Search Query Formation by Strategic Consumers

Jia Liu and Olivier Toubia*

March 9, 2018

Abstract

Submitting queries to search engines has become a major way in which consumers search for information and products. The massive amount of search query data available today has the potential to provide valuable information on consumer preferences. In order to unlock this potential, it is necessary to understand how consumers translate their preferences into search queries. In this paper, we argue that if consumers are strategic, then their search queries should *not* be a direct representation of their preferences: strategic consumers should formulate queries that are likely to retrieve the content they are searching for, rather than being merely similar to that content. We present secondary and primary evidence that is consistent with a strategic view of query formation. Using secondary field data, we illustrate the benefits for consumers from strategically formulating queries which contain only a subset of the terms they are interested in, but which are effective at retrieving other terms of interest. Our incentive-aligned lab experiment confirms that consumers have at least some ability to be strategic when formulating search queries.

*Jia Liu is a Post-doctoral researcher at Microsoft Research New York Lab. Email: jiliu3@microsoft.com. Olivier Toubia (corresponding author) is Glaubinger Professor of Business, Graduate School of Business, Columbia University. Email: ot2107@columbia.edu.

1 Introduction

Every minute, more than 3 million queries are submitted to Google (InternetLiveStats, 2016). Each of these queries is some expression of one consumer's preferences (Pirolli, 2007), which is voluntary and incentive-compatible (to the extent that the consumer is motivated to find useful content). As such, online search query data present a massive opportunity for revealed preference measurement. Such data may not only be a source of consumer insights, they are also essential for search marketing, an industry expected to be worth $80 billion by 2020 in the U.S. only (Borrell, 2016). For example, the amount a marketer should be willing to bid on a particular keyword is a function of how well their content matches with the preferences of consumers who usually type that keyword.

Despite the ubiquity of search engines and the size of the search marketing industry, to the best of our knowledge no revealed preference framework has been developed for inferring consumer preferences from their search queries. A long literature in marketing and economics has modeled search by utility-maximizing consumers (e.g., Stigler, 1961; Weitzman, 1979; Gabaix et al., 2006). Some of this literature has even studied search in the context of search engines (e.g., Kim et al., 2010; Rutz and Trusov, 2011; Jeziorski and Segal, 2015). However, this literature has been primarily limited to situations where search is performed by a series of discrete choices (e.g., purchases, clicks). Text-based search is *not* a straightforward special case of discrete-choice search in which consumers would select from a very large universe of queries. In particular, search terms are semantically related to each other and to the search results, which creates a rich set of dependencies between queries and their results.

In order to extend utility-based search models to online search queries, it is necessary to specify some assumptions on how consumers formulate queries given their preferences. That is, what is the link between the content that a consumer wants to consume, and the queries they formulate on a search engine? Unfortunately, the Information Retrieval (IR) literature does *not* provide answers to this question, because it has focused on optimizing the search results given a query, from the perspective of the search engine (e.g., Manning et al., 2008; Ruthven, 2003; Li and Xu, 2013; Santos et al., 2015). In contrast, we focus on the question of how consumers translate their content preferences into search queries, taking the search engine as given. We define a consumer's "content preferences" as the textual content that this consumer desires to "consume." Depending on the context, content may consist of news stories, information about products and services, or information about any domain of knowledge.

A naive view of query formation would be that consumers simply formulate queries that contain the most relevant terms they are searching for. For example, consider a consumer typing the following query: "affordable sedan made in America." If the naive assumption were correct, then inferring a consumer's preferences from their queries would be straightforward, as their search queries would be a direct representation of their preferences. However, if consumers are strategic in the formulation of their queries, then queries should not be direct expressions of preferences, but rather a tool used by consumers to retrieve certain content. Using the same example, it might be possible that the most important attributes for this consumer are in fact safety, comfort, and made in America, and that affordability is of lesser importance. This consumer might have decided to type the query "affordable sedan made in America" because they believe that cars made in America are generally safe and comfortable, but not necessarily affordable. In that case, the consumer anticipated that they would find results that match their preferences efficiently (i.e., with short queries) by only including "made in America" and "affordable" in their queries, but not "safe" or "comfortable," although these are important attributes.

In this paper, we illustrate some benefits from strategic query formation using field data; and we explore whether consumers indeed have the ability to be strategic in query formation, using an incentive-aligned experiment. In doing so, we hope to lay some foundations for the future development of utility-based search models that include search queries.

2 Field Data

It is well known empirically that consumers have a strong preference for short queries (Jansen et al., 2000; Spink et al., 2001). This suggests that consumers would benefit from formulating short queries that contain only a subset of the terms they are interested in, but that are effective at retrieving other terms of interest. In this section, we illustrate that such queries not only exist, but that they are also popular among online users.

2.1 Data

Shorter queries. We collect, in the food domain, the 100 most popular search queries among online users from a few major platforms (including Google, Bing, and Yahoo!) in March 2016. The ranking is obtained from the website http://tools.seobook.com. Each query contains on average 2.42 terms with a standard deviation of 0.70. We label these popular queries as "shorter queries," because we compare them to longer queries formed by combining them with related terms. (Note however that these queries were selected based on their popularity, not their length.) We form a vocabulary consisting of 98 words for this domain, using all the unique words that appear in these queries.

Related terms. For each shorter query q, we identify terms from the vocabulary that tend to be of interest to consumers who use query q, by leveraging data publicly available from Google Trends. For any target query, Google Trends reports a list of "related queries" that users who search the target query also tend to search for (typically 25 related queries are provided for each query). Accordingly, as a measure of whether word w is related to query q, we use whether word w is part of any "related queries" of query q on Google Trends. We construct sets of relevant words $\{G\}$ by combining each of the top 100 search queries q with all its related words. For instance, if the target query is "food network" and the related terms are "recipes" and "tv," then there are two relevant sets of words corresponding to query q: $\{$food, network, recipes$\}$ and $\{$food, network, tv$\}$. In total, we have 382 unique sets of relevant words, each containing 3.38 words on average.

Longer queries. For each set of relevant words $G = \{q, w\}$, we compare the results from the shorter query q with the results from the queries that combine word w with query q. We consider all possible ways to combine word w with query q, e.g., in the first previous example we would consider the following queries: "recipes food network," "food recipes network," and "food network recipes."

Search results. Consumers generally tend to focus on the top results when evaluating search results from a query (Narayanan and Kalyanam, 2015). Hence, for each query in our data, we collect its top 10 search results, using the Google customer search API. We use a script to automatically download the actual webpage content of these links. We record which of the words in the corresponding set G are contained in each webpage. We record whether the word appears anywhere on the actual webpage associated with the search result, not just the title and snippet provided by Google. Note that we record *whether* the word appears on the webpage associated with a result, not how many times it appears.

2.2 Evaluation Metrics

We consider two factors that may impact the value that a consumer derives from a set of search results provided by a search engine. Our first factor is whether the consumer evaluates each webpage in a *compensatory* or *non-compensatory* fashion. Under a compensatory evaluation rule, the consumer derives value from the presence of each relevant word on the page, and the presence of one word may make up for the absence of another. For the sake of illustration, we assume that all words have the same value, equal to 1. Under a non-compensatory evaluation rule, in contrast, the value of a page is a function of the entire set of relevant words present in the page. Here we consider one simple non-compensatory evaluation rule: a conjunctive rule. Under this rule, a page has value (also set to 1) only if it contains all the words in G, and it has no value otherwise.

Our second factor concerns the way in which the results provided by the search engine for a given query are combined. We consider the average value across the top results (i.e, the consumer derives value from all search results), as well as the maximum value (i.e., the consumer derives value from the best search result only). Research has shown that users tend to click only on a few search results (Yoganarasimhan, 2015). This suggests that the maximum value across search results may be the more relevant metric, but we consider both metrics nonetheless.

Combining these two factors gives us four evaluation metrics: Compensatory-Average, Compensatory-Max, Conjunctive-Average, and Conjunctive-Max.

2.3 Results

Table 1 reports the proportion of times the shorter query q performs at least as well as all longer queries corresponding to the set G. We report the average across all sets, and across sets of sizes $|G| = 3, 4, 5$, for each of the four evaluation metrics. We see that when queries are evaluated based on the best search results (which is more consistent with the empirical evidence referenced above), it is possible to achieve the same performance with the shorter query as with any of the longer queries, in 61% of the cases. Across all scenarios studied here, we see that the benefits from using the shorter queries tend to be higher when the number of relevant words is greater, the consumer cares about the maximum value across results rather than the average, and the consumer has a compensatory utility function rather than a conjunctive utility function. We also see that performance is very similar under the "Compensatory-Max" and "Conjunctive-Max" metrics, because the best queries and corresponding search results are usually the same under both metrics.

[Insert Table 1 Here]

In order to provide more intuition for these results, Table 2 displays the candidate queries for two sets of relevant words as examples, along with their activation probabilities and performance metrics. In our context, activation probabilities are the probabilities of finding specific words in the top results of a search engine, given a specific search query. In the first example, a consumer interested in recipes from the food network magazine would be weakly better off simply using the query "food network magazine" rather than a longer query that includes "recipes," even without taking into account the general preference for shorter queries. This is because 70% of the results from the shorter query already contain the term "recipes." Adding "recipes" into the query increases this proportion to 80%, at the expense of the other terms "network" and "magazine," i.e., the results may contain recipes that do not come from the food network magazine. All queries retrieve at least one page that contains all the relevant terms, i.e., they all perform the same on the "Compensatory-Max" and "Conjunctive-Max" metrics. The shorter query performs slightly better on the two metrics based on the average across search results. In the other example, a consumer who is looking for information on Kitchenaid and/or Cuisinart food processors, may be better off using the query "Kitchenaid food processor" rather than a longer query that includes "Cuisinart." This is because Cuisinart food processors are often compared to Kitchenaid, and the shorter query will retrieve at least one search result that contains all relevant terms. Including "Cuisinart" into the query greatly increases the proportion of results that contain this term, but this comes at the expense of other terms, in particular "Kitchenaid." The shorter query already performs better under our simple implementation of the "Compensatory-Average" metric that weighs all terms equally, but the difference would be more pronounced if the consumer were primarily interested in Kitchenaid over Cuisinart. In that case, omitting "Cuisinart" from the query would ensure that most results mention "Kitchenaid," while some results still mention "Cuisinart."

[Insert Table 2 Here]

In sum, our field data suggest that consumers indeed stand to benefit from being strategic in query formation, as shorter queries may be at least as effective at retrieving desired content, compared to queries that contain all the terms the consumer is searching for. The fact that consumers tend to prefer shorter queries (Jansen et al., 2000; Spink et al., 2001) makes these shorter queries even more attractive.

3 Experiment

The previous section provided indirect, correlational evidence from the field that is consistent with strategic query formation. Recall that our shorter queries were selected as the 100 most popular search queries in the food domain. The fact that these queries are indeed popular among users is consistent with users actually being strategic when formulating queries. However, this evidence is purely correlational, and this does not prove that consumers indeed have the ability to be strategic in query formation. Consequently, we developed and implemented an incentive-aligned experimental paradigm to directly test and measure consumers' ability to leverage activation probabilities.

3.1 Design

We built our experimental paradigm as a "search query game," with the following specifications in mind: (i) the relevant words G should be set exogenously and provided to participants; (ii) the game should be incentive-aligned, i.e., participants' payment should be a function of the performance of their queries; (iii) the performance of a query should be independent of the particular computer on which the game is played; (iv) in order to focus exclusively on query formation, any other type of search behavior such as evaluating results and clicking on links should be excluded from the game; (v) the game should capture the essence of query formation on search engines; (vi) the game should be easy to explain to participants.

Taking the above into consideration, our search query game asks each participant to form search queries on Google to win a cash bonus. In order to test participants' ability to leverage activation probabilities, we developed a setup in which leveraging activation probabilities is financially optimal in some situations, irrespective of any cognitive cost. Each participant completed 10 independent search tasks in a random order, i.e., 10 rounds of the game. We selected nouns in the food domain again to form the sets of relevant words $\{G\}$, because this is a very common domain on which we expect all participants to have at least some knowledge. We formed the 10 overlapping sets of 3 words corresponding to each task, using 14 unique words.[1] See Table 3. In each task, the participant was asked to form a query consisting of any subset of the words in G in any order. We assumed a "Compensatory-Max" evaluation metric, consistent with the empirical finding that users tend to click on a few links only (Yoganarasimhan, 2015). We varied the values of the three

[1]We randomized the order in which the three words were displayed to participants in each task, in order to avoid any potential ordering effect.

words $(\beta_1, \beta_2, \beta_3)$ by selecting randomly (with equal probabilities) one of the following four sets of values for each of the 10 tasks and each participant: $(\$2, \$2, \$2)$, $(\$1, \$2, \$2)$, $(\$2, \$1, \$2)$, and $(\$2, \$2, \$1)$. Like in the field data, we consider the pages associated with the top 10 results of each query. The utility of each page is based on whether each word in G appears anywhere on the actual webpage or the title/snippet provided by Google, regardless of the number of times it appears. The performance score associated with each query is the utility of its best result minus the cost of the query in dollars. In our experiment, this cost was simply set to \$1 times the number of words in the query. For each participant, we chose at the end of the game the score from one of the 10 tasks randomly and paid that amount as a bonus to the participant, in addition to a \$3 show-up fee (the score per task could range from \$2 to \$5).[2]

Figure 1 shows an example in which $G = \{$milk, cheese, tea$\}$, valued respectively at \$1, \$2, and \$2. In Figure 1a, the participant is forming their query by deciding which words to use and in which order. In this case, although the words “cheese” and “tea” are worth more than the word “milk,” only “milk” has a strong association with both of the other two words. This implies that forming the one-word query “milk” may achieve the highest score. After submitting a search query, on the next page the participant is shown the url of the link with the highest score, the list of words that were found on that page, and the performance score for this task. For example, Figure 1b is the result after a participant submits the query “tea cheese” in which they pick the two words with a higher value; Figure 1c displays the result after submitting the query “milk.”

[Insert Table 3 and Figure 1 Here]

The actual instructions of the game shown to participants are displayed in the Appendix. To ensure that participants understood the instructions, they were given a short quiz after reading the instructions. Participants proceeded to the game only after having answered all quiz questions correctly. While playing the game, participants were not allowed to use any other website. We

[2] Before running the study, we obtained all the activation probabilities using the same approach as with our field data (see “Search results” in Section 2.1). We ran all queries on a single computer to ensure that the results given to participants during the game would not be dependent on the computer on which the query was run. We used these results during the game, i.e., we did not actually run any query during the game. We also re-ran these queries using different computers, and the optimal queries and results were mostly consistent.

enforced this by running the study in a lab in which we could observe and control the sites accessed by participants.

We formed the 10 sets of words so that different types of queries would be optimal across tasks. Table 3 presents the optimal queries for each set of words. There are seven tasks in which there exists one "trigger" word that can activate both other two words in the search results. For these cases, it is optimal to form a query that leverages activation probabilities by using the "trigger" word alone, irrespective of the set of word values. The words in the remaining three tasks have weaker activation probabilities with each other, and forming queries using two words is optimal in these cases. In these three tasks, different queries may be optimal based on the particular set of word values, and more than one query may be optimal for a given set of values. Note that the same word may be a "trigger" word in one task and "non-trigger" word in another task.

3.2 Results

We obtained results from N=108 participants recruited at a large university in the northeast of the United States. Table 4 summarizes the distribution of the length of participants' queries, crossed with whether the query is optimal. Participants were most likely to form queries with two words (56%), followed by one word (30%) and three (14%). Queries were more likely to be optimal conditional on having one word: 65% of the one-word queries were optimal. This suggests that at least some participants were able to leverage the activation probabilities between words to increase their scores, and were able to recognize some cases in which a single-word query was optimal. Additional evidence in support for activation probabilities being leveraged may be found by looking specifically at words that were valued at $1. Recall that with probability 0.25, all three words were valued at $2, and with probability 0.75, two words were valued at $2 and one was valued at $1. We find that in 21% of the cases in which one word was assigned a value of $1, participants formed a one-word query containing the $1 word. In these situations, the participants favored the $1 word over both $2 words, which is inconsistent with including only the most valued words in the query. In addition, when one word was assigned a value of $1, 32% of the two-word queries contained the \$1 word, which was favored over the third word valued at \$2. However, one may wonder whether this pattern of results may be the results of participants forming their queries randomly. We find that participants formed shorter queries in tasks in which the optimal query had only one word. The average query length was 1.76 when the optimal query had one word, vs. 2.01 when the optimal query had two words (*p-value*< 0.01). This is *not* consistent with participants forming their queries completely randomly.

[Insert Table 4 Here]

To further test for respondents' ability to leverage activation probabilities, we compare how frequently participants used each word when its value was \$2 versus \$1, depending on whether it was optimal to use the word. We find that among all cases in which a word was valued at \$1 and it was optimal to use this word, the word was actually used in 65.19% of the queries. This proportion dropped to 47.17% among cases in which a word was valued at \$1 and it was *not* optimal to use it. A Chi-square test reveals that these two proportions are significantly different (*p-value*< 0.01), confirming that consumers have at least some ability to leverage activation probabilities in search. However, the fact that the proportions are quite far from 100% and 0% respectively also suggests that participants did not leverage activation probabilities to their full potential. Among all cases in which a word was valued at \$2 and it was optimal to use this word, the word was used in 63.95% of the queries. This proportion dropped only slightly to 61.42% when considering cases in which a word was valued at \$2 and it was *not* optimal to use the word. The difference in proportions is not significant (*p-value*$= 0.20$). The fact that the use of \$2 words was not significantly affected by whether it was optimal to use them confirms that participants leveraged activation probabilities to some extent, but not to their full potential.

Finally, we can compare how likely the same word was to be used when it was the ”trigger” (i.e., when it could activate the other two words) vs. not. In our design, five words were used in two different tasks, and were triggers in only one of these tasks. For two out of these five words, we observe a significant increase in the probability of being included in the query when the word was a trigger vs. not (candy: 73% vs. 37%, *p-value*< 0.01; Easter: 94% vs. 69%, *p-value*< 0.01).

However there was no significant difference for two words (sugar: 43% vs. 37%, p-$value = 0.40$; tomato: 35% vs. 35%, p-$value = 1.00$), and one word was actually significantly less likely to be used when it was a trigger (cake: 57% vs. 67%, p-$value < 0.05$). This further suggests that although consumers have some ability to leverage activation probabilities, this ability is somewhat limited.

To sum up, the behavior we observe suggests that participants are able to strategically leverage activation probabilities between words, at least to some extent. Because our study uses a somewhat artificial lab setting, we do not claim that the *extent* to which consumers leverage activation probabilities in the real world is the same as in our study. Instead, we view our results as proof of existence that consumers have some ability to strategically formulate queries that contain only a subset of the terms they are interested in, but that are effective at retrieving the other terms. An analogy may be made to the experimental economics literature. Games such as the dictator game are used in this literature to show that individuals have the potential to behave in ways that are inconsistent with maximizing their own financial well-being, although these games do not quantify the extent of such behavior in real life.

4 Discussion, Implications, and Future Research

We have proposed that the queries formed by consumers are not necessarily straightforward expressions of their content preferences, but rather the outcome of a strategic attempt to leverage activation probabilities in order to retrieve relevant content. In particular, our results are consistent with the argument that consumers formulate queries that are more likely to *retrieve* the content they are searching for, rather than merely being *similar* to that content.

Our findings have implications both for practitioners and researchers. For researchers, our findings pave the way for the development of utility-based models that link consumers' content preferences to their textual queries. In particular, in today's environment search is primarily text-based, and marketing models of search should be adapted to capture this reality. We argue that utility-based models capturing text-based search should allow consumers to be strategic when formulating search queries given their preferences.

Our findings also have direct implications for practitioners engaging in Search Engine Optimization (SEO) or Search Engine Marketing (SEM). These practitioners are faced with the challenge of identifying queries and keywords on which to promote their content, i.e., on which they should make an effort to appear as a top organic result (in the case of SEO) or for which they should bid higher (in the case of SEM). These queries and keywords should reflect content preferences that are well aligned with the content the firm is trying to promote. Our results suggest that the set of queries on which a piece of content should be promoted should not be limited to queries that have more words in common with the target content, but also include queries that are more likely to retrieve the target content. In addition, our results are relevant for practitioners interested in leveraging online search queries as a source of consumer insights. Our findings suggest that the content that is of interest to consumers is not simply the content mentioned in their search queries, but also the content retrieved by their search queries.

We close by highlighting additional areas for future research. First, the extent to which consumers leverage activation probabilities may be compared across types of search, including voice search (e.g., Siri and OK Google). Second, future research may study the impact of query auto-completion (i.e., the search engines auto-completes the user's queries) on the link between content preferences and query formation. Auto-complete suggestions reflect co-occurrence of words in queries across all users. In contrast, leveraging activation probabilities between queries and results allows consumers to improve the quality of the search results given their *own* particular content preferences. Hence, from the perspective of consumers, auto-completion should not be a substitute for leveraging activation probabilities between queries and search results.

References

Borrell (2016). Forecast says seo-related spending will be worth $80 billion by 2020. *Available online from http://searchengineland.com/forecast-says-seo-related-spending-will-worth-80-billion-2020-247712*.

Gabaix, X., D. Laibson, G. Moloche, and S. Weinberg (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review 96*(4), 1043–1068.

InternetLiveStats (2016). Google search statistics. *Available online from http://www.internetlivestats.com/google-search-statistics/*.

Jansen, B. J., A. Spink, and T. Saracevic (2000). Real life, real users, and real needs: a study and analysis of user queries on the web. *Information processing & management 36*(2), 207–227.

Jeziorski, P. and I. Segal (2015). What makes them click: Empirical analysis of consumer demand for search advertising. *American Economic Journal: Microeconomics 7*(3), 24–53.

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2010). Online demand under limited consumer search. *Marketing Science 29*(6), 1001–1023.

Li, H. and J. Xu (2013). Semantic matching in search. *Foundation and Trends in Informational Retrieval 7*(5), 343–469.

Manning, C. D., P. Raghavan, and H. Schütze (2008). *Introduction to information retrieval*, Volume 1. Cambridge university press Cambridge.

Narayanan, S. and K. Kalyanam (2015). Position effects in search advertising and their moderators: A regression discontinuity approach. *Marketing Science 34*(3), 388–407.

Pirolli, P. L. (2007). *Information foraging theory: Adaptive interaction with information*. Oxford University Press.

Ruthven, I. (2003). Re-examining the potential effectiveness of interactive query expansion. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pp. 213–220. ACM.

Rutz, O. J. and M. Trusov (2011). Zooming in on paid search ads?a consumer-level model calibrated on aggregated data. *Marketing Science 30*(5), 789–800.

Santos, R. L., C. Macdonald, and I. Ounis (2015). Search result diversification. *Foundations and Trends in Information Retrieval 9*(1), 1–90.

Spink, A., D. Wolfram, M. B. Jansen, and T. Saracevic (2001). Searching the web: The public and their queries. *Journal of the American society for information science and technology 52*(3), 226–234.

Stigler, G. J. (1961). The economics of information. *Journal of political economy 69*(3), 213–225.

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica: Journal of the Econometric Society*, 641–654.

Yoganarasimhan, H. (2015). Search personalization using machine learning. *Available at SSRN 2590020*.

Tables

Table 1: Benefits of Leveraging Activation Probabilities

Metric		$\lvert G \rvert$	Prob (shorter query $\geqslant$ all longer queries)
Compensatory	Average	All	0.175
		3	0.142
		4	0.250
		5	0.296
	Max	All	0.610
		3	0.558
		4	0.720
		5	0.593
Conjunctive	Average	All	0.099
		3	0.067
		4	0.140
		5	0.296
	Max	All	0.613
		3	0.558
		4	0.730
		5	0.593

Note: "Shorter query" refers to a query q that is among the 100 most popular queries in the food domain (according to seobook.com). "Longer queries" refer to queries that contain an additional word w which is relevant to query q (as indicated by Google Trends). Queries are evaluated based on their ability to retrieve the words in $G = \{q, w\}$. Compensatory: the consumer derives value from the presence of each word on a page. Conjunctive: a page has value only if it contains all words in G. Average: the consumer derives value from all search results. Maximum: the consumer derives value from the best search result only.

Table 2: Examples of Activation Probabilities and Evaluation Metrics

	Prob(w_1)	Prob(w_2)	Prob(w_3)	Prob(w_4)	Compensatory Avg.	Compensatory Max	Conjunctive Avg.	Conjunctive Max
$G_1 = \{$"food network magazine," "recipes"$\}$	food	network	magazine	recipes				
food network magazine	1.0	1.0	0.9	0.7	3.6	4	0.7	1
food network magazine recipes	1.0	1.0	0.7	0.7	3.4	4	0.6	1
food network recipes magazine	1.0	0.9	0.7	0.8	3.4	4	0.6	1
food recipes network magazine	1.0	0.9	0.6	0.7	3.2	4	0.5	1
recipes food network magazine	1.0	0.9	0.7	0.8	3.4	4	0.6	1
$G_2 = \{$"kitchenaid food processor," "cuisinart"$\}$	kitchenaid	food	processor	cuisinart				
kitchenaid food processor	0.8	1.0	0.9	0.1	2.8	4	0.1	1
kitchenaid food cuisinart processor	0.5	0.8	0.7	0.4	2.4	4	0.2	1
kitchenaid cuisinart food processor	0.5	0.9	0.8	0.5	2.7	4	0.3	1
cuisinart kitchenaid food processor	0.5	0.9	0.8	0.5	2.7	4	0.3	1
kitchenaid food processor cuisinart	0.5	0.9	0.8	0.5	2.7	4	0.3	1

Note: For each query, the first four or five columns report its activation probabilities to each relevant word, and the last four columns report its performance on each evaluation metric. Compensatory: the consumer derives value from the presence of each word on a page. Conjunctive: a page has value only if it contains all words in G. Avg.: the consumer derives value from all search results. Maximum: the consumer derives value from the best search result only.

Table 3: Search Tasks in Experiment and Optimal Queries

Task	w_1	w_2	w_3	Optimal Query
1	candy	caffeine	sugar	"candy"
2	fish	tea	tomato	two words*
3	milk	cheese	tea	"milk"
4	Easter	candy	egg	"Easter"
5	tomato	drink	pizza	"tomato"
6	Easter	caffeine	ketchup	two words*
7	sugar	cake	pizza	"sugar"
8	egg	candy	drink	two words*
9	cake	cheese	Easter	"cake"
10	ketchup	cake	tomato	"ketchup"

Note: For the seven tasks in which the optimal query has a single word, this trigger word is labeled as w_1. In the study, words were always shown to participants in a random order. * indicates that the optimal query depends on the value assigned to each word.

Table 4: Number of Queries with Different Lengths and Optimality in Experiment

Query Length	Not Optimal	Optimal	Row Total	Percentage
1	180	149	229	30%
2	498	106	604	56%
3	147	0	147	14%
Column Total	825	255	1,080	
Percentage	76%	24%		100%

Figures

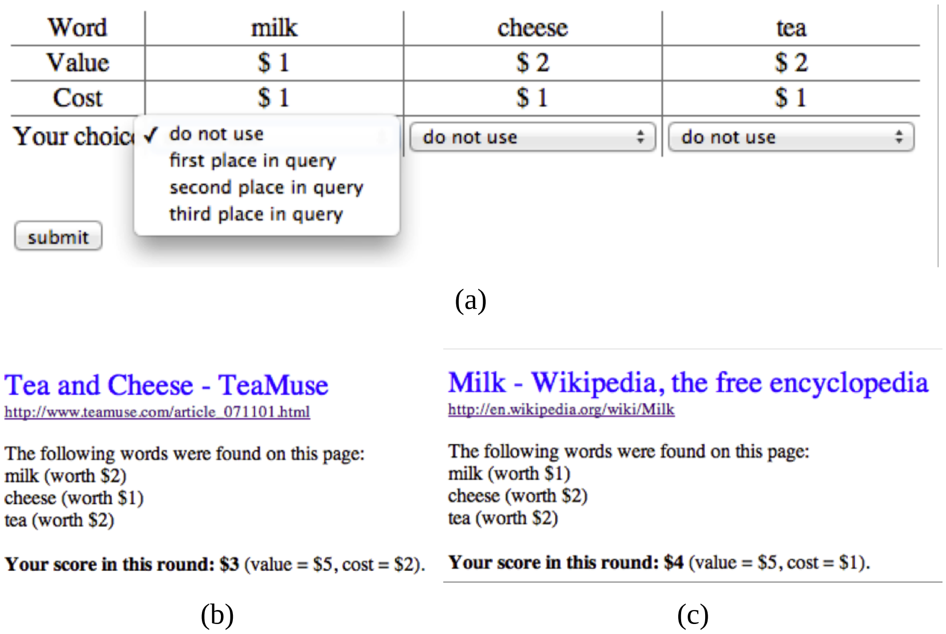

Figure 1: Search Query Game Interface in Experiment

Figure (a) is the game interface where a participant forms a search query given the set of words, their values ($1 or $2 per word) and costs ($1 per word). The participant decides which word(s) to use and in which order. Figure (b) and (c) show the screens the participant will see after submitting the queries "tea cheese" (b) and "milk" (c). The participant is shown the search result with the highest score (score=value-cost), the list of relevant words found on its webpage, and the corresponding score.

Appendix: Instruction Page for Search Query Game

Welcome to our Search Query Game!

Overview
You will be playing a web **search query game**. You will submit **search queries** and you will score **points** based on the **search results** from these queries. After you finish the game, you will fill out a short survey. The whole study will take you about 20 minutes.

Rules of the Game
You will play the game for **10 rounds**. In each round, you will be given **3 words** to search for. Each word has some **value** ($1 or $2) if you find it. In order to find these words you will form a **search query** that contains any of these 3 words in any order. Each word also **costs $1** if you use it in your query. That is, the fewer words you use in your query, the lower its cost.

Your task is to decide which words to use in your query and in which order.

For example, suppose the following 3 words are given to you in a round:

Word	fruit	salad	chicken
Value	$2	$1	$2
Cost	$1	$1	$1

You may decide to only use ‘fruit’, i.e., your query is ‘fruit’ and it costs $1; or you may decide to use ‘fruit’ first followed by ‘salad,’ i.e., your query is ‘fruit salad’ and it costs $2.

After you submit your search query, in the background we will automatically **run this query on Google and scan the webpages associated with the top 10 links**. Each link has a different value **based on which of the 3 words it contains.**

Your score in each round is the value of the best link minus the cost of your search query:
Score = Value of best link – Cost of the query

In other words, **the best queries are those that are short but allow you to find many of the words you are looking for.**

In the above example, suppose your query is ‘chicken fruit’ and we find that the best link contains the words ‘chicken’ and ‘fruit’ but not 'salad.' Then your score is $2 ($2 for 'chicken' plus $2 for 'fruit' minus $2 because there are 2 words in the query). Suppose now that your query is ‘chicken salad’ and we find that the best link contains the words ‘chicken’ and ‘fruit’ and 'salad.' Then your score is $3 ($2 for 'chicken' plus $1 for 'salad' plus $2 for 'fruit' minus $2 because there are 2 words in the query).

Your Final Bonus
At the end of the study, one round will be selected randomly. As a bonus, you will receive your score in this round in cash. For example, if your score is $4, you will receive a $4 bonus. This bonus will be in addition to your show up fee.

Note:

- You score points as long as the word appears anywhere on the actual webpage associated with the link, not just the url or the blurb given by Google.
- It doesn’t matter how many times each word appears on the webpage associated with a link, as long as it appears at least once.
- Your score cannot be negative, that is, you cannot loose money.
- You are not allowed to use any other web site while playing the game. Doing so will exclude you from the study.