When More Is Less: The Paradox of Choice in Search Engine Use

Antti Oulasvirta, Janne Hukkinen
Helsinki Institute for Information Technology HIIT
Helsinki University of Technology TKK

Barry Schwartz
Psychology Department
Swarthmore College

ABSTRACT

In numerous everyday domains, it has been demonstrated that increasing the number of options beyond a handful can lead to paralysis and poor choice and decrease satisfaction with the choice. Were this so-called *paradox of choice* to hold in search engine use, it would mean that increasing recall can actually work counter to user satisfaction if it implies choice from a more extensive set of result items. The existence of this effect was demonstrated in an experiment where users (N=24) were shown a search scenario and a query and were required to choose the best result item within 30 seconds. Having to choose from six results yielded both higher subjective satisfaction with the choice and greater confidence in its correctness than when there were 24 items on the results page. We discuss this finding in the wider context of "choice architecture"—that is, how result presentation affects choice and satisfaction.

Categories and Subject Descriptors

H.1.2 [**User/Machine Systems**]: Human information processing. H.3.3 [**Information Search and Retrieval**]: Information filtering.

General Terms

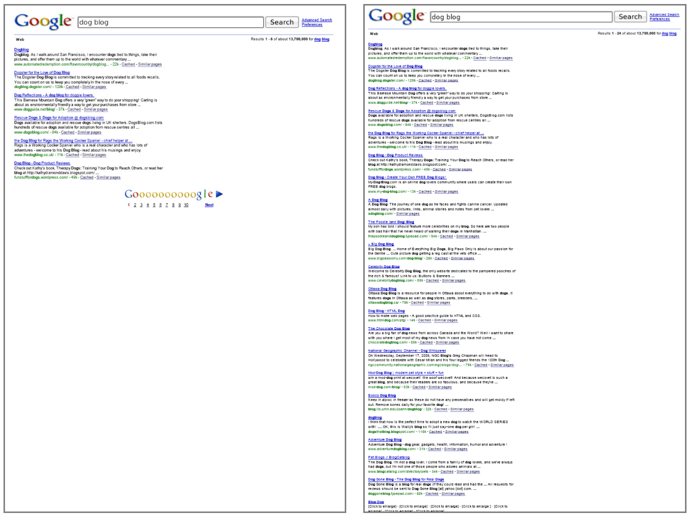
Design, Human Factors.

Keywords

Search engines, relevance judgments, satisfaction, user interfaces.

1. INTRODUCTION

If you type in your favorite pop singer's name to Google, you will be presented with a result set of possibly millions of items. Items within a single page may have perceivable differences, yet the better the engine has done its job, the greater the number of items that will appear relevant. In such a case, *can* you be content with the link you finally choose, given that you could not consider even an iota of the full number of results available? At the time of writing, Google offered 99,500,000 results for the query "Britney Spears." The situation is not that different from what Westerners face daily in the offline domain: massive choice. For example, wanting to buy breakfast cereal at a grocery store forces a choice from among some 273 products [28].



Results **1** - **6** of about **13,700,000** for dog blog.

Dogblog
Dogblog. As I walk around San Francisco, I encounter **dogs** tied to things, take their pictures, and offer them up to the world with whatever commentary ...
www.automatedredemption.com/flavorcountry/dogblog... - 22k - Cached - Similar pages

Dogster for the Love of Dog Blog
The Dogster **Dog Blog** is committed to tracking every story related to all foods recalls. You can count on us to keep you completely in the nose of every ...
dogblog.dogster.com/ - 129k - Cached - Similar pages

Dog Reflections - A dog blog for doggie lovers.
This Bernese Mountain **Dog** offers a very "green" way to do your shopping! Carting is about as environmentally friendly a way to get your purchases from store ...
www.dogguide.net/blog/ - 37k - Cached - Similar pages

Rescue Dogs & Dogs for Adoption @ dogsblog.com
Dogs available for adoption and rescue **dogs** living in UK shelters. DogsBlog.com lists hundreds of rescue **dogs** available for adoption from rescue centres all ...
www.dogsblog.com/ - 84k - Cached - Similar pages

the Dog Blog for Rags the Working Cocker Spaniel - chief helper at ...
Rags is a Working Cocker Spaniel who is a real character and who has lots of adventures - welcome to his **Dog Blog** - read about his musings and enjoy.
www.thedogblog.co.uk/ - 11k - Cached - Similar pages

Dog Blog - Dog Product Reviews
Check out Kathy's book, Therapy **Dogs**: Training Your **Dog** to Reach Others, or read her **blog** at http://kathydiamonddavis.blogspot.com/ ...
funstuffordogs.wordpress.com/ - 49k - Cached - Similar pages

Results **1** - **24** of about **13,700,000** for dog blog.

Dogblog
Dogblog. As I walk around San Francisco, I encounter **dogs** tied to things, take their pictures, and offer them up to the world with whatever commentary ...
www.automatedredemption.com/flavorcountry/dogblog... - 22k - Cached - Similar pages

Dogster for the Love of Dog Blog
The Dogster **Dog Blog** is committed to tracking every story related to all foods recalls. You can count on us to keep you completely in the nose of every ...
dogblog.dogster.com/ - 129k - Cached - Similar pages

Dog Reflections - A dog blog for doggie lovers.
This Bernese Mountain **Dog** offers a very "green" way to do your shopping! Carting is about as environmentally friendly a way to get your purchases from store ...
www.dogguide.net/blog/ - 37k - Cached - Similar pages

Rescue Dogs & Dogs for Adoption @ dogsblog.com
Dogs available for adoption and rescue **dogs** living in UK shelters. DogsBlog.com lists hundreds of rescue **dogs** available for adoption from rescue centres all ...
www.dogsblog.com/ - 84k - Cached - Similar pages

the Dog Blog for Rags the Working Cocker Spaniel - chief helper at ...
Rags is a Working Cocker Spaniel who is a real character and who has lots of adventures - welcome to his **Dog Blog** - read about his musings and enjoy.
www.thedogblog.co.uk/ - 11k - Cached - Similar pages

Dog Blog - Dog Product Reviews
Check out Kathy's book, Therapy **Dogs**: Training Your **Dog** to Reach Others, or read her **blog** at http://kathydiamonddavis.blogspot.com/ ...
funstuffordogs.wordpress.com/ - 49k - Cached - Similar pages

Dog Blog - Create Your Own FREE Dog Blogs!
My-**Dog-Blog**.com is an online **dog** lovers community where users can create their own FREE **dog** blogs.
www.my-dog-blog.com/ - 13k - Cached - Similar pages

A Dog Blog
A **Dog Blog**: The journey of one **dog** as he faces and fights canine cancer. Updated almost daily with pictures, links, animal stories and notes from pet lovers ...
adogblog.com/ - Similar pages

The Poodle (and Dog) Blog
My son has told I should feature more celebrities on my **blog**. So here are two people with bad hair that I've never heard of walking their **dogs** in Manhattan. ...
thepoodleanddogblog.typepad.com/ - 54k - Cached - Similar pages

» Big Dog Blog
Big **Dog Blog**. ... Home of Everything Big **Dogs**, Big Paws Only is about our passion for the Gentle ... Cute picture **dog** getting a leg cast at the vets office ...
www.bigpawsonly.com/dog-blog/ - 26k - Cached - Similar pages

Celebrity Dog Blog
Welcome to Celebrity **Dog Blog**, the only website dedicated to the pampered pooches of the rich & famous! Link to us: Buttons & Banners ...
www.celebritydogblog.com/ - 55k - Cached - Similar pages

Ottawa Dog Blog
Ottawa **Dog Blog** is a resource for people in Ottawa about everything to do with **dogs**. It features **dogs** in Ottawa as well as **dog** stores, parks, breeders, ...
ottawadogblog.ca/ - 78k - Cached - Similar pages

Dog Blog - HTML Dog
How to make web pages - A good practice guide to HTML and CSS.
www.htmldog.com/ptg/ - 14k - Cached - Similar pages

The Chocolate Dog Blog
Are you a big fan of **dog** news from across Canada and the World? Well I want to share with you where I get most of my **dog** news from in case you have not come ...
chocolatedogblog.com/ - 55k - Cached - Similar pages

National Geographic Channel - Dog Whisperer
On Wednesday, September 17, 2008, NGC **Blog's** Greg Chapman will head to Hollywood to celebrate with Cesar Milan and his four legged friends the 100th **Dog** ...
ngccommunity.nationalgeographic.com/ngcblogs/dog-... - 75k - Cached - Similar pages

Mod-Dog Blog | modern pet style + stuff + fun
win a mod-**dog** print at wecovet! We woof wecovet! And because wecovet is such a great **blog**, and because their readers are so fabulous, and because they've ...
mod-dog.com/blog/ - 53k - Cached - Similar pages

Bosco Dog Blog
Keep in ziploc in freezer as these do not have any preservatives and will get moldy if left out. Remove bones daily for your favorite **dog**! ...
blog.lib.umn.edu/joanh/dogblog/ - 32k - Cached - Similar pages

dogblog
I think that now is the perfect time to adopt a new **dog** to watch the WORLD SERIES with! .... OK, this is Wally's **blog** so I'll just say--one **dog** per girl! ...
dogsthatblog.blogspot.com/ - 116k - Cached - Similar pages

Adventure Dog Blog
Adventure **Dog Blog** - **dog** gear, gadgets, health, information, humor and adventure !
www.adventuredogblog.com/ - 31k - Cached - Similar pages

Pet Blogs // BlogCatalog
The **Dog Blog**. I'm not a **dog** lover. I come from a family of **dog** lovers, and we've always had **dogs**, but I'm not one of those people who adores animals at ...
www.blogcatalog.com/directory/pets - 34k - Cached - Similar pages

Dog Gone Blog - The Dog Blog for Real Dogs
Dog Gone Blog is a **blog** for real **dogs** (if they could read and had the ... All requests for reviews should be sent to **Dog** Gone **Blog** [at] yahoo [dot] com. ...
doggoneblog.typepad.com/ - 80k - Cached - Similar pages

Blog Dog
[Click to enlarge] · [Click to enlarge] · [Click to enlarge] · [Click to enlarge] · [Click to enlarge] · [Click to enlarge] · [Click to enlarge]

Figure 1: ***Does it matter how many search results are presented? Six-item (left) versus 24-item (right) result listings in Google, materials used in the experiment. Note: In the 24-item list, the final three items are shown on a second page.***

Recent research in cognitive psychology has revealed an interesting effect of choice overload:

> **The paradox of choice**: providing more options—particularly if they are highly relevant and success is personally important—will lead to poorer choice and degrade satisfaction [28].

Experimental demonstrations of this paradox are quite compelling and bespeak its generality. For example, passersby are more likely to buy jams on display, and more satisfied as customers when there are six jams to choose from than 24 [15]. University students are more likely to write an extra-credit essay, and write better essays, when they have six topics to choose from than 30 [15]. Employees are more likely to participate in 401(k) retirement plans when there are two rather than 59 funds to choose from [16]. But would the same apply to Google with, say, six versus 24 items? Figure 1 illustrates the situation.

The existence of this phenomenon could have important implications for how we think about search engine use. One presumption has been that *if* the user has the persistence to go through the result set, or a sufficient part of it, a larger number of items on the list indicates greater likelihood that (s)he has encountered an item of higher relevance as the end is reached. Ergo, the more results, the higher the effectiveness. If this assumption turns out to be questionable, we can ask whether search engines should be less like slavish "reporters" and more akin to personal assistants who guide customers to the most reasonable options in a store.

However, anyone can imagine a number of reasons for the effect not appearing in search engine use. For example, if users jump to

SIGIR'09, July 19–23, 2009, Boston, Massachusetts, USA.
Copyright 2009 ACM 978-1-60558-483-6/09/07...$5.00.

the first-ranked items (e.g., [10][12][24][26][31]), they may ignore others, no matter how many there are left. Or, if users are effective in using cues such as snippets (e.g., [9][10]), they may narrow down their options to one or two candidates.

This paper presents the first experiment to directly test the existence of this paradox in information retrieval. We focus on search engines as a major category of present-day information retrieval, and Google in particular. The experimental paradigm to study the paradox in consumer choice (e.g., [15]) was "translated" into the context of search engine use. Participants were given a realistic search task (three types) and asked to mark from a search result list the item that best answers the question, within 30 seconds. In half of the trials, 24 items were presented, and six were used in the rest, a situation similar to that depicted in Figure 1. After choosing the best item (and without seeing the actual page), the participants then evaluated their choice in four dimensions, including satisfaction and confidence. Two search engine layouts were used: Google (with some cues removed) and a fake engine ("RCD4000"). The results follow the pattern predicted according to the paradox: When provided with only six items to choose from, users are more satisfied with their choice, more confident that it is pertinent to the task, and more likely to think they were more careful. To conclude the paper, we frame the discussion with the notion of "architecture of choice."

2. CHOICE AMONG RESULTS

In this section, we discuss *why* the paradox of choice would emerge in search engine use. Before we present the hypothesis, let us first describe the choice situation in more concrete terms. First, *objective descriptors of the number of result items* must distinguish among:

- **Result set**, the set of all search result items (total "hits"). The higher the recall, the larger the result set.
- **Presentation set**, the total set of items the user can view. Normally, the presentation set is the result set.
- **Page set**, the set of items presented on an individual page.

Second, *subjective* descriptors must distinguish the following:

- **Perceived set**, the set of items perceived by the user during choice, including those glanced at very quickly.
- **Consideration set**, the set of alternatives the user considers in making a choice.
- **Remembered set**, the set of items remembered after the choice (i.e., without the user seeing the result page).

2.1 Hypothesis

The paradox of choice refers to the effect of increasing the consideration set size, not the presentation set size. Given that most searches produce more than a handful of items, it is more rule than exception that users cannot consider presentation sets exhaustively but must narrow them down. However, since this variable sets a ceiling to consideration set size, the effect is likely to be associated with it.

The paradox manifests itself as an inverse U-shaped relationship between consideration set and subjective satisfaction with choice [30]. Having a few options is often not enough, and is thus associated with lower satisfaction; then there is an intermediate range where subjective satisfaction is higher (say, 4–10 items); and the final part shows decreasing satisfaction as the number of options increases. In the present study, we are not interested in charting the whole continuum but focus on a sample of two points, one from the middle and the other from the end portion.

According to a recent theory, the paradox of choice is caused not by a single factor but by the interplay of many [28]. These can be broken down into three chronologically ordered phases.

2.1.1 *Phase 1: Attraction*

When one first sees a result page, a large presentation or page set size can actually bear two positive effects:

1. **Increased attraction** to a page: Seeing *more* items on a display increases its perceived attractiveness and may make the user start scanning it (e.g., [15]).

2. **Increased expectations**: Seeing that there are more items available increases the expectation that an excellent item will be found [28]. You are supposed to get a perfect answer when you have many options to choose from.

2.1.2 *Phase 2: Choice*

However, negative effects emerge when the user enters the phase in which the choice is made. According to the theory, increasing consideration set size can:

3. **Paralyze** the user in the process of entertaining alternatives. If finding the best item is linear in N, the total decision time for 24 items is four times greater than with six items. If choice requires pair-wise comparisons, N(N-1)/2 comparisons are needed. For six items, 15 comparisons are needed, but for 24 one needs 276—over 18 times more!

4. Result in **poorer choice**: As a result of the choice task being more difficult, and paralysis, users with more options are more prone to choosing suboptimally.

2.1.3 *Phase 3: Evaluation*

Increasing remembered set size can do the following in relation to evaluation of the chosen item:

5. Cause **dissatisfaction** by creating a discrepancy with expectations: With higher expectations, the item ultimately selected will be perceived as further away from the standard. This effect is boosted by an effect predicted by prospect theory [19]: the felt magnitude of discount caused by an item *not* meeting the expectation is larger than the felt magnitude of gain caused by an item surpassing the expectation by the same amount [28].

6. Cause **regret** by increasing the perceived opportunity cost: With more and better options to choose from, the ones that were *not* chosen are also better. Since users are likely to not be able to recall the options they considered, they may amalgamate items in the remembered set to form a "super option," overly emphasizing positive features [28].

2.2 Alternative accounts

Research on information retrieval has addressed two problems that are closely related to the paradox of choice: 1) how many items should be presented and 2) in which order. Roughly speaking, there are two categories of theories. Both predict *diminishing* returns for increasing presentation set size, yet they do not predict *negative* returns as the paradox of choice.

The first category predicts a *decrease* in subjective relevance as one advances in a list. For example, Brookes [6] proposed that perceived utility of a document list should decrease in going through the result list, because the more documents one has seen, the lower the informational value of each new item will be.

The second category predicts that judgments of relevance will *fluctuate* between positive and negative according to a standard that is held and updated in mind as one traverses the list. General theories from cognitive psychology have been referred to in discussions of possible explanations for order effects in the use of information retrieval results [14]. For example, the social psychologist Asch [1] was the first to propose that reordering items in a list could change how they are judged (see also [23]). The explanation was that initial appraisal (or impression) would shape the judgment of subsequently processed items. Hogarth and Einhorn [13] proposed an anchoring-adjustment hypothesis: people hold a primary belief (anchor) that guides the processing of the subsequent items in the list. This anchor is continually adjusted with new conflicting or complementary information. Clancy and Wachsler [8] discussed a phenomenon that would lead to inconsistent and inaccurate judgments toward the end of a list: fatigue. As people process more and more items, they would be too fatigued to make proper judgments toward the end of the list.

Both theories entail, in the case where the user has the persistence to go through the results, greater likelihood as the end is reached of the user having encountered an item of higher relevance as the number of items in the list increases. Hence, a larger presentation set should yield better choice, although with high costs. We suspect that this kind of logic may underlie the intuitive appeal of making the presentation set as large as the result set.

Why is it that the paradox of choice points to an opposite prediction? Several reasons can be identified. The experimental paradigm for studying order effects has been based on *a step-by-step relevance judgment task*, where the participant is to provide a relevance rating to all documents presented one at a time [14]. However, in real-life choice, people think about not "relevance" but the suitability of an item for a goal or action. Therefore, the assessment of options is often multidimensional. The consequence of multidimensional choice is that the user can be overwhelmed with a comparison of only two items, if the number of choice-relevant dimensions is large. Anyone who has bought a PC or car knows this issue. Moreover, each item examined can reveal new aspects that were not previously considered, which may require revisiting the earlier items (see also [20]). As users may be unable to keep in their minds all intermediate results of comparisons, their search is often not linear. Indeed, eye-movement studies of search use have shown that users do not examine results in a linear fashion but go back and forth between items (e.g., [24]).

Interestingly, the only two studies known to us that vary presentation set size in conjunction with order effect provide tentative support for our prediction. The original study reported by Eisenberg and Barry [11] showed that when documents are reordered according to relevance (low to high), users overestimate the importance of the documents. More interestingly, in a later study, Parker and Johnson [27] showed that order effects do not appear with fewer than 15 documents. Huang and Wang [14] found that with only five items, there are no order effects, whereas with more items (15–75) there are. These studies utilized the step-by-step paradigm, in which the users were, however, able to see all items on a page. To sum up, when the presentation set has been varied, the order effect disappears with small set size.

3. METHOD

The study is based on a direct "translation" of the experimental paradigm used to study the paradox of choice. In a nutshell, the core choice task consists of three pages presented to a participant: Page 1, a search scenario/task and the associated search query; Page 2, the result page; and Page 3, evaluation. On Page 2, participants have to mark the item ("click") that best corresponds to the scenario. They do not need to explain their choice or provide an answer to the question. Three kinds of tasks were used: simple fact-finding, problem-solving, and subjective opinions. To better simulate search engine use, which tends to be rapid [9][12], we guided the users to make their choice within 30 seconds. The critical manipulation was the number of results presented: in half of the trials, 24 items were presented and in the other half only six (Figure 1). To ensure that the results did not reflect a familiarity/preference effect, fake search engine RCD4000 (Figure 2) was used in addition to a Google replica. Had we obtained a negative result (no effect), it could have been due to users being skilled with Google no matter how many results are presented. Moreover, including two engines, familiar vs. unfamiliar, gave us a way to assess the effect's generalizability.

Paper form was used for technical difficulties in replicating the performance of Google in RCD4000 in real-time. However, previous studies have used paper form as well [14], and there is no effect reported that casts into doubt the validity of paper form.

3.1 Participants

With three exceptions who were recruited via personal networks, all 25 participants (13 M, 11 F) were recruited on-site on two university campuses in Helsinki. The participants were between 19 and 28 years of age, with an average of 22.1. They had the following majors: 6 physics, 5 chemistry, 4 geography, 2 humanities, 1 mathematics, 1 computer science, 1 biology, 1 biochemistry, 1 astronomy, 1 craft science, 1 agricultural politics. One participant was excluded because of unwillingness to follow instructions.

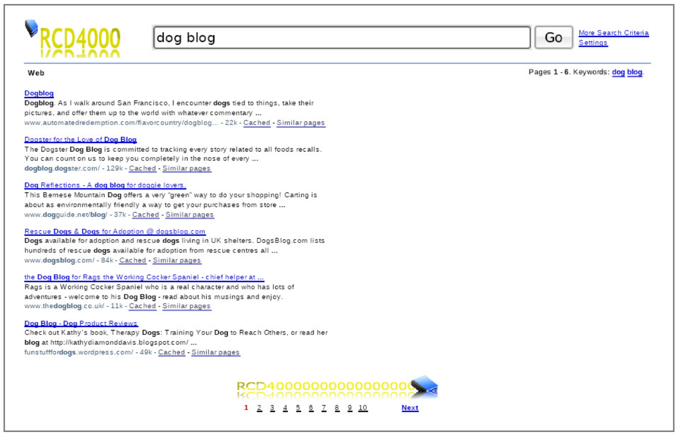
No compensation was given other than the option of obtaining one's own data later. All participants were native speakers of Finnish, the language used in the study materials.

3.2 Experimental design

The experiment followed a 2 (number of result items: 6 vs. 24) x 2 (search engine: Google vs. RCD4000) within-subjects design. The 24 search tasks were divided into four blocks of six tasks each, each block with the same engine but with the number of result items alternating. For counter-balancing, the order of these blocks was rotated over the set of participants. The dependent variables are reported upon below in subsection 3.3.4.

3.3 Materials

The set of papers handed out to a participant consisted of 99 A4 pages in total, including the following: 1) welcome, 2) task introduction, 3) practice, 4) preparation for the experiment, 5) blocks of tasks and questionnaires, 6) post-block questionnaires, 7) post-experiment questionnaire.



Pages **1** - **6**. Keywords: **dog** **blog**.

Dogblog
Dogblog. As I walk around San Francisco, I encounter **dogs** tied to things, take their pictures, and offer them up to the world with whatever commentary ...
www.automatedredemption.com/flavorcountry/dogblog... - 22k - Cached - Similar pages

Dogster for the Love of **Dog Blog**
The Dogster **Dog Blog** is committed to tracking every story related to all foods recalls. You can count on us to keep you completely in the nose of every ...
dogblog.**dog**ster.com/ - 129k - Cached - Similar pages

Dog Reflections - A **dog blog** for doggie lovers.
This Bemese Mountain **Dog** offers a very “green” way to do your shopping! Carting is about as environmentally friendly a way to get your purchases from store ...
www.**dog**guide.net/**blog**/ - 37k - Cached - Similar pages

Rescue **Dogs** & **Dogs** for Adoption @ dogsblog.com
Dogs available for adoption and rescue **dogs** living in UK shelters. DogsBlog.com lists hundreds of rescue **dogs** available for adoption from rescue centres all ...
www.**dogsblog**.com/ - 84k - Cached - Similar pages

the **Dog Blog** for Rags the Working Cocker Spaniel - chief helper at ...
Rags is a Working Cocker Spaniel who is a real character and who has lots of adventures - welcome to his **Dog Blog** - read about his musings and enjoy.
www.the**dogblog**.co.uk/ - 11k - Cached - Similar pages

Dog Blog - **Dog** Product Reviews
Check out Kathy’s book, Therapy **Dogs**: Training Your **Dog** to Reach Others, or read her **blog** at http://kathydiamonddavis.blogspot.com/ ...
funstufffor**dogs**.wordpress.com/ - 49k - Cached - Similar pages

RCD4000000000000000000

1 2 3 4 5 6 7 8 9 10

Figure 2: ***An additional fake search engine was used to test for effects of familiarity or expectations. The same content was used as in Google, but layout and terminology differed.***

3.3.1 Search tasks

Each search task consisted of the task page, the search results page(s), and a satisfaction questionnaire. There were three different types of search tasks, each with two topics:

1. **Simple facts**, with a short and unambiguous answer (a factual claim). Topics: sports, geography. Examples: 1) “Find out which country is located at the highest altitude,” query “country highest altitude”; 2) “Find out which Olympic-record-holding athletes were busted for doping in the 2008 Olympic Games” using the query “doping beijing OR 2008.”
2. **Problems**, with ambiguous answers. These require understanding a mechanism or an open set of causal criteria. Topics: Reasons behind international conflicts, price structure of goods. Examples: 1) “What determines the cost of railway tickets in Europe,” query: “railway ticket price formation Europe”; 2) “Why are Russia and Georgia at war?” with the query: “russia georgia war.”
3. **Preferences**, addressing subjective opinions. Topics: Where would you like to buy X online? What is the most interesting instance of X? Examples: 1) “Find a promising pizzeria in [city],” query: “pizza [city name]”; 2) “Find your favorite Swedish novelist’s homepage,” query: “Swedish writer.”

3.3.2 Search results

All search result listings were generated via the Google search engine. However, after piloting, we decided to remove the following elements: blogs, news, sponsored links, related articles, cited by, items without quotes, tabbed menus, YouTube videos, Google books, Wikis, indented materials, advertisements, “Did you mean?” prompts, tips. These changes were made firstly to remove information that is not useful or necessary for the given task and secondly with the purpose of decreasing variance in data due to individual strategic differences. Although we realize that this simplification may seem to compromise the ecological validity of the materials, an issue we return to in the “Discussion” section, we believe that the materials were realistic enough to retain the crux of search engine use. In fact, none of the participants spontaneously complained about missing cues.

In the six-item condition, the listing was manually stripped down. In the case of 24 items, typically only 20–21 items fit on the first A4 page, and the rest were printed on a second page. This means that there are probably more relevant items in the 24 item condition than in the 6 item condition.

Turning a paper page introduces a break that is somewhat analogous to loading the second result page by clicking “Next” on the first result page [12]. Many search engine users never go to the next page [9], and doing so was very rare in our experiment also.

3.3.3 The fake search engine (RCD4000)

Two search engine “replicas” were used: Google and RCD4000. They used the same search result contents; only layout differed.

In creating the layout of RCD4000, we aimed for recreating the look and feel of an engine that could seriously compete with Google. Google was used as the basis for the design. The layout was similar to Google’s with the following exceptions (see Figure 2): different colors, fonts, and logo were used. Moreover, we changed some of the terminology, such as “Results” -> “Pages,” “Go” -> “Find,” and “Preferences” -> “Settings.” The positions of elements were the same as in Google.

3.3.4 Ratings

After each choice, the following claims were rated on a Likert scale (1–7):

- **Satisfaction**: “I am satisfied with my choice.”
- **Confidence**: “I am confident that my choice is correct.”
- **Carefulness**: “I made my choice carefully.”
- **Suitability**: “I think that the search results were suitable for the task.”

In addition, at the end of each task block (six tasks with the same search engine), slightly modified items were rated, as follows, but this time they referred to the search engine (see Figure 3).

- Satisfaction: “I am content with the search results.”
- Confidence: “I am confident that my choices were correct.”
- Suitability: “I think that the search results were suitable for the tasks.”
- Preference: “I would choose this search engine for my use.”

3.3.5 Satisficing versus maximizing scale

After the tasks, we used the brief version of the Maximization Scale, which is a 13-item scale to measure tendency to try to maximize the outcomes of one’s choices. This scale presents several questions about everyday choice behavior. For example, those people who are likely to search for an even better radio station though already listening to a satisfactory one are more often “maximizers” than “satisficers.”

Previous work has associated high score with decreasing overall happiness and subjective well-being. Satisficers, by contrast, are generally happier with their choice, although they do a little less well in objective terms [28][29]. The prediction here is that maximizers would be more susceptible to the choice overload effect than satisficers. An analogous idea has been explored recently in an eye-movement study of search engine use [3].

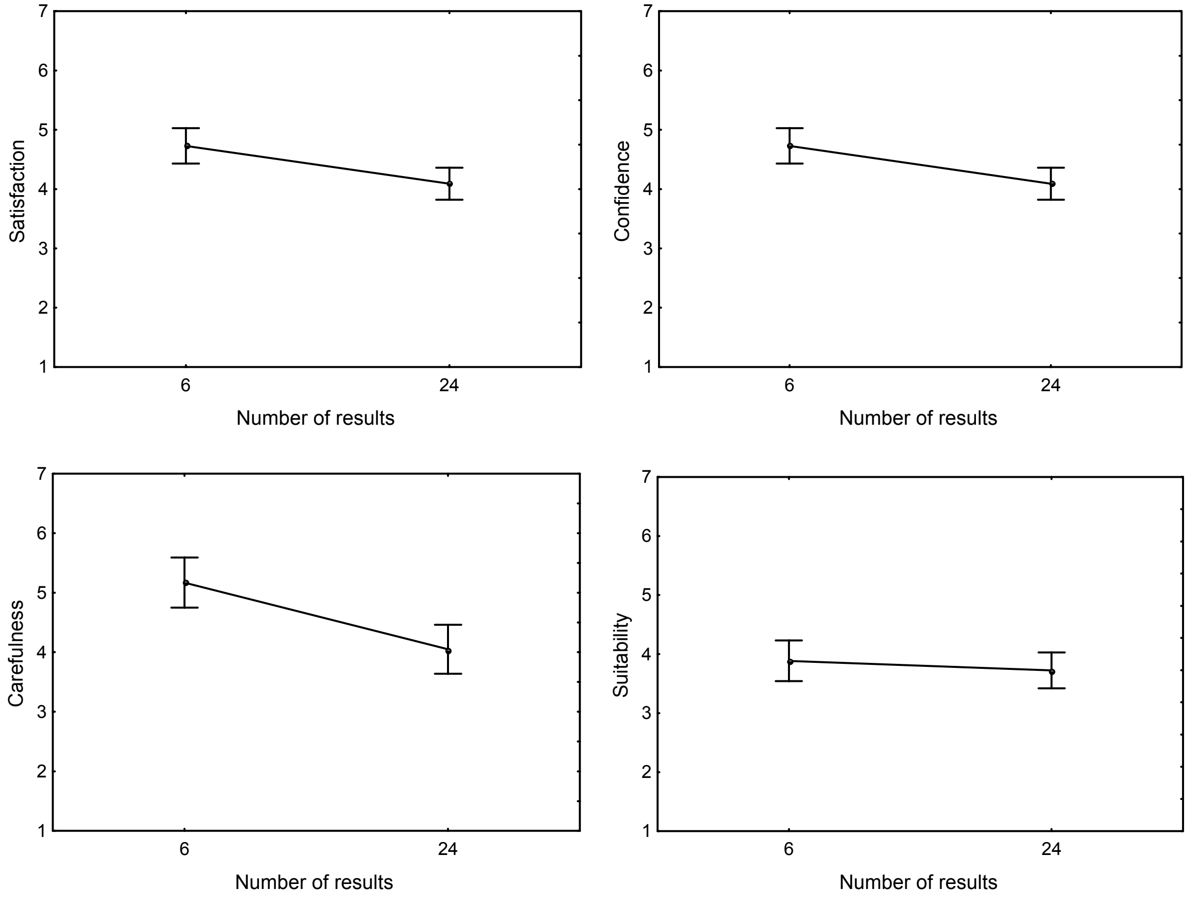


Figure 3: ***Satisfaction and confidence with the choice made (upper row), carefulness in making the choice, and suitability of the results (bottom row). Scale for all items: 1–7, where 7 is strongest. Error bars denote 95% confidence intervals.***

3.4 Procedure

Potential participants were approached with the pretext of doing research on search engine attitudes. The experiments were run in a campus cafeteria, lobby, or student lounge table environment while other people and variable background noise were present. However, care was taken to prevent direct interruptions. RCD4000 was introduced as "a new, exciting engine that can provide results as good as or better than Google's." Participants were not told that RCD4000 was in fact Google in disguise.

After brief practice, the 24 tasks were carried out. The experimenter's instruction explained that the scenario, query, and result page are provided "as is" and cannot be changed, and that it may feel that the items are not the best for the scenario. After providing the answer, and without seeing the result page anymore, the participants rated their experience before turning to the next task. Post-block questionnaires were administered at the end of each block of six tasks.

To prevent unusually long thinking times, the timeframe was limited to 30 seconds, after which choice was forced. A timing program for mobile devices (Egg Timer) was used, allowing precision of about one second. Notwithstanding a handful of exceptions, all participants followed this instruction.

To eliminate multi-experimenter bias, the second author ran all the participants. Each experiment lasted about 35–45 minutes.

4. RESULTS

The main results are reported in section 4.2 and in Figure 3. The results follow the pattern predicted by the paradox of choice: users were less satisfied with their choice when there were 24 items in the search result listing than when there were six.

For statistical testing, we utilize a 2x2 repeated measures analysis of variance (RM-ANOVA) with Number of Results and Search Engine as the two factors. Throughout, we use an alpha of .05.

4.1 Choosing an item

With fewer items, users tended to choose an item appearing higher in the list. With six items in the result listing, participants marked, on average, the 2.8th item (SE 0.1), but with 24 items the 6.0th item (SE 0.4). This difference was statistically significant, $F_{1,23}$=61.3, p<.001. The effect of Search Engine was non-significant, $F_{1,23}$=.3, p=.62, as was the interaction between the two, $F_{1,23}$=2.8, p=.11.

4.2 Evaluation immediately after choice

The main results follow the pattern of the paradox of choice as presented in Figure 3:

- **Satisfaction.** Participants were more satisfied with their choice in the six-item condition than in the 24-item condition, $F_{1,23}$=21.6, p<.001. This effect held for 22 out of the 24 participants (92%).

- **Confidence.** Participants were more certain about the correctness of their choice in the six-item condition than in the 24-item condition; $F_{1,23}$=14.1, p<.01. The effect held for 22 of the 24 participants (92%).
- **Carefulness.** Participants thought they made their choice more carefully in the six-item condition than in the 24-item condition, $F_{1,23}$=54.6, p<.001. The effect held with 22 out of the 24 participants (92%).
- **Suitability.** Users thought they had as good a pool to choose from in the 24-item as in the six-item condition. The difference between the six-item and 24-item condition was not significant: $F_{1,23}$=1.5, p=.24.

We did not find any effect of search engine; the effects reported above held for both Google and RCD4000.

4.2.1 Effect sizes

Analysis of effect sizes positions the findings in a small-to-medium size range. Following the recommendation of Keppel and Wickens for a two-factor within-subject design [22], omega-squared was used for estimation of effect sizes. Ranges for effect sizes by significant variable are as follows.

- Satisfaction: $.18 < \omega^2 < .30$, indicating a small to medium effect size
- Confidence: $.12 < \omega^2 < .21$, indicating a small effect size
- Carefulness: $.36 < \omega^2 < .53$, indicating a medium effect size

4.2.2 Task differences

Echoing previous results (e.g., [21]), task type had an effect on evaluation. The preferences task type was associated with the highest absolute scores for all four variables. For satisfaction, the mean for preference tasks was 5.5 (SE .14), 3.9 (SE .15) for problems, and 4.3 (SE .16) for simple fact-finding tasks.

To examine whether the paradox-of-choice effect holds for the three task types, we ran a 2 (number of results) x 3 (task type) RM-ANOVA. The effect of task type was significant on satisfaction ($F_{2,46}$=50.9, p<.001), confidence ($F_{2,46}$=61.4, p<.001), carefulness ($F_{2,46}$=9.6, p<.001), and suitability ($F_{2,46}$=89.4, p<.001). The interaction effect was borderline-significant for satisfaction ($F_{2,46}$=3.0, p=.06) but non-significant for confidence ($F_{2,46}$=2.0, p=.15), carefulness ($F_{2,46}$=.5, p=.64), and suitability ($F_{2,46}$=1.0, p=.39). Although there were no reliable interaction effects for any of the dependent variables (DVs), *post hoc* tests with Bonferroni correction on the borderline-significant variable "satisfaction" showed that both simple facts and problems task types manifest the paradox of choice (both p<.005), but the preferences type does not (p=.93). This was a little surprising, given that most research on the paradox of choice has been done in consumer domains where the tasks are very subjective. However, since the omnibus interaction effect was only borderline-significant, and the same effect was not repeated for the other DVs, we do not explore the finding further here.

4.3 Post-block evaluation of the search engine

In questionnaires positioned after each block (of six tasks), we asked about satisfaction *with the particular search engine* used in that block. Overall, we found no differences between the two engines in satisfaction, confidence, or suitability. While suitability and choice were in favor of Google, this trend was not statistically significant; both $F_{1,23}$<3.1, p>.09. However, the preference variable manifested a difference: participants felt that they would choose Google for their tasks; $F_{3,63}$=7.7, p<.001.

4.4 Other analyses

There were no practice effects on the variables of interest, all $F_{3,69}$<.32, p>.80.

None of the participants was a maximizer (max. 4.6, M=3.7, SD=.73) if compared against criteria used by Schwartz (e.g., maximizer group M=5.3 in Study 2 of [29]). Nevertheless, a *post hoc* median split was done to divide participants into high- and low-score groups. A t-test revealed no difference between the groups in satisfaction (t_{22}=1.6, p=.47) nor any of the other variables. We correlated all DVs with this score but found nothing reliable (all rs < .1). This lack of effects is disappointing but hardly surprising, since the subject pool was so homogenous on the Maximization Scale.

5. DISCUSSION

We know from cognitive psychology that choice overload can have three unfortunate effects: it can paralyze, it can lead to poor choices, and it can lead to dissatisfaction with even good choices. The power of modern search tools is extraordinary, but if they result in users feeling paralyzed and powerless, they become self-defeating. Putting "all the world's information" in front of people may solve one problem, but it creates another.

Virtually all of the research on choice overload done thus far has been in connection with consumer goods. The present study extends the phenomenon to the domain of information. We found that a six-item search result list was associated with higher satisfaction, confidence, and perceived carefulness than a 24-item list. The effect was robust; it held for all three task types and for 22 out of the 24 participants, although none was a maximizer [29]. Why the effect has not been reported before may be due to the effect size: Our effect size analysis revealed that the phenomenon is perhaps too small to be obvious to the naked eye, though it still is large enough to have ecological significance.

But will the effect occur outside the confines of the laboratory, or is it more a small blemish on the face of search engine use? The experimental method, as does any lab study, subscribes to a set of assumptions that may or may not hold water in real-world situations. We have tried to summarize key factors—known and yet unexplored—in Figure 4. These may pose boundary conditions for generalizability. However, we do wish to note that, in the present experiment, paralysis was essentially precluded by the demands of the experiment, and participants got virtually no feedback to inform them of whether their choices were good or bad. In real life, we might expect paralysis and poor choices to contribute to dissatisfaction. Therefore, we believe that our results, if anything, may understate the magnitude of the choice overload effect, since we show effects on satisfaction even when paralysis and poor-choice information are not available.

We suggest that, instead of pooling of results from disconnected lab experiments, the question of generalizability is ideally tackled in a large-scale, controlled online experiment that correlates presentation set size with subjective satisfaction and objective behavior (e.g., clickthrough curves).

	Variable	In the present experiment	Possible influence on the paradox-of-choice effect
1	Personality	Participants not maximizers	The effect may increase for maximizers and decrease for satisficers [29].
2	Skill in using search engines	Participants “average,” neither novices nor experts	Skill may decrease the effect. Skilled users may learn to control expectations, accept “good enough” results, and not worry about unsearched items.
3	Domain-related prior knowledge	Prior knowledge not measured	Pre-knowledge may help decrease the effect by collapsing a multidimensional choice into a one-dimensional choice, by increasing selectivity in use of cues.
4	Type of task	Only informational and transactional [5][7]	For example, navigational tasks may be less about choice and more about recognition of familiar items. A task is likely to have indirect effects via other variables known to be associated with the effect, such as time spent [21].
5	Solution ambiguity	Both closed- and open-ended tasks used	If the user is not certain about what it is that needs to be found, (s)he is more likely to evaluate his or her choice and therefore trigger the effect.
6	Importance of choice for user	Experimenter-imposed tasks	Increased subjective importance of succeeding in the task will boost the effect [28].
7	Availability of cues	Many Google-specific cues removed	Irrelevant cues will increase the effect by bloating the number of choice dimensions. Relevant cues will aid in narrowing down the consideration set and help in inter-item comparisons (see, e.g., [10]).
8	Query control	Queries given, could not be changed	Users may be able to formulate better queries after one failure [18]. A “divide and conquer” strategy may help limit the consideration set.
9	Time in task	Experimenter-imposed 30 s time limit	Increasing time in the task can 1) boost regret and “sunk cost effect” by increasing the remembered set but 2) can also help to resolve the paradox of choice if a satisfactory consideration set can be properly examined.
10	Extraneous tools	Not allowed	In real-world use, users can utilize extraneous means (bookkeeping, multiple windows for comparison; e.g., [2]) that may aid in overcoming paralysis
11	Opportunity to view the selected page	Link selected but page not viewed	Feedback for poor/good choice was excluded from the experiment. In real-life use, seeing the page may either increase or decrease satisfaction.
12	Opportunity to use the choice		Actually using the page for something further increases the probability that it will be evaluated.
13	Choice-evaluation interval	Evaluation forced immediately after the choice	Increasing the time interval between choice and evaluation makes evaluation more dependent on the remembered set, which may overly emphasize the top-ranked and clicked items (see, e.g., [32]).
14	Reversibility of choice	Not allowed	In contrast to many consumer domains, in search engine use the choices are reversible. The ability to easily redo the search and choose something else may decrease the paradox by reducing opportunity costs.
15	Multiple-option trials	Not allowed	Going back to the result page and trying another result is common (e.g., [17] [31]). Evaluation of multiple alternatives may help to limit the consideration set but equally well may boost the sunk cost effect if the options tried were good.

Figure 4: ***Factors that can contribute to choice overload in real-world use.***

Interestingly, recent results on search behavior point toward choice overload, although the issue has not been put on the table. One such signal is users’ “*over*-reliance” on ranking information. A recent study found that although users spend about the same amount of time looking at the *second* abstract in results as the first one, they nevertheless choose the first almost three times as often [12]. Something akin to paralysis has been reported when top-ranked items are not perceived as reliable: When top-ranked items were put at the end of the list, users spent more time checking each item, exhibited diffuse click patterns, and were less likely to eventually locate the best items [26]. But these effects are what one would expect as a consequence of choice overload: It is only natural to rely on extraneous information suggested by other people and authoritative sources—these can provide the only available means out of paralysis.

The paradox of choice has important implications for the design of search engine results. It calls for distinguishing result set from presentation set in search engine development (see also [6]). The study indicates that increasing recall can hamper user experience, unless considered in conjunction with presentation of items. What if, instead of 99.5 million Britney Spears links, the user were to be given only, say, six? However, we want to avoid the simplistic conclusion that six would be somehow optimal. We sampled only two points on the continuum of presentation set size, and there are unknown factors at play (Figure 4). It may turn out that Google’s default page set size of 10, combined with effective ranking and additional cues, is good enough to prevent choice overload in most searches. Future work will have to address this issue.

Beyond the obvious implication of limiting the page/presentation set, the next generation of search result design should focus on what others have called choice architecture [33]. We do not have room to examine the full argument here, but the general points are to 1) help narrow the consideration set, 2) aid in spotting diagnostic features of items, and 3) make comparisons more effective. For example, it is known that “defaults,” such as “I’m Feeling Lucky,” can be powerful, especially in the context of almost limitless choice. And the result page can be designed flexibly so that users can “opt in” to however many hits they want. In addition, there is evidence that the amount of choice people perceive is governed not only by the actual number of tokens that are present but also by how the tokens are organized into categories [25]. Through introduction of categorical structure into search results, small numbers of hits may be made to seem large or vice versa (see also [4]). In addition, more diagnostic cues can be designed. A recent study indicates that information in general-purpose cues like snippets may help to overcome the problem of over-reliance on ranking [9].

6. ACKNOWLEDGEMENTS

We thank Miikka Miettinen, Petteri Nurmi, Martti Mäntylä, Matt Wallaert, and Ken Rimey for comments. This work was supported by the Academy of Finland project ContextCues and the Fulbright Technology Industries Finland Grant for a Junior Scholar.

7. REFERENCES

[1] Asch, S. *Forming Impressions of Personality*. Taylor & Francis Group, 2004 (Re-print, original printed 1946.)

[2] Aula, A., Jhaveri, N., and Käki, M. Information search and re-access strategies of experienced web users. In *Proc. WWW* 2005, ACM Press (2005), New York, USA, pp. 583-592.

[3] Aula, A., Majaranta, P., and Raiha, K. Eye-tracking reveals the personal styles for search result evaluation. *Lecture Notes in Computer Science 3585* (2005), 1058.

[4] Beitzel, S., Jensen, E., Chowdhury, A., Grossman, D., and Frieder, O. Hourly analysis of a very large topically categorized web query log. In *Proc. SIGIR'04,* ACM Press (2004), New York, USA, pp. 321-328.

[5] Broder, A. A taxonomy of web search. *SIGIR Forum 36*, 2 (2002), 3-10.

[6] Brookes, B. Measurement in information science: objective and subjective metrical space. *Journal of the American Society for Information Science 31*, 4 (1980), 248-255.

[7] Byrne, M., John, B., Wehrle, N., and Crow, D. The tangled web we wove: A taskonomy of WWW use. In *Proc. CHI'99*, ACM Press (1999), New York, USA, pp. 544-551.

[8] Clancy, K., and Wachsler, R. Positional effects in shared-cost surveys. *Public Opinion Quarterly 35*, 2 (1971), 258-265.

[9] Clarke, C., Agichtein, E., Dumais, S., and White, R. The influence of caption features on clickthrough patterns in web search. In *Proc. SIGIR'*07, ACM Press (2007), New York, USA, pp. 135-142.

[10] Cutrell, E., and Guan, Z. What are you looking for?: an eye-tracking study of information usage in web search. In *Proc. CHI'*07, ACM Press (2007), New York, USA, pp. 407-416.

[11] Eisenberg, M., and Barry, C. Order effects: A study of the possible influence of presentation order on user judgments of document relevance. *JASIS 39*, 5 (1988), 293-300.

[12] Granka, L., Joachims, T., and Gay, G. Eye-tracking analysis of user behavior in www search. In *Proc. SIGIR'*04, ACM Press (2004), New York, USA, pp. 478-479.

[13] Hogarth, R., and Einhorn, H. Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology 24*, 1 (1992), 1-55.

[14] Huang, M., and Wang, H. The influence of document presentation order and number of documents judged on users' judgments of relevance. *JASIST 55*, 11 (2004), 970-979.

[15] Iyengar, S., and Lepper, M. When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology 79*, 6 (2000), 995-1006.

[16] Iyengar, S., Jiang, W., and Huberman, G. How much choice is too much? contributions to 401 (k) retirement plans. *Pension Design and Structure: New Lessons from Behavioral Finance* (2004), 83-96.

[17] Jansen, B., and Spink, A. An analysis of web documents retrieved and viewed. In *Internet Computing Conference* (2003).

[18] Jansen, B., Spink, A., and Pedersen, J. A temporal comparison of altavista web searching. *Journal of the American Society for Information Science and Technology 56*, 6 (2005), 559-570.

[19] Kahneman, D., and Tversky, A. *Prospect theory: An analysis of decision under risk*. Bradford Books, 2004.

[20] Katzer, J., and Snyder, H. Toward a more realistic assessment of information retrieval performance. In *Proc. ASIS 1990*, pp. 80-85.

[21] Kellar, M., Watters, C., and Shepherd, M. A goal-based classification of web information tasks. In *Proc. ASIST'06.*

[22] Keppel, G., and Wickens, T.D. *Design and Analysis: A Researcher's Handbook* (4th International Ed.). Upper Saddle River, NJ: Pearson Prentice-Hall.

[23] Kochen, M. *Principles of information retrieval*. Los Angeles, CA: Melville, 1974.

[24] Lorigo, L., Pan, B., Hembrooke, H., Joachims, T., Granka, L., and Gay, G. The influence of task and gender on search and evaluation behavior using google. *Information Processing and Management 42*, 4 (2006), 1123-1131.

[25] Mogilner, C., Rudnick, T., and Iyengar, S. The mere categorization effect: How the presence of categories increases choosers' perceptions of assortment variety and outcome satisfaction. *Journal of Consumer Research 35* (2008), 202-215.

[26] Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., and Granka, L. In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication 12*, 3 (2007), 801-823.

[27] Parker, L., and Johnson, R. Does order of presentation affect users' judgment of documents? *JASIS 41*, 7 (1990), 493-494.

[28] Schwartz, B. The *Paradox of Choice: Why More Is Less*. Harper Perennial, 2005.

[29] Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., and Lehman, D. Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology 83*, 5 (2002), 1178-1197.

[30] Shah, A., and Wolford, G. Buying behavior as a function of parametric variation of number of choices. *Psychological Science 18*, 5 (2007), 369-370.

[31] Silverstein, C., Marais, H., Henzinger, M., and Moricz, M. Analysis of a very large web search engine query log. *ACM SIGIR Forum* 33 (1999), 6-12.

[32] Teevan, J. How people recall search result lists. In *Proc. CHI'06*, ACM Press (2006), New York, USA, pp. 1415-1420.

[33] Thaler, R., and Sunstein, C. *Nudge*. New Haven: Yale University Press, 2008