# Public Bookmarks and Private Benefits: An Analysis of Incentives in Social Computing

**Rick Wash, Corresponding Author**

School of Information, University of Michigan   rwash@umich.edu

**Emilee Rader**

School of Information, University of Michigan   ejrader@umich.edu

**Users of social computing websites are both producers and consumers of the information found on the site. This creates a novel problem for web-based software applications: how can website designers induce users to produce information that is useful for others? We study this question by interviewing users of the social bookmarking website del.icio.us. We find that for the users in our sample, metadata reflecting who bookmarked a webpage better supports information seeking than free-form keyword metadata (tags). We explain this finding by describing differences in the way that the design of del.icio.us motivates users to contribute by providing personal benefits for bookmarking and tagging.**

## Introduction

Websites like flickr (http://www.flickr.com) YouTube (http://www.youtube.com), and del.icio.us (http://del.icio.us) belong to a growing category of Internet applications broadly referred to as *social computing* sites. The information presented by these sites is generated by the users themselves, rather than by an agency officially tasked with that responsibility, like a publishing company or news distributor. Users of social computing websites play two roles with respect to the information the sites contain: they can act as *producers* (contributors of information) or *consumers* (seekers of information). Consider the social bookmarking website del.icio.us http://del.icio.us, which provides the capability for users to bookmark web pages and associate user-generated metadata, or tags, with them (Marlow, Naaman, boyd, & Davis, 2006). Information consumers are able to discover new web pages using del.icio.us only because the actions of information producers caused the web pages to be stored by del.icio.us in the first place, through the action of saving the web pages as bookmarks. This aspect of social computing – that the community of information consumers benefits from the contributions of information producers – presents an interesting research question. How is it that producers, through actions intended to serve purely personal goals, come to generate information that is beneficial for the larger community?

An *incentive* is something that influences a person to choose one course of action over the alternatives. Often, incentives are thought of in terms of money: a bonus for meeting performance goals, or a subject payment in an experiment. However, incentives can be anything that induces expectations of a future positive outcome or benefit. For example, the capability of del.icio.us to store one's bookmarks in such a way that they are available from any computer connected to the Internet provides an incentive to use del.icio.us rather than a web browser's built-in bookmark tool. Incentives are important in social computing because they motivate or induce certain types of actions and not others. By identifying the pattern of incentives, it is possible to better understand user behaviors that create *efficient,* or optimal, outcomes both for individuals and for the community of users. We say that incentives are *aligned* when the incentives for one group of users to maximize their own benefit, result in circumstances that are also beneficial for the other group.

In this paper, we focus on del.icio.us as a case study of the incentives in social computing. The literature investigating user behavior and overall usage patterns in social bookmarking and tagging systems like del.icio.us is

growing rapidly. A widely-held belief is that in such systems, tags are the primary means by which information producers organize their own bookmarks, and also how information consumers seek and discover new information that has been bookmarked by others on topics that interest them (Marlow, Naaman, boyd, & Davis, 2006; Udell, 2005). Through analyzing semi-structured interviews with twelve regular users of del.icio.us, we discovered that metadata reflecting the identity of the user who saved a web page, which is automatically associated with each bookmark when it is created, was more useful for consumers' information seeking than the user-generated tags. Producers created public bookmarks and tags for their own private reasons. While consumers received benefit from the web pages that were added to del.icio.us, the associated tags did not help their information seeking. In other words, we found that *incentives are aligned* for bookmarking actions, but not for tagging.

This insight is important because it allows us to consider how alternate incentive designs might shape behavior differently, and consequently the corpus of information available in del.icio.us. For instance, many people talk about the potential for a grass roots taxonomy, or *folksonomy*, to emerge from tagging behaviors on del.icio.us. As we will discuss later in this paper, we believe that the incentives are not currently aligned for this to take place. However, by understanding how incentives shape the corpus, future designs may be more successful. Also, it highlights how public metadata acts as the "glue" which holds the del.icio.us corpus together. Without the user, tag, and date-time metadata associated with each bookmark, individuals in the role of information consumer would find the corpus much less useful.

## An Overview of Del.icio.us

We begin with a description of the interface and functionality of del.icio.us, and some important definitions. Users of del.icio.us willing to create an anonymous user account are able to save web pages as bookmarks. When bookmarks are created, it is possible to associate tags with them. In del.icio.us, tags are restricted to a single word, and plurals or different spellings of the same word are treated as different tags. Each bookmark has the following metadata associated with it when it is created: the username of the information producer, the tags selected, and the date and time the bookmark was created. Users may also associate a "note", or text string, with each bookmark they create. See Figure 1 for an example of the del.icio.us interface for creating, or *posting*, bookmarks.

Information consumers browsing del.icio.us view subsets of the corpus, or *collections*, consisting of bookmarks that have been filtered according to certain metadata characteristics. For example, each user account is a collection, delimited by the metadata indicating the username of the person who created the bookmarks. A collection can also consist of all web pages associated with a given tag. Clicking the tag "library" in the list of popular tags on del.icio.us displays the collection of all web pages bookmarked by any del.icio.us user having the tag "library" associated with them. Clicking on a username displays the collection of web pages bookmarked by that person. The metadata for each web page is visible along with the title of the web page, as shown in Figure 2. For example, this particular user posted the ASIS&T call for papers to del.icio.us one day ago using the tags "asist, conference, 2007, paper." That specific web page has been bookmarked by 12 other people. All metadata items are also links that filter collections of bookmarks.

Figure 1. The interface for posting bookmarks                    Figure 2. One user's collection of bookmarks

url http://www.asis.org/
description ASIS&T The American Society for Information Sci
notes
tags information cool
save

**recommended tags**
information technology
**your network**
for:joshua
**popular tags**
associations library information IA informationscience techn

del.icio.us / bierdoctor /
your bookmarks | your network | subscriptions | links for you | post

All your items (143)

« earlier | later »     page 1 of 3

2007 ASIS&T Annual Meeting Call for Papers edit / delete
to asist conference 2007 paper ... saved by 12 other people ... 1 day ago

A Journey to the Dark Side - New York Times edit / delete
to for:rwash beer ... on oct 19

WGN Weather Weblog edit / delete
to weather chicago blog skilling ... saved by 16 other people ... on oct 18

Index - Labradoodle Discussion Forum edit / delete
to dog labradoodle information forum ... on oct 18

view.jpg (JPEG Image, 1280x720 pixels) - Scaled (85%) edit / delete
to photos webcam banff ... saved by 2 other people ... on oct 18

Flickr: Photos tagged with labradoodle edit / delete
to dog photos labradoodle ... on oct 18

If a registered user wishes to follow the latest bookmarks saved by a certain person, or having a certain tag, it is possible to "subscribe" to users, or tags, or a combination. New bookmarks satisfying these metadata criteria will then appear in the user's account. For example, if user A subscribes to user B, every time B posts a new bookmark it will appear when user A clicks the "your network" link in her account (see Figure 2). In addition, registered users can subscribe to all bookmarks associated with a given tag, or all bookmarks posted by a given user with a given tag, which can be accessed by clicking on the "subscriptions" link. Finally, the "links for you" link returns bookmarks that have been saved by one user specifically for another user, using a special tag format supported by del.icio.us.

A number of researchers have studied del.icio.us and the phenomenon of tagging.  Golder and Huberman and Halpin et al. found (for del.icio.us) that the frequency distribution of tags used on a given site tends to stabilize over time, with a definite most appropriate tag and a power-law distribution of tags (Golder & Huberman, 2006b; Halpin, Robu, & Shepherd, 2007).  A number of researchers have tested designs to improve tagging systems like the one in del.icio.us  (Rivadeneira, Gruen, Muller, & Millen, 2007; Sen et al., 2006a; Storey, Cheng, Bull, & Rigby, 2006; Xu, Fu, Mao, & Su, 2006).  Ames and Naaman studied motivations for tagging in Flickr.  They found tags were used both for organization and to communicate, and were used both for selfish and for social purposes (Ames & Naaman, 2007).

By default, bookmarks in del.icio.us are public information, meaning that any user may browse any other user's bookmarks and tags without logging in to the system. This makes the del.icio.us corpus a *public good*, meaning that the information in the corpus can be accessed simultaneously by everyone without ever being used up (Mas-Colell, Whinston, & Green, 1995). According to economists, one of the distinguishing features of public goods is that is that most of the time individuals are insufficiently motivated to contribute to public goods relative to what would be best for the community or society as a whole. The standard solution is to have the government provide the good, as is the case for the provision of national defense (the army). Voluntary provision of public goods by individuals is an open research problem (Andreoni, 2006); one of the reasons del.icio.us is an interesting case study is that this problem appears to have been solved in this setting.

The problem of motivating people to voluntarily provide information has been studied in other public goods settings. Forte and Bruckman studied why people contribute to Wikipedia, a popular website encyclopedia.  They found a variety of motivations, both selfish and social (Forte & Bruckman, 2005).  Ling et al. and Harper et al. studied contributions of ratings to an online movie rating system (MovieLens), finding that these contributions could be increased by making users more aware of some social information  (Harper, Li, Chen, & Konstan, 2005; K. Ling et al., 2005).  Feldman studied incentives for users to contribute information goods to peer-to-peer file sharing systems.(Feldman & Chuang, 2005)  Finally, von Ahn and Dabbish motivate users to provide words that accurately describe images by making the task into a game  (von Ahn & Dabbish, 2004).

## Method and Participants

We conducted twelve 1.5 hour semi-structured interviews with users of del.icio.us during the summer of 2006. All had used del.icio.us for multiple months, and had posted bookmarks to del.icio.us about once a week on average, for a number of months before the study took place. Five of the twelve users were masters students or recent graduates at a local university, three were PhD students, one was an undergraduate, and three were information technology professionals. A number of del.icio.us users responded to fliers posted around campus and to Internet postings on del.icio.us. Our respondents were selected from that group to represent a wide variety of usage patterns. For example, the total number of bookmarks that had been saved by a single respondent varied from a low of 60 to a high of approximately 3000. Table 1 provides descriptive statistics on respondents' use of del.icio.us. Our selection of these particular respondents for participation in the study reflects our desire to collect data on a range of possible user behaviors.

Table 1: Descriptive statistics about our respondents

| N = 12 (8 Men, 4 Women) | Mean | Std. Dev |
|---|---|---|
| Total Bookmarks | 950 | 1030.77 |
| Total Unique Tags | 400 | 356.075 |
| Months Using del.icio.us | 15.90 | 8.73 |
| New Bookmarks/Week | 16.15 | 25.34 |
| New Bookmarks/Day | 2.31 | 3.62 |

Our sample consisted of highly educated, tech-savvy students and professionals, meaning that our respondents were likely to be more sophisticated and self-aware in their use of technology than the average home computer user. As a result, our respondents were probably more likely to have attempted to optimize their personal usage of del.icio.us, to expend more effort when taking advantage of the available functionality, and to tolerate usability problems. In addition, respondents self-selected for participation when they responded to our advertisements, which indicates that their use of del.icio.us was salient and important enough to them that they were willing to volunteer to participate. We believe that by requiring that respondents be regular users of del.icio.us for several months before the study took place, such bias was unavoidable. Our choice to use this criterion was motivated by a desire to interview people who could recall many past instances when they had used del.icio.us. Our sample therefore consists of users likely to have explored different features and uses, biasing our results toward a greater variety of activities than might be seen in a sample obtained based on different criteria.

The interviews were comprised of three phases. In the first phase, the interviewer asked general questions about respondents' use of del.icio.us: how often do you use it?, what do you bookmark?, how do you choose tags?, and similar high-level questions. The second phase consisted of ten search tasks. Respondents were sequentially presented with ten printouts of web pages found in del.icio.us, five bookmarked by the respondent and five bookmarked by others, and were to find them using only del.icio.us. They were instructed to think aloud during this task (results from the search tasks will not be presented in this paper). Finally, the interviewer looked through the respondent's bookmark history and asked questions designed to trigger retrospective accounts of past actions, such as "Tell me about that bookmark. What were you doing when you posted it? Tell me about the tags you chose." The interviewer also asked detailed questions about respondents' use of their subscriptions.

The interviews were recorded, transcribed and coded using Atlas.ti. The analysis was conducted in a similar fashion to Miles and Huberman (Miles & Huberman, 1994). Informal coding began with a list of classes of behavior upon which to focus, and the code list developed as we proceeded. We identified the stated *motivations* of the respondents, their *actions* undertaken using del.icio.us, and the *outcomes* of those actions, including inferences about benefits they received from using del.icio.us. Summary matrix displays were created showing which users exhibited similar motivations, subsequently undertook which actions, and received which benefits. Coding for these three things (among others) enabled us to locate and analyze possible instances where incentives had influenced behavior. It is important to note that this is not an exact science; unlike monetary incentives that can be explicitly

identified and measured, identifying factors that induce users to behave in certain ways is an interpretive process. Nevertheless, and despite the varied usage patterns of our respondents, patterns emerged that present intriguing evidence for the role incentives play in del.icio.us. These patterns are described below.

### Results: Producer and Consumer Incentives

In this section we present data describing three categories of activities users engage in with del.icio.us: bookmarking, tagging, and information seeking. We found that information producers have many different motivations for bookmarking and tagging, and that consumers browse del.icio.us for several different types of information. We then describe evidence supporting our claim that incentives for bookmarking are aligned to produce positive outcomes for both producers and consumers, but the incentives for tagging are not.

### Why Bookmark?

Respondents reported three motivations for bookmarking web pages:

- To keep track of useful or interesting web pages
- To access bookmarks from multiple computers
- To achieve recognition from other users of del.icio.us

The primary motivation for bookmarking web pages reported by all respondents, was a desire to have ready access to information they found useful or interesting. As one respondent, Fred said: "Any web page that I see, I basically ask myself this question: Would I ever have a need to find this again? And if I do I just bookmark it." Seven respondents valued the ability to access their bookmarks from multiple computers. Zoe liked del.icio.us "because I'm working on so many computers and so many different places, it's just made me so much more efficient. It's a lifesaver." The action of bookmarking leads to the benefit of future access to web pages that contain information important enough to save. The websites that respondents bookmarked can be divided into a few categories:

- *Topics of specific interest to the respondent*. Alice bookmarked PhD programs she was interested in, Bob bookmarked library-related links, Charlie liked web pages on sustainability, Oscar looked for programming skills, and Marvin was into community informatics. There is a different list of topics for every respondent in our study.
- *Pages the respondent hasn't finished reading*. Bob described this well: "Umm, I get to that situation where I have eight different tabs open in Firefox and I don't really have the time to read them all. [I'll] get up and do something, and so I'll bookmark a bunch of them so that I will go back and in theory read them later." Half of these respondents expressed discontent over rarely actually returning to these web pages.
- *Reference information or internet tools*. Nine respondents reported this type of bookmark. Examples include new search engines (Zoe), manuals for the Perl programming language (Oscar and Eve), and collaborative text editors (Eve).
- *Novelty or funny web pages*. Fred described these well: "something funny. Like a video of a monkey sniffing itself or something. [...] Or if something is just, oh wow cool, a new story or an amusing rant or a blog post, those get added [...] as well." Another example was when Victor bookmarked a web page "because I thought the title was so ridiculous."

In addition, eight out of twelve respondents were motivated to bookmark in order to share web pages with other people. They used a variety of actions for sharing: using del.icio.us' built-in tag convention for sending a bookmark to another user (the "for:username" tag); using a previously agreed-upon tag; through knowing that a particular person subscribes to their bookmarks; or just by explicitly instructing the person to go look for it. As Trent explained, he does this "in lieu of [having] sent an e-mail with a link in it." Five of those eight, however, reported that this motivation only rarely influenced them to bookmark. Two respondents bookmarked they had created in del.icio.us, hoping that other del.icio.us users would find them. Also, one respondent mentioned that he believes some search engines index del.icio.us, and therefore bookmarks pages to increase their Google PageRank (Brin & Page, 1998).

Social recognition only functions as a motivation when the respondent is aware of other users' behavior, and in our

sample this awareness varied widely. One respondent seemed to have no knowledge of others looking at his bookmarks. Seven respondents had directly told other people to look at their bookmarks, either indicating a specific bookmark, as in: "I couldn't remember the Dog Judo link but I wanted him to check it out, so I sent him a thing that said Go to Dog Judo on my del.icio.us" (Eve), or directing them to a certain tag. Six respondents were aware of other people who subscribed to their bookmarks; often this awareness came from conversations with friends. Isaac was aware of his friends' subscription because, "Like every so often [a friend] will say I noticed you bookmarked that or [another friend] will say that."

Half of our respondents mentioned a general awareness that the bookmarks they post are public information, unless they specify that they should be private. However, they reported that this awareness rarely affected their actions. For example: "I do make a conscious decision of whether or not I want it to be available for everybody but 98% of the time I don't care" (Charlie); and, "Even though [my use of del.icio.us] is oriented primarily towards myself, the awareness that it is public never goes away totally" (Trent).

## Why Tag?

The primary motivation respondents reported for tagging was to organize their bookmarks and make it easier to find them if the need arose in the future. Respondents generally used one or more heuristics for choosing tags:

- Reuse tags he or she has applied before
- Create and adhere to mental rules or definitions for specific tags
- Choose terms he or she expects to search on

These are only heuristics, do not apply in all cases, and were not necessarily applied consistently. In addition, these data are self-reports, and other unconscious factors could also be a factor in the choice of tags. We have attempted to verify the reports by manually looking through the respondents' tagging history and by eliciting multiple instances of the heuristics during the interviews. Respondents' observable bookmarking behaviors indicate a fair amount of compliance with these heuristics.

To reuse tags, respondents placed priority on choosing tags they had used in the past. "I will not add a new tag until I have a group of things that I think it goes with," said Zoe. Reusing old tags made bookmarks easier to find by minimizing the length of respondents' tag lists, which most respondents reported searching visually when they wanted to find a bookmark. Victor described the problem his tag reuse solves: "One of my friends, his tag section goes way down below the fold [...] I'm like 'How on earth do you sort through all these?' And he said, 'I don't.'"

Respondents often 6 reported that they had created mental rules or definitions a number of their tags. For example, Peggy described some of her rules about tags related to blogs: "So 'blogs' are usually other people's blogs. 'Blogging' would be something that's about usually research about blogging. And then if it's something like Blogger for instance or LiveJournal then that would be a 'bloggingtool'." The creation of such rules is also observed in the creation of folders in which to store documents (Whittaker & Sidner, 1996). One advantage of tags over folders is that a single web page can be associated with multiple tags, reducing the effort involved with selecting one and only one location for the information. However, the more tags one has, the more overhead is involved with remembering the mental rules necessary for distinguishing among tags. People in general have a hard time being consistent within themselves with the tags they use (Golder & Huberman, 2006a), and our respondents were no different. Trent described how he handles this problem: "I've been sloppy in the past about 'collaborative' and 'collaboration,' so this one got tagged as both. Just to make sure that I got coverage." Eve's tags showed the same characteristics: "So, apparently I'm using funny and humor interchangeably. And not reliably. So I should remember that when I'm looking for something funny I also label it humor. [...]" Respondents also had problems with singular/plural tags and with misspelled tags, all of which create the situation where multiple tags have the same logical meaning. Respondents with this problem speak of their tags as "dirty," and the occasional act of fixing this as "cleaning" their tags.

Seven respondents reported choosing tags by trying to guess what terms they might search on in the future to find the bookmark. Eve described her thought process:

**Interviewer**: So, on the librarian video, how did you choose the tags that you have?

**Eve**: [...] If I were looking for this again [...] I'd be like "What was that video about the girl in the library with that guy?" But girl and guy is not very helpful, so library and video won.

In addition to the general heuristics for tag choice, respondents used tags to represent personally meaningful categories. Five respondents used tags to represent projects. Whenever they bookmarked a web page related to the project, one of the tags they applied to that web page was the project name. Marvin said this was "so I can just type in [the project name] and the things related to that project should show up if I did it right." Ten of the twelve respondents used tags for purely personal purposes. This is similar to the "functional purpose" tags of (Golder & Huberman, 2006a) and the personal tags of (Sen et al., 2006b). Rather than placing icons in a specific location on the desktop to serve as reminders, as seen in some personal information management studies (Barreau & Nardi, 1995; Bruce, Jones, & Dumais, 2004), respondents were using special tags as reminders within del.icio.us. The "toread" tag is one example, used by at least four of the respondents. Fred had a "wishlist" for items he would like to purchase. Both Alice and Oscar used the tag "research" to refer to web pages they wanted to remember to return to, because they might be useful for their respective research projects. The meaning of all of these tags is highly subjective and personal, and can only be correctly interpreted and understood by someone who knows the context.

Four out of twelve respondents mentioned consciously trying to build a collection of links on a specific topic that would benefit the larger community of del.icio.us users. Alice said, "I tag everything on [topic of interest] I can find. I was so frustrated when I started working with this stuff that I just couldn't find information about it. [...] There aren't many places for it so I have probably collected one of the larger lists out there." For Alice, wanting to be known as an expert on the topic of the collection seemed to be an additional motivation beyond personal organization and community benefit.

### Information Seeking

Information consumers are the beneficiaries of others' bookmarking activities when they browse del.icio.us to find new information. Browsing has been defined as, "a kind of searching in which the initial search criteria or goals are only partially defined or known in advance" (Chang & Rice, 1993). We found that respondents' goals for discovering new information fell into three categories:

- *Novelty information*: "something entertaining" (Alice)

  *Topical information*: web pages relevant to specific topics

  *Social information*: updates on friends' interests and activities via following their bookmarks

The most common action for *novelty discovery*, undertaken by seven respondents, was subscribing to someone they knew personally. Eve said: "I check out [my friend, he] always goes to really interesting places," and, Bob reported that he subscribed to a friend because he, "like[s] to pick his brain for cool stuff." Respondents' *topical discovery*, or seeking web pages that contain information on specific topics, was either a one-time seeking behavior, or it was due to a continuing interest in a particular topic. For one-time seeking, the most common action was to click on the "Saved by X other people" link. Seven respondents reported doing this for topical discovery, and two for novelty discovery. Bob described his reasoning:

"If I've got something bookmarked myself and it says 'Saved by X other people,' then it's more intriguing to me if there are very few people who have saved it. Because that means I belong to this elite group of people who actually find this stuff interesting. [...] And then maybe I'll take a look at what else they've bookmarked because if they are interested in something that I'm interested in maybe they've got other stuff that I'd be interested in."

For continuing topical discovery, respondents again reported subscribing to someone they knew in real life, or occasionally someone who is famous, or was found using "Saved by X other people." As Zoe discovered, "certain people tend to tag the same things I'm interested in." Five respondents reported looking on del.icio.us for other users with similar interests, and then subscribing to those users. Finally, *social discovery* was used to keep tabs on friends.

Peggy reported, "It's just interesting to see what it is they're up to. So like my friend Matt who's not in the area anymore, I think I get a sense of what it is he's doing." Mostly, this occurred through subscriptions. Charlie gives a good example:

"One of my friends [...] just got, you know, just got a job in San Francisco. I believe she went out to interview and then all of a sudden there's like 50 links to apartment search in San Francisco, and a few days later she tells me oh I got the job in San Francisco and I was like I know."

Interestingly, most respondents reported taking advantage of user metadata for information seeking and discovery, but not tag metadata. Tags were rarely mentioned in the context of novelty and social discovery. Because assessments of how interesting or entertaining a web page might be is a subjective judgment, it is reasonable to expect that it would be difficult to find a tag which would capture this assessment accurately. Only one respondent reported browsing tags (like "funny"), or searching del.icio.us, for novelty purposes. In social discovery, it is the user and not the topic that is of interest. Since bookmarks are always automatically associated with the user who bookmarked them, tags are not needed. Finally, a number of respondents struggled with using tags for topical discovery; three explicitly mentioned trying to do so and failing find the information they wanted. Only one respondent (Trent) subscribed to any tags, and he was careful to block users (using del.icio.us's built-in blocking mechanism) who post too many bookmarks for which "none of [them] fit my definition." Only four respondents mentioned browsing del.icio.us for one-time topical discovery by looking at specific tags, and they reported doing it only rarely.

## Discussion

### Incentive Alignment

In the previous section, we described how information producers and consumers use del.icio.us to meet their individual, private needs. It requires no extra effort for producers to perform the actions that make information seeking in the del.icio.us corpus effective for consumers: bookmarks are automatically public, and as we have shown, the user information that is essential for all types of information discovery is automatically associated with every bookmark that is created. User metadata is the means by which bookmarks are *discoverable* by information consumers. We say that the incentives are aligned because bookmarks created by producers to maximize their own benefit from using del.icio.us also make possible positive information seeking outcomes for consumers.

For example, when a user bookmarks web pages to save them for later, the public record of his bookmark history reveals information about his preferences and interests, allowing other users to make inferences about how similar his interests might be to theirs. Other users with similar interests stumble upon such collections through the "Saved by X other people" link. Fred indicated some level of awareness of this incentive alignment: "[I have] mostly just a general sense of the network, of people. If someone tracks the stuff I post enough, then I assume they care about the eigenvalues of the things that I like. And so I figure if I like it, they'll like it." However, as we have shown, producers use tags to organize their own collections of bookmarks, but consumers rarely use tags when engaging in novelty or social discovery, preferring instead to browse by user (see Figure 3). We also found that consumers dislike using tags for topical discovery. We believe that this is due to four characteristics of language use that make it difficult for people to be consistent in the tags they use, even within their own collections of bookmarks.

### Tags and Language Use

First, the *vocabulary problem*(Furnas, Landauer, Gomez, & Dumais, 1987) states that random pairs of people choose the same label for an object on average about 20% of the time. This robust tendency results from humans' imprecise and flexible use of language in conversational settings, where meaning is determined by the surrounding context and complex communication processes. This suggests that if two random users having different knowledge and situational contexts create a tag for the same web page, there is an 80% chance on average that they will NOT choose the same tag. Similarly, if an information consumer attempts to imagine what tags might be applied to the information she is looking for, chances are low that she will end up using the same words to represent the same concepts in the same way as others have used them.

Second are *synonymy* (multiple words that can be used interchangeably in the same context) and *plurals/tenses* of the same word. Eight respondents noticed this problem just within their own bookmarks! Trent mentioned being "sloppy in the past about 'collaborative' and 'collaboration'" and Charlie "at one point had 'recipe' and 'recipes'." Third is *polysemy*, or one word that has many meanings or senses. One example of this is the word "python" which can represent a snake, a programming language, or a comedy troupe from Great Britain. Expertise can also contribute to polysemy: people with different levels of expertise may end up using the same word to represent different concepts – their internal rules for what that tag represents are different. Fred subscribed to the 'security' tag, but as a computer security expert he found many of the bookmarks with that tag too basic for him. However, a novice user would likely find the web pages he bookmarked under 'security' too advanced. Finally, users of del.icio.us tend to create tags that have personal or figurative meanings only they have the requisite contextual knowledge to understand, like the "research" or tag several respondents used to refer to their respective projects.

These characteristics of language use produce a great deal of variability in the collections of bookmarks delimited by specific tags. It requires a considerable degree of effort to use tags consistently in one's own collection of bookmarks, where the incentive of maintaining a well-organized collection exists. We did not find evidence that our respondents attempted to match their internal rules for tag use with the ways in which others applied the same tags, which leads us to conclude that producer incentives for tagging are not aligned with consumer information seeking. There is little incentive to expend the effort required to combat these facts of language use; people seem able to tolerate this in their personal list of tags, and they have found other methods to discover new information without using tags.

## Constraints and Locus of Control

There is yet another way that incentives play a role. Discovery via user metadata works because there are constraints in place on a user's own collection of bookmarks and tags such that they remain personally efficient. Users are motivated to keep their bookmarks and tag lists relevant to their needs and interests, and small enough to be manageable. People are also motivated, as we saw in our respondents, to create tags that are personally efficient so that they can remain organized. Other constraints such as limits to attention and time spent bookmarking help to control the number of bookmarks and tags an information producer is able to create. Our respondents created between 1 bookmark per week and 13 bookmarks per day, and averaged slightly over 2 bookmarks a day. This is much smaller than the hundreds of bookmarks a day associated with some popular tags (like "music").

In the case of bookmarks, coincidentally or by design, these constraints result in a corpus populated with content that others want to see. However, in the case of tags, the only constraints that exist are those that might be imposed by the technical capabilities of the system. No constraints or happy coincidences cause them to conform in ways that are efficient for the community, meaning that any user can apply any word at any time as a tag for any web page! This is particularly important for ongoing discovery through subscriptions, where the rate of incoming bookmarks depends on how prolific others are; too many bookmarks make it difficult to keep with the influx of web pages. Fred reported experiencing this when he subscribed to the tag "security": "I'd go through it [the list of bookmarks], but I end up skipping a lot." Ultimately, he stopped using that subscription altogether.

One way to think about the difference between bookmarks and tags is that the collection of bookmarks created by a given user is under a single *locus of control* (the user), while no central authority oversees the way words are used as tags. When choosing tags for a new bookmark, none of our respondents expressed concern about increasing the number of bookmarks system-wide already associated with a tag. This difference is similar to the architectural difference between the Internet and Cable TV (Mackie-Mason, Shenker, & Varian, 1995). *Content-aware* architectures, like cable television, have the capacity for editorial control over the content they deliver. Cable TV providers have an incentive to make it easier for information consumers to find something interesting to watch: the larger the audience, the more money they make. They do so by limiting their offerings to those that appeal to a majority of their audience. *Content-blind* architectures like the Internet have no single locus of control, and therefore there is no agent for whom an incentive to limit content offerings might exist. Anybody can create a web page or a podcast and make it available on the Internet, but there is no guarantee that it will ever be noticed. Bookmarks in del.icio.us are content-aware: because an information producer has control over his collection of bookmarks, he also

has an incentive to place limits on that collection so that it meets his needs. No such incentive exists for tags, which are content-blind.

Figure 3: This chart depicts the number of respondents who used these strategies for information discovery. The light bars are tag-related strategies, and the dark bars are user-metadata-related strategies
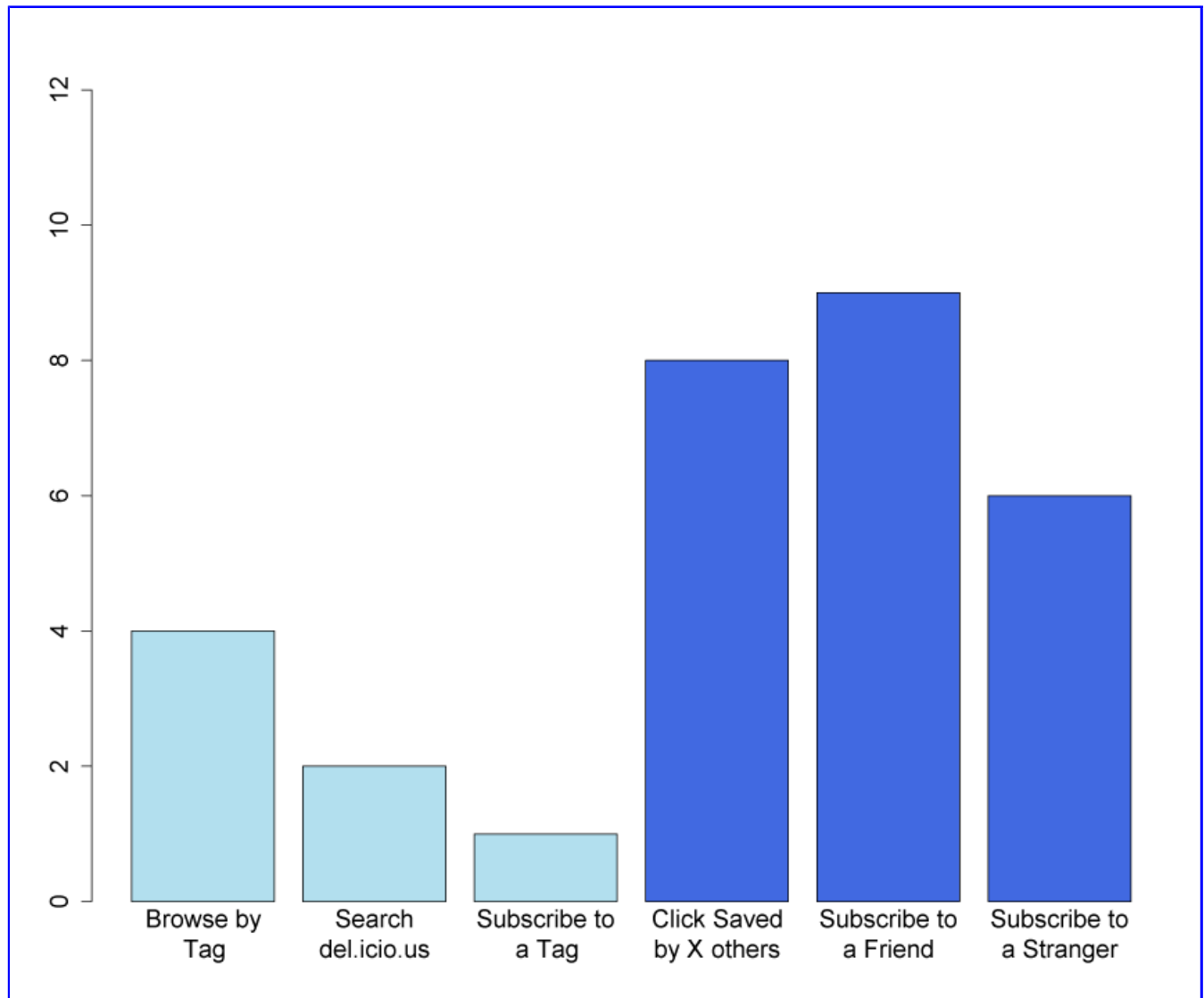


## Tag Convergence and Folksonomy

It might be possible to manipulate incentives in order to promote *tag convergence*, and folksonomy. Tag convergence can be defined in two different ways. The first refers to the convergence on a particular set of popular tags for a given web page (Golder & Huberman, 2006a), which does not control for any of the language use issues pointed out earlier in this section. The second way of thinking about 'tag convergence' is the convergence on the meaning of a particular tag such that it is used only in certain contexts or to refer to specific concepts. This is the sense in which tag convergence can support the creation of a folksonomy. Convergence on the meaning of specific tags could greatly benefit consumers engaged in information seeking by providing a consistent keyword vocabulary.

Online tagging systems offer unique opportunities for users to converge by publicly exposing the previous tag choices of others. Yew, Gibson and Teasley (Yew, Gibson, & Teasley, 2006) found through interviews with users of a required class blog with tags that conventions did in fact emerge among members of a class for tags they applied to the class blog. Sen et al. (Sen et al., 2006b) found that differences in the way the system suggests tags to the user can affect tag choices and cause some level of tag convergence. Both results indicate that increased awareness of other users' tagging choices is likely to increase convergence.

However, there is yet one more wrinkle to be considered. Bookmarks represent an implicit endorsement of a web page; when an information producer chooses to post a web page to del.icio.us, she is in effect making a statement that this particular web page is valuable enough to save for later. There is no similar assumption for tags. Del.icio.us provides information on which tags and bookmarks are currently popular based on aggregate usage statistics; however, this method of creating a valuation for tags fails for topical discovery because popular tags are often the noisiest.

## Conclusion

Del.icio.us users control their bookmark history, and receive benefits for using del.icio.us to store their bookmarks. This has effectively aligned the incentives for information producers with the needs of information consumers. However, no one controls the uses of a tag, and the natural tendencies of language use preclude the applications of tags in a way that is beneficial for information seeking. In this paper we have shown how the design and architecture of a social computing website can influence the choices that users make.

As social computing becomes more pervasive, it is important to understand how such systems can induce users to contribute. Future systems that rely on user-contributed content will need to provide users with an incentive to contribute, and might be able to learn from del.icio.us by providing a private usefulness and having the social nature of the contributed information be an intentional side effect. To better understand incentives in user-contributed content, it would be useful to compare how contributions are motivated across a number of tagging systems, such as del.icio.us, Flickr.com, and Amazon.com. What properties do these incentive mechanisms have in common, and how do their differences influence the amount of contribution? Also, such systems will need to deal with fundamental properties of human language like the vocabulary problem; how might system designers align incentives such that users are motivated to choose tags like others, and form consensus definitions of the meaning of tags?

## Acknowledgements

## References

Ames, M., & Naaman, M. (2007).    Why we tag: Motivations for annotation in mobile and online media.    *Paper presented at the CHI 2007,*   San Jose, CA.

Andreoni, J. (2006).    Philanthropy.    *In S.-C. Kolm & J. M. Ythier (Eds.), Handbook of Giving, Reciprocity and Altruism*   (pp. 1201-1269). Amsterdam: North Holland.

Barreau, D., & Nardi, B. A. (1995).    Finding and reminding: File organization from the desktop.    *SIGCHI Bulletin* 27(3), 39-45.

Brin, S., & Page, L. (1998).    The Anatomy of a Large-Scale Hypertextual Web Search Engine.    *Paper presented at the 7th International World Wide Web Conference*   Brisbane, Australia.

Bruce, H., Jones, W., & Dumais, S. (2004).    Keeping and Re-Finding Information on the Web: What Do People Do And What Do They Need?    *Paper presented at the ASIST 2004: Proceedings of the 67th ASIST annual meeting,* Chicago, IL.

Chang, S.-J., & Rice, R. E. (1993).    Browsing: A multidimensional framework.    *Annual Review of Information Science and Technology,*   28, 231-271.

Feldman, M., & Chuang, J. (2005).    Overcoming Free-Riding Behavior in Peer-to-Peer Systems    *ACM Sigecomm Exchanges,*   6(1).

Forte, A., & Bruckman, A. (2005).    Why Do People Write for Wikipedia? Incentives to Contribute to Open-Content Publishing.   Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.

Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1987).    The vocabulary problem in human-system communication    *Commun. ACM,*  30(11), 964-971.

Golder, S., & Huberman, B. A. (2006).    Usage Patterns of Collaborative Tagging Systems.    *Journal of Information Science,*  32(2), 198-208.

Halpin, H., Robu, V., & Shepherd, H. (2007, May 8-12, 2007).    *The Complex Dynamics of Collaborative Tagging.  Paper presented at the 16th International World Wide Web Conference (WWW 2007),*  Banff, Alberta, Canada.

Harper, F. M., Li, X., Chen, Y., & Konstan, J. (2005).    An Economic Model of User Rating in an Online Recommender System.    *Paper presented at the 10th International Conference on User Modelling*  Edinburgh, UK.

K. Ling, G. Beenen, P. Ludford, X. Wang, K. Chang, D. Cosley, et al. (2005).    Using social psychology to motivate contributions to online communities    *Journal of Computer-Mediated Communication,*  10(4).

Mackie-Mason, J. K., Shenker, S. J., & Varian, H. R. (1995, September 30-October 2, 1995).    Network architecture and content provision: An economic analysis.    *Paper presented at the Telecommunications Policy Research Conference,*  Solomons, MD.

Marlow, C., Naaman, M., boyd, d., & Davis, M. (2006).    Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead.    *Paper presented at the WWW 2006 Collaborative Web Tagging Workshop,*  Edinburgh, Scotland.

Mas-Colell, A., Whinston, M., & Green, J. (1995).    Microeconomic Theory:    *Oxford University Press.*

Miles, M. B., & Huberman, M. (1994).    Qualitative Data Analysis: An Expanded Sourcebook (2nd Edition ed.):    *Sage Publications, Inc.*

Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., et al. (2006, November 4-8, 2006).    tagging, communities, vocabulary, evolution.    *In the Proceedings of CSCW 2006,*  Banff, Alberta, Canada.

Udell, J. (2005, October 20, 2005).    Managing Metadata.   Retrieved September 27, 2006, from http://www.infoworld.com/article/05/10/20/43FEmetadata_1.html

Whittaker, S., & Sidner, C. (1996).    Email overload: exploring personal information management of email.    *In the Proceedings of CHI '96: Human factors in computing systems,*  Vancouver, British Columbia.

Yew, J., Gibson, F. P., & Teasley, S. D. (2006).    Learning By Tagging: Social tagging as a means of group knowledge formation.    *Ann Arbor, MI: University of Michigan.*

Barreau, D., & Nardi, B. A. (1995).    Finding and reminding: File organization from the desktop.    *SIGCHI Bulletin,*  27(3), 39-45.

Brin, S., & Page, L. (1998).    The Anatomy of a Large-Scale Hypertextual Web Search Engine.    *the Proceedings of 7th International World Wide Web Conference,*  Brisbane, Australia.

Bruce, H., Jones, W., & Dumais, S. (2004).    Keeping and Re-Finding Information on the Web: What Do People Do And What Do They Need?    *In the Proceedings of ASIST 2004: Proceedings of the 67th ASIST annual meeting,*  Chicago, IL.

Chang, S.-J., & Rice, R. E. (1993).    Browsing: A multidimensional framework.    *Annual Review of Information Science and Technology,*  28, 231-271.

Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1987).    The vocabulary problem in human-system communication.    *Commun. ACM,*  30(11), 964-971.

Golder, S., & Huberman, B. A. (2006).    The Structure of Collaborative Tagging Systems.    *Journal of Information Science,*  32(2), 198-208.

Mackie-Mason, J. K., Shenker, S. J., & Varian, H. R. (1995, September 30-October 2, 1995).    Network architecture and content provision: An economic analysis.    *In the Proceedings of Telecommunications Policy Research Conference,*  Solomons, MD.

Marlow, C., Naaman, M., boyd, d., & Davis, M. (2006).    Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead.   *In the Proceedings of WWW 2006 Collaborative Web Tagging Workshop,*  Edinburgh, Scotland.

Mas-Colell, A., Whinston, M., & Green, J. (1995).    Microeconomic Theory:    *Oxford University Press.*

Miles, M. B., & Huberman, M. (1994).    Qualitative Data Analysis: An Expanded Sourcebook (2nd Edition ed.):   *Sage Publications, Inc.*

Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., et al. (2006, November 4-8, 2006).   tagging, communities, vocabulary, evolution    *In the Proceedings of CSCW 2006,*  Banff, Alberta, Canada.

Smith, V. (1980).    Experiments with a decentralized mechanism for public goods decision.    *American Economic Review,*  70, 584-599.

Udell, J. (2005, October 20, 2005).    Managing Metadata.  Retrieved September 27, 2006, from  http://www.infoworld.com/article/05/10/20/43FEmetadata_1.html

Whittaker, S., & Sidner, C. (1996).    Email overload: exploring personal information management of email.    *In the Proceedings of CHI '96: Human factors in computing systems,*  Vancouver, British Columbia.

Yew, J., Gibson, F. P., & Teasley, S. D. (2006).    Learning By Tagging: Social tagging as a means of group knowledge formation.    *Ann Arbor, MI: University of Michigan.*

Corresponding Author

[i]In del.icio.us, all bookmarks and associated metadata are public information by default, and can be browsed by anyone with an Internet connection and web browser.

[ii]http://www.atlasti.com/

[iii]It is not immediately clear to users how one can make best use of del.icio.us. Indeed, all twelve respondents talked of an exploratory period where they attempted to figure out how to make del.icio.us useful for them. Actions that were undertaken during this period and never since (only once or twice, and in the past) are excluded from the findings reported below. We are primarily concerned with how people regularly use del.icio.us, not this learning period. The effects of this learning period are an open question.

[iv]All names in this paper are pseudonyms, and statements have been anonymized to protect the identities of our respondents.

[v]A full explanation can be found at http://del.icio.us/help/for

[vi]Del.icio.us now makes this information available in the "your fans" section on the network page, but this was a new addition to the interface not long before the interviews took place.