MCF). Globally on both sessions, a total of 42 items come from RS (60%) versus 28 from MCF (40%).

Gaze data further shows that users consulted the recommender system as much as the description box or the list view. They even look at the recommender twice as much as the multi-criteria filtering tool. Our user study allowed us to understand the distinctive roles of each of the four entities. In particular, the results highlighted that users more frequently pay attention to the recommendation categories that provide additional information relative to what they were looking at in the description box. Users were more interested in recommendations related to popularity, which is one of the least considered dimensions in the description box, or providing the biggest diversity (2 times more clicks than average for ”More popular and cheaper”, 1.5 time more for ”People who like this also like”).

On a general level, this paper shows how to conduct an eye tracking study to understand users interaction behavior with intelligent systems, and is thus a relevant topic to the IUI community. We plan to use these results to improve agents' ability to help users make decisions, inspire consumer confidence to purchase and maximize the technology adoption among potential users.

REFERENCES

1. W. Albert. Do web users actually look at ads? a case study of banner ads and eye-tracking technology. In *In Proceedings of the Eleventh Annual Conference of the Usability Professionals Association*, Orlando, Florida, USA, 2002.
2. L. J. Ball, E. J. Lucas, J. N. V. Miles, and A. G. Gale. Inspection times and the selection task: What do eye-movements reveal about relevance effects? *Quarterly Journal of Experimental Psychology*, 56A:1053–1077, 2003.
3. S. Castagnos, N. Jones, and P. Pu. Recommenders' influence on buyers' decision process. In *Proceedings of the 3rd ACM Conference on Recommender Systems (RecSys 2009)*, New-York, USA, October 2009.
4. S. Castagnos, N. Jones, and P. Pu. When recommenders lead to confidence: a story of how technology changes online search behaviors. Technical report, 2009.
5. L. Chen and P. Pu. Preference-based organization interfaces: Aiding user critiques in recommender systems. In *In Proceedings of International Conference on User Modeling (UM'07)*, pages 77–86, Corfu, Greece, June 2007.
6. T. H. Davenport and J. C. Beck. *The Attention Economy: Understanding the New Currency of Business*. Harvard Business School Press, 2001.
7. J. H. Goldberg and X. P. Kotval. Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*, 24:631–645, 1999.
8. M. A. Just and P. A. Carpenter. Eye fixations and cognitive processes.
9. G. L. Lohse. Consumer eye movement patterns on yellow pages advertising. *Journal of Advertising*, 26:61–73, 1997.
10. P. Maes, R. Guttman, A. Moukas, and R. Moukas. Agents that buy and sell: Transforming commerce as we know it. *Communications of the ACM*, 42:81–91, 1999.
11. J. Nielsen. F-shaped pattern for reading web content. Technical report, 2006.
12. A. Poole and L. Ball. *Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects*. Idea Group, 2006.
13. P. Pu, M. Zhou, and S. Castagnos. Critiquing recommenders for public taste products. In *Proceedings of the 3rd ACM Conference on Recommender Systems (RecSys 2009)*, New-York, USA, October 2009.
14. D. Yoon and N. H. Narayanan. Mental imagery in problem solving: An eye tracking study. In *In Proceedings of the Eye Tracking Research and Applications Symposium 2004*, pages 77–83, New York, USA, 2004.
15. G. Zelinsky and D. Sheinberg. *Why some search tasks take longer than others: Using eye movements to redefine reaction times*. North-Holland: Elsevier, 1995.