Effects of Search Success on Search Engine Re-Use

Victor Hu
Department of Statistics
Stanford University
Stanford, CA 94305
vhu@stanford.edu

Maria Stone and Jan Pedersen
Microsoft Bing
65 La Avenida
Mountain View, CA 94043
{mariast, jpederse}@microsoft.com

Ryen W. White
Microsoft Research
One Microsoft Way
Redmond, WA 98052
ryenw@microsoft.com

ABSTRACT
People's experiences when interacting with online services affects their decisions on re-use. Users of Web search engines are primarily focused on obtaining relevant information pertaining to their query. Search engines that fail to satisfy users' information needs may find their market share to be negatively affected. However, despite its importance to search providers, the relationship between search success and search engine re-use is poorly understood. In this paper, we present a longitudinal log-based study with a large cohort of search engine users that quantifies the relationship between success and re-use of search engines. We use time series analysis to define two groups of users: stationary and non-stationary. We find that recent changes in satisfaction rate do correlate moderately with changes in rate of return for stationary users. For non-stationary users, we find that satisfaction and rate of return change together and in the same direction. We also find that some effects are stronger for a smaller player on the market than for a clear market leader, but both are affected. This is the first study to explore these issues in the context of Web search, and our findings have implications for search providers seeking to better understand their users and improving their experience.

Categories and Subject Descriptors
H.3.3 [**Information Search and Retrieval**]: *search process*.

General Terms
Experimentation, Measurement.

Keywords
Search success, user satisfaction, longitudinal log analysis.

1. INTRODUCTION
Search success is the primary goal of information seeking. Satisfaction may be a critical determinant of continued use of search engines, or a range of other goods or services, over time [1][17]. Indeed, research on search engine switching behavior has shown that satisfaction is important in predicting switching [12][30]. Research on the success-reuse relationship is necessary to help search providers better understand factors affecting their usage and improve the search experience for their users, retain and grow their user base, and increase revenue from search advertising.

While it is commonly assumed that user satisfaction with a search engine will lead to more frequent returns to that engine, there have been no previous attempts to characterize and document this relationship. General literature on the relationship between customer satisfaction and market share finds that some relationship exists [1], but the relationship between satisfaction and rate of return was never studied for Web search specifically. There has been some research on modeling engine usage patterns over time [19][24][30][33]. This work has focused on modeling engine loyalty, or predicting switching events or encouraging the use of multiple search engines. However, research is needed on the relationship between search satisfaction and engine re-use.

In this paper, we analyze log data gathered from the toolbar of a large commercial search engine for a random sample of users over a period of six months. We estimated search satisfaction from result clicks over 30 seconds in duration [7][8], and compute satisfaction ratio for each user for each week for each search engine. We also compute the rate of return to a search engine (total number of searches conducted on each search engine by each user for each week). We first use time series analysis to define two groups of users: (i) *stationary users* whose rate of return and satisfaction ratios remain stable over the course of 25 week duration (the vast majority of users), and (ii) *non-stationary users*, users for whom probability distribution for both rate of return and satisfaction ratios change over the study duration (a small minority of users). We investigate the relationship between satisfaction ratios and return behavior for stationary users, and users whose satisfaction ratio and rate of return are both changing. We also compare success and re-use dynamics across the major and minor engines.

The remainder of this paper is structured as follows. Section 2 describes related work. Section 3 describes the study that we performed, and the results are presented in Section 4. These results and their implications are discussed in Section 5.

2. RELATED WORK
There are four areas of prior work that are most relevant for this investigation: (i) customer satisfaction and loyalty, (ii) online metrics to measure the quality of user experience, (iii) defining search success based on online signals, and (iv) characterizing and predicting short- and long-term search engine switching behavior.

***Customer Satisfaction and Loyalty*:** There is a large body of work regarding product or brand switching and the relationship between satisfaction and loyalty (e.g., [23]). Although customer satisfaction is the predominant metric used by companies to detect and manage defections to competitors [1], more recent research has found that knowledge of competitors and attitudinal and demographic factors, among other influences, can also play an important role [24]. Research in these areas has also shown that complete satisfaction is significantly more important than just satisfaction, and even satisfied customers may still defect [17].

***Online User Experience Metrics*:** Since the business community focuses on user return rather than user satisfaction [4], it is not surprising then that most commercially available Web analytics packages (e.g., [9][25]) do not include metrics related to user satisfaction. Rodden *et al.* [27] refer to traditional business analytics-driven metrics as PULSE metrics (page views, uptime, latency, seven day active users, and earnings). While these metrics

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CIKM'11, October 24–28, 2011, Glasgow, United Kingdom.
Copyright 2011 ACM 1-58113-000-0/00/0010…$10.00.

may tell us how well a product or service is doing, it does not facilitate any links between user success and business success. To address this shortcoming, Rodden *et al.* propose an alternative framework for user-centric metrics, called HEART (happiness, engagement, adoption, retention, and task success). This suite of metrics focuses instead on user happiness. The authors recommend a process by which specific goals for a product can be articulated and measured, but this approach still does not make the connection between measures of user satisfaction (e.g., task success) and measures of retention and engagement (e.g., number of visits a week per user).

Online experimentation is useful to evaluate new features or new ranking methodologies [21][28], but it is unclear what metrics should be used. It is not unusual for these experiments to observe an increase in one kind of metric (e.g., task success), but also observe no change or a decrease in other metrics such as engagement. It is important to investigate whether an increase in task success leads to increased usage, for what users and under what circumstances. That is the goal of the research that we present.

***Search Success in Online Experimentation*:** One significant obstacle to overcome in using online metrics is selecting a reliable measure of success using only online signals. Search success is subjective and frequently cannot be inferred from logs alone [11]. Many different sources of information need to be used to establish with certainty that a search was successful. Extensive literature exists on trying to derive indicators of task success or failure from online user behavior. In search, specifically, several fruitful approaches have been tried. One approach is to correlate behavior with either self-reported success [8] or labels of success provided by expert judges [7][15]. Early investigations correlated self-reported measures of search satisfaction with implicit signals, such as clicks and dwell time for clicks [8]. More recently, a sequence of user actions in a session was presented to an independent judge, who determined if a user was successful [15]. These labels can then be learned from online signals, such as click order, click dwell time, and click sequence. A similar methodology was employed to define user frustration [7]. One clear association that emerged from these early investigations is that longer clicks are more likely indicative of search success than shorter clicks.

***Search Engine Switching*:** Heath and White [16] developed models for predicting switching within search sessions. White and Dumais [18] used log analysis and a survey to characterize search engine switching behavior, and used query, session, and user features to predict engine switching. Guo *et al.* [13] used a browser plugin to capture people's motivations for engine switching. Their findings show that dissatisfaction was the most common rationale for switching. Mukhopadhyay *et al.* [24] found that dissatisfaction with search engine results had both short-term and long-term effects on search engine choice. Juan and Chang [19] looked at weekly search engine switching behavior as a predictor of market share. They found that user engagement, engine preferences, market share and number of search sessions are positively correlated.

The question left unanswered by Juan and Chang's study is what will drive engagement for a search engine. Many different factors impact the rate of user return—there are seasonality changes and sudden, unexpected disruptions in usage patterns due to users' history or events around them. With so few search engine choices, most users are settled in a stable pattern of usage and may have learned a seemingly optimal way to search, such that neither their satisfaction rates nor rate of return change significantly. A better understanding of the relationship between search success and rate of return can help search engines better understand users' decision making processes and help those engines more effectively tailor the services they offer and more accurately measure their impact.

We now present a novel methodology for detecting and describing the relationship between satisfaction and engine rate of return[1].

3. DATA AND METHODOLOGY

We describe the log data used to perform our longitudinal study, the methods used to estimate searcher satisfaction from those data, and the methods to study the satisfaction-return relationship.

3.1 Log Data

The data that we use in this analysis is the same as the data underlying prior work by White and Dumais [30]. We analyzed six months of interaction logs from September 2008 through February 2009 inclusive. The logs came from hundreds of thousands of consenting users through a widely-distributed browser toolbar. To remove variability caused by geographic and linguistic variation in search behavior, included entries were from the English speaking United States locale. From these logs we extracted *search sessions*. Every session began with a query issued to the Google, Yahoo!, or Microsoft Live Search Web search engines, and could contain further queries or Web page visits, including search engine result page (SERP) clicks and post SERP click navigation. A session ended if the user was idle for a period of more than 30 minutes. Similar criteria have been used previously to demarcate search sessions in search log data (e.g., [30]).

3.2 Estimating Searcher Satisfaction

Given the large number of searchers and search queries, we needed an automated way of estimating search satisfaction with search engine results. To do this, we use a searcher's dwell time on a clicked result as our primary indicator of satisfaction. Dwell time was computed as the difference between page load times in the logs of subsequent pages in a search session. This definition of search satisfaction is justified by prior findings that longer clicks are more likely to be successful clicks [7][8][21]. Search satisfaction in this study is defined as at least one click of 30 or more seconds following a query. The *satisfaction ratio* for each user for each week for each search engine is the ratio of the number of queries with search satisfaction and the total number of queries for that week, for that user, and for that engine. We also compute total number of queries issued on each engine by each user for each week as number of returns to the search engine. We refer to these query counts as the *return rate* in the rest of the paper.

3.3 Methodology

We focus on how the satisfaction with the engine itself impacts the rate of return to that engine. It is likely that satisfaction with a competitor may also be a factor in search engine usage, as well as many other external factors that may be near impossible for us to study remotely using search engine logging. We wanted to first examine satisfaction with the search engine itself and rate of return to that engine. Our overarching goal was to develop a methodology for understanding how a short-term behavioral marker (search satisfaction) impacts long-term return behavior.

As a first step, we performed time series analysis of both search satisfaction ratio and number of returns (number of searches a week) for all users over the nearly six-month duration of the study

[1] Note that for the purposes of our analysis we use search satisfaction as a proxy for search success in the remainder of this paper.

for each search engine. This was necessary to understand whether both or either of the variables changed systematically over the course of 25 week duration of the study. The data from all users of the major engine and all users of minor engine were included into the analysis. The minor and major engines were a subset selected from the three search engines present in our data (i.e., Google, Yahoo!, and Live Search). We computed summary indices of search satisfaction and rates of return for every user for every week in the logs. There were a total of 111,908 users of the major search engine who had non-zero number of returns and non-zero satisfaction rate for at least one of the 25 weeks included in this data analysis, and 95,969 users of the minor search engine who met the same standard. We also identified the two user groups—*stationary* and *non-stationary*—that are central to our study of the satisfaction-return relationship. Stationary users have unchanging search satisfaction ratio and number of returns processes; non-stationary users have both processes changing over the 25 weeks.

As a second step, we performed separate analyses on stationary and non-stationary users. Rate of return and satisfaction both constitute time series data. When both time series are stationary processes (that is, they are fluctuating but not changing over time around a fixed value), standard regression analysis is appropriate to understand whether the two series are related. However, when time series data are changing over time (non-stationary processes), regression analysis is not appropriate because the assumptions behind standard regression and correlation analysis are violated, and the expected value of the correlation coefficient under the null hypothesis is no longer zero [10]. We defined groups as follows:

Stationary users: We used a general linear model constructed with each week's data to understand if changes in satisfaction led to changes in rates of return. We computed a change in satisfaction ratio for each user for each week relative to the mean satisfaction ratio, obtained over the study duration. We also computed change in rate of return, both relative to the mean return rate for each user and relative to previous week's rate of return. The formulae that were used are presented in the analysis section later in the paper.

Non-stationary users: We performed co-integration analysis to understand if changes in satisfaction ratio and rate of return are related (co-occur in time-series) [6][10]. Co-integration analysis was created specifically to deal with spurious correlation when dealing with changing time series data. This analysis determines whether or not any linear combination of these two variables is stationary. Suppose $Y(t)$ is the number of return process, $X(t)$ is the satisfaction ratio process. A test for co-integration will first regress $Y(t)$ on $X(t)$ to get a coefficient ß. If ß is above zero, then there is a positive relationship between $X(t)$ and $Y(t)$. Once ß is determined for each user, co-integration will test whether $Y(t) - ß \cdot X(t)$ is stationary. That is, co-integration will test if the leftover residual error in the regression exhibits any trends (up or down). If the error does *not* exhibit a trend, then all change in $Y(t)$ is accounted for by change in $X(t)$. For users for whom this is true, we can assume that the success-reuse correlation is real.

4. RESULTS OF DATA ANALYSIS

4.1 Aggregate Data Description

To understand whether or not overall user satisfaction ratio and the rate of return were associated, we looked at some basic descriptive statistics for satisfaction ratio and number of returns. The correlation between average satisfaction ratio and average return rate for all users across the full 25-week period is moderate and positive for the minor search engine (r=0.39, p<0.001), and weak and positive for the major search engine (r=0.12, p<0.001). The very presence of such differences suggests that satisfaction and usage may relate differently depending on the search engine.

Table 1 summarizes our data set and aspects of the search activity of those who used the major and/or minor search engine.

Table 1. Summary statistics of the 25-week data set. Standard deviation values are parenthesized where appropriate.

Statistic	Major engine	Minor engine
Number of users	111,908	95,969
Percentage of all queries	80.8%	19.2%
Avg. return rate [queries]	78.1 (sd=23.2)	33.3 (sd=16.6)
Avg. satisfaction ratio	0.34 (sd=0.04)	0.26 (sd=0.03)

There is an imbalance between the major and minor engines in the fraction of all queries observed in our data. Users also issue more queries on average per week to the major engine, and interestingly for this study, users of the major engine appear to experience significantly higher levels of satisfaction than the minor engine. All differences between the major and minor search engine are statistically significant at p < 0.01 using independent measures t-tests.

4.2 Time Series Analysis

The first component of our analysis involved treating both return rate and satisfaction ratio as time series. Time series are a set of observations arranged in chronological order according to the time at which they occurred. The characteristic property of a time series is the fact that the data are not generated independently, their dispersion may vary in time, they may exhibit a trend, and/or they may have cyclic components [11]. Since our data match this description, statistical procedures supposing independent and identically distributed data cannot be used to analyze our data.

The first task in our time series analysis is establishing if either of the variables in question (satisfaction or return) exhibits a trend. We performed time series analysis for each of the time series involved—return rate for each of the two search engines and satisfaction ratio for each of the two engines. We did this for each user to determine what proportion of the users meet our definition of stationary users: that the distribution for a given user remains time invariant, suggesting that the joint probability distribution is stable with a time shift. The behavioral implication is that the data fluctuates around this user's mean, but does not change over time.

We used the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test [22], with the alpha for statistical significance testing set to 0.05, to determine whether our 25-week data was stationary. Table 2 reports the percentage of users who had significant trends in both variables (non-stationary users), the percentage of users with no significant trends in either (stationary users) or who had significant trend in only one variable (satisfaction or return). "SAT" is used in the table and elsewhere an abbreviation for satisfaction.

Table 2 shows that most of the users in our data set (around 70%), on both of the engines are stationary with no changes in the satisfaction ratio of the return rate in the course of our study. We assign those users whose success and re-use patterns remain time invariant for the duration of the study to our groups of *stationary* users (one group per engine). Those users with trends in satisfaction and returns over time (around 5% of users on both engines) form our groups of *non-stationary* users (one group per engine).

Table 2. Results of time series analysis.

Analysis Outcome	Major Engine	Minor Engine
Both SAT ratio and return rate stationary	70.1%	76.6%
Only return rate changes	16.2%	13.9%
Only SAT ratio changes	8.9%	5.8%
Both SAT ratio and return rate non-stationary	4.6%	4.8%

We now present a more detailed analysis of two of the four groups identified in this section: stationary and non-stationary users. These groups differ the most in terms of the satisfaction and return relationship, and are hence most likely to offer meaningful differences out of the possible pairwise comparisons of groups.

4.3 Analyzing Stationary Users

Stationary users are those for whom neither the rate of return nor satisfaction ratio change directionally over time. For these users, simple linear regression is appropriate to understand the relationship between satisfaction and return. Given that the mean satisfaction itself does not change and neither does rate of return for these users, correlation between these values over time is likely to be near zero. An interesting question about these users is whether slight changes in satisfaction from their respective means correlates with slight changes in return relative to their respective means, either this week or in previous weeks. Another question is whether or not this week's return changes relative to previous week's return in response to changes in satisfaction ratio.

We defined two response variables ($R1$, $R2$) for each user:

$$R1 = \frac{Current\ Week\ Num.Returns - Mean\ Num.Returns}{Mean\ Num.Returns}$$

$$R2 = \frac{Current\ Week\ Num.Returns - Previous\ Week\ Num.Returns}{Mean\ Num.Returns}$$

The response variables are variants of the normalized changing number of returns for each week. $R1$ lets us study changes in the current week relative to the mean, whereas $R2$ lets us study changes in the current week relative to the previous week.

To capture the satisfaction ratio at different times, allowing us to look for time lag effects where changes in satisfaction may not be immediately reflected in return, we defined three interest variables ($F1$, $F2$, and $F3$) using the normalized changing satisfaction ratio:

$$F1 = \frac{Current\ Week\ SAT\ ratio - Mean\ SAT\ ratio}{Mean\ SAT\ ratio}$$

$$F2 = \frac{Previous\ Week\ SAT\ ratio - Mean\ SAT\ ratio}{Mean\ SAT\ ratio}$$

$$F3 = \frac{SAT\ ratio\ Two\ Weeks\ Ago - Mean\ SAT\ ratio}{Mean\ SAT\ ratio}$$

To examine the effect of the features on the response variables, we use simple linear regression. $R1$ is a change in rate of return this week relative to mean rate of return for this user. The analysis for the major engine and the minor engine on the market are summarized in the Table 3. In the table, R^2 denotes the standard coefficient of determination in a standard regression model and ß denotes the coefficient in the regression equation. ß is positive when increases in satisfaction ratios are associated with increases in return rate (when ß is negative it means the opposite).

Table 3. Overall effect of SAT change on return change in stationary users relative to 25-week mean.

Engine	R^2 for *F1*	ß for *F1*	R^2 for *F2*	ß for *F2*
Major engine	0.23	0.49	>0.000	-0.001
Minor engine	0.26	0.40	>0.000	0.006

For $R1$, only $F1$ is significant, and neither $F2$ nor $F3$ have any effect for either the major search engine or the minor engine. That means that a positive relationship between current week's normalized satisfaction ratio and current week's normalized return exists, but normalized satisfaction ratio from the previous week (or earlier) has no impact on this week's normalized return rate. Table 3 does not include $F3$ because coefficient is virtually zero and p-value well exceeds α (meaning non-significant).

We also performed similar analysis related to the change in return rate relative to previous week's return ($R2$). This additional analysis tells us how normalized changes in rate of return from week to week respond to changes in satisfaction ratio. The findings are summarized in Table 4 in the same format as the $R1$ analysis.

Table 4. Overall effect of satisfaction change on return change in stationary users relative to previous week's return.

Engine	R^2 for *F1*	ß for *F1*	R^2 for *F2*	ß for *F2*
Major Engine	0.11	0.48	0.14	-0.53
Minor Engine	0.13	0.40	0.14	-0.41

As we can see in the table above, for the response variable $R2$, the effects of $F1$ and $F2$ are nearly opposite (as indicated by the different signs for), and once again $F3$ has no effect. For stationary users, the current week's increase in satisfaction ratio helps lift the return, and the previous week's increase in satisfaction ratio makes return rate drop (perhaps since it causes previous week's return to be higher), and their effects almost cancel each other out. Any changes in satisfaction rate prior to the previous week appear to have no effect. Essentially, we can take from this analysis that although the satisfaction ratio and rate of return fluctuate together, the effect of any fluctuations is very short-lived.

4.4 Analyzing Non-Stationary Users

Given that around 5% of users of each search engine were non-stationary, if satisfaction ratio and return rate both changed, to what extent did they change together and in the same direction?

Figure 1 illustrates average satisfaction ratios and average rates of return for co-integrated and not co-integrated non-stationary users. The charts illustrate clearer trends for co-integrated users, especially on the minor search engine, suggesting (at least visually) that the co-integration analysis is successfully identifying users with a clearer relationship between satisfaction and return. However, these plots of averages across all users are only a summary, and do not tell the full story, especially since the co-integration analysis was performed individually for each user.

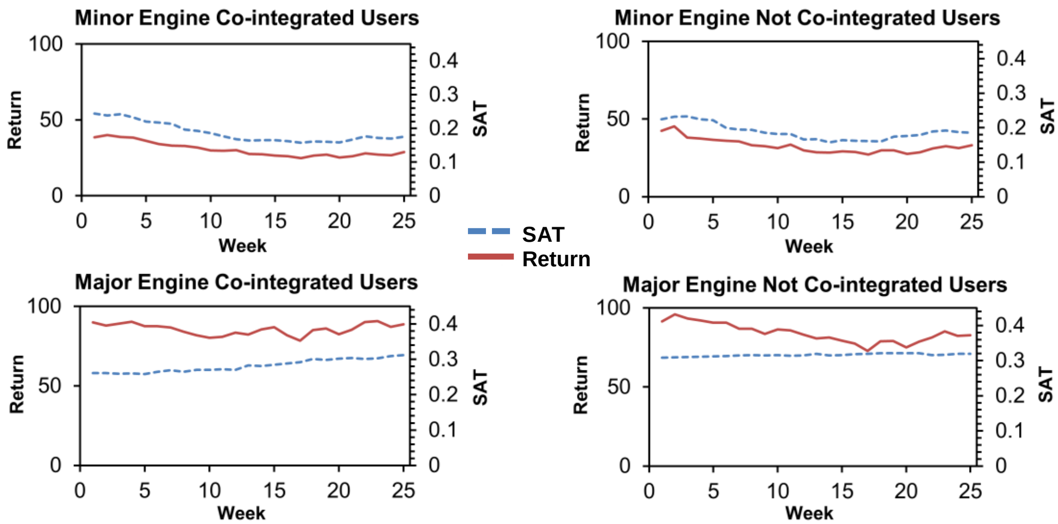

Figure 1. Average SAT and Return over Time for Co-integrated and Not Co-integrated users.

Table 5 shows the comparison between stationary users and non-stationary users for both search engines. The percentage in the left column is the fraction of users in each (user group, search engine) pair with a positive correlation between satisfaction and return. In the right column we show the percentage of users who were co-integrated, and for whom there may be a real relationship between satisfaction and return (per co-integration analysis).

Table 5. Results of regression analysis for all users and co-integration analysis for non-stationary users.

User Group and Search Engine	% of ß > 0 (positive correlations)	% of $Y(t) - ß \cdot X(t)$ (co-integrated users)
Non-stationary, minor engine	100.0%	70.6%
Non-stationary, major engine	100.0%	41.8%
Stationary, major engine	33.9%	n/a
Stationary, minor engine	74.6%	n/a

For non-stationary users, the relationship between satisfaction and return is uniformly positive. In contrast, stationary users do not react consistently to changes in satisfaction ratio over 25 weeks. For the minor search engine, 70.6% of the non-stationary processes show the pattern of co-integration. That means that for those users we can expect changes in rate of return and changes in satisfaction to co-occur. If we are tracing market share, and given that 4.8% of the users are changing their search behavior, we can be reasonably confident that 4.8% × 70.6% = 3.4% of all the users of the minor search engine will increase (or decrease) their long term rate of return as a result of increase (or decrease) in their satisfaction ratio. The result for major engine (the market winner) is a lot weaker. If roughly 4.6% of users are non-stationary, only 4.6% × 41.8% = 1.9% of that engine's users will increase (decrease) their long-term return rate as a result of increase (decrease) in success.

This analysis may be useful to carry out periodically to understand how many users of each search engine fall into the non-stationary category, and for how many of them the two variables are co-integrated. This underscores the opportunity that the minor engine has to gain more traffic by improving search satisfaction for its users. It also suggests that improving search satisfaction may not result in similar gains for the major search engine in the market.

5. IMPLICATIONS AND FUTURE WORK

We characterized the success-reuse relationship for two groups of users. We showed that changes in search satisfaction ratio and changes in rate of return are positively correlated when these two variables are relatively stable over time (as they are for the vast majority of *stationary* users, who comprise 70-75% of our all users) and when both variables change (as they are for a small minority of *non-stationary* users, totaling 5% of all users). For these non-stationary users, the satisfaction ratio and rate of return change in the same direction. For a sizeable proportion of the changing users, the change in satisfaction ratio variable completely accounts for the change in rate of return. This proportion is higher for the minor player on the market than for a market winner. This is a place where the minor engine could gain (or lose) more market share than the major engine.

One explanation for why a stronger relationship was observed for the minor player may be that the average satisfaction ratio for the minor player was much lower. It may very well be that once the minor market player improves search performance to be equivalent to the market winner, it will be harder to see movements in rates of return in response to changes in satisfaction.

An important limitation of the research presented is that we relied upon a single, rudimentary definition of search success, namely result clicks of at least 30 seconds in duration. Although this definition has been used extensively in related work, online signals can be noisy and this measure in particular not normalized relative to user expectation or query difficulty. More difficult queries may result in fewer satisfied clicks, but users may adjust their expectations for such queries. To make further progress, we need to better understand how to adjust on-line signals of satisfaction for different levels of query difficulty. The information retrieval community has already studied the automatic estimation of query difficulty that could be useful here (e.g., [12]).

One area for future work is to understand whether more and less restrictive automatic definitions of search satisfaction (e.g., search engine switching behavior) exhibit the same relationship with rates of return. Another, more laborious approach, may be to identify satisfaction by manually labeling clicked search results, rather than relying on a purely online definition. Specific metrics that take query difficulty into account already exist for relevance labels (such as *normalized discounted cumulative gain* [18]). Alternatively, we could deploy a browser plugin to gather satisfaction judgments from volunteers *in-situ* at session time, as has been used previously in similar scenarios in the search domain [8][13].

In this paper we demonstrate that satisfaction with a search engine can contribute to users' propensity to use and re-use its service. There are many factors that drive increases in market share and search volume that are not connected to satisfaction with search. These factors may include advertising campaigns, browser defaults and toolbars that alter the convenience and availability of the search engine services when and where users may need this service, etc. Although the analysis shows a correlation between satisfaction and engine re-use, it does not prove causation. Only through comprehensive experimentation can we offer a convincing demonstration of causality. Such experiments are not always possible, especially remotely at Web scale, and a notion of *statistical causality* [10] might have useful applications here.

Tracking rates of return and better understanding searchers' rationales is critically important for engine success. Reliable changes in rates of return are slow to accumulate and take a long time to track. Finding short-term surrogate measures that are associated with long term changes could be useful to rapidly estimate long-term implications of feature additions or ranking enhancements.

REFERENCES

[1] Anderson, E.W. and Sullivan, M.W. (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing Science* 12: 125–143.

[2] Anderson, R.E. and Srinivasan, S.S. (2003). E-satisfaction and e-loyalty: a contingency framework. *Psychology and Marketing*, 20(3): 123–138.

[3] Bolton, R.N. (1998). A dynamic model of the duration of the customer's relationship with a continuous service provider: the role of satisfaction. *Marketing Science*, 17(1): 45–65.

[4] Burby, J.Y. and Atchson, S. (2007). *Actionable Web Analytics*. Wiley Publishing, Inc., Indianapolis, IN.

[5] Capraro, A.J., Broniarczyk, S., and Srivastava, R.K. (2003). Factors influencing the likelihood of customer defection: the role of consumer knowledge. *Journal of the Academy of Marketing Science*, 31(2): 164–175.

[6] Engle, R. F. and Granger, C.W.J. (1987). Co-integration and error correction: representation, estimation, and testing, *Econometrica*, 55: 251–276.

[7] Field, H., Allan, J., and Jones, R. (2010). Predicting searcher frustration. *Proc. SIGIR*, 34–41.

[8] Fox, S., Karnawat, K., Mydland, M., Dumais, S.T., and White, T. (2005). Evaluating implicit measures to improve the search experience. *ACM TOIS*, 23(2): 147–168.

[9] Google Analytics (http://www.google.com/analytics).

[10] Granger, C.W.J. (2003). The Nobel Prize Lecture. http://nobelprize.org/nobel_prizes/economics/laureates/2003/granger-lecture.html

[11] Grimes, C., Tang, D., and Russell, D.M. (2007). Query logs alone are not enough. *Proc. Query Log Analysis Workshop*.

[12] Guo, Q. White, R.W., Dumais, S., Wang, J. and Anderson, B. (2010). Predicting query performance using query, result, and user interaction features. *Proc. RIAO*.

[13] Guo, Q., White, R.W., Zhang, Y., Anderson, B., and Dumais, S.T. (2011). Why searchers switch: Understanding and predicting engine switching rationales. *Proc. SIGIR*, 335–344.

[14] Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, NJ.

[15] Hassan, A., Jones, R., and Klinkner, K. (2010). Beyond DCG: user behavior as a predictor of a successful search. *Proc. WSDM*, 221–230.

[16] Heath, A.P. and White, R.W. (2008). Defection detection: predicting search engine switching. *Proc. WWW*, 1173–1174.

[17] Jones, W.E. and Sasser, T.O. (1995). Why satisfied customers defect. *Harvard Business Review* (November-December)

[18] Järvelin, K. and Kekäläinen, J. (2000). IR evaluation methods for retrieving highly relevant documents. *Proc. SIGIR*, 41–48.

[19] Juan, Y.F. and Chang, C.C. (2005). An analysis of search engine switching behavior using click streams. *Proc. WINE*, 806–815.

[20] Keaveney, S.M. and Parthasarathy, M. (2001). Customer switching behavior in online services: an exploratory study of the role of selected attitudinal, behavioral, and demographic factors. *Journal of the Academy of Marketing Science*, 29(4): 374–390.

[21] Kohavi, R., Henne, R.M., and Sommerfield, D. (2007). Practical guide to controlled experiments on the web: listen to your customers not to the hippo. *Proc. SIGKDD*, 259–267.

[22] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin (1992): Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54: 159–178.

[23] Mittal, B. and Lassar, W.M. (1998). Why do customers switch? the dynamics of satisfaction versus loyalty. *Journal of Services Marketing*, 12(3): 177–194.

[24] Mukhopadhyay, T., Rajan, U., and Telang, R. (2004). Competition between internet search engines. *Proc. Hawaii International Conference on System Sciences.*

[25] Omniture (http://www.omniture.com).

[26] *Pew Internet and American Life Project: Search Engine Users* (2005) (accessed December 2010)

[27] Rodden, K., Hutchinson, H., and Fu, X. (2010). Measuring the user experience on a large scale user-centric metrics for web applications. *Proc. SIGCHI*, 2395–2398.

[28] Tang, D., Agrawal, A., O'Brien, D, and Meyer, M. (2010). Overlapping experiment infrastructure: more better faster experiments. *Proc. SIGKDD*, 17–26.

[29] White, R.W. and Drucker, S.M. (2007). Investigating behavioral variability in Web search. *Proc. WWW*, 21–30.

[30] White, R.W. and Dumais, S.T. (2009). Characterizing and predicting search engine switching behavior. *Proc. CIKM*, 87–96.

[31] White, R.W., Kapoor, A., and Dumais, S.T. (2010). Modeling long-term search engine usage. *Proc. UMAP*, 28–39.

[32] White, R.W. and Morris, D. (2007). Investigating the querying and browsing behavior of advanced search engine users. *Proc. SIGIR*, 255–262.

[33] White, R.W., Richardson, M., Bilenko, M., and Heath, A.P. (2008). Enhancing web search by promoting multiple search engine use. *Proc. SIGIR*, 43–50.