An Elementary Social Information Foraging Model

Peter Pirolli
Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto CA, 94304
pirolli@parc.com

ABSTRACT
User interfaces and information systems have become increasingly social in recent years, aimed at supporting the decentralized, cooperative production and use of content. A theory that predicts the impact of interface and interaction designs on such factors as participation rates and knowledge discovery is likely to be useful. This paper reviews a variety of observed phenomena in social information foraging and sketches a framework extending Information Foraging Theory towards making predictions about the effects of diversity, interference, and cost-of-effort on performance time, participation rates, and utility of discoveries.

Author Keywords
Social information foraging theory.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

At the end of World War II, the United States' Director of Scientific Research and Development, Vannevar Bush, published an article entitled *As we may think* in the popular magazine *Atlantic Monthly* [5]. The article was both reflective and visionary: The war had been won with the help of the coordinated efforts of scientists, and those resources were available to be applied to the development of new tools to extend the powers of the human mind through more facile command of recorded knowledge. *As we may think* presented seminal ideas about personal computing and hypermedia that inspired computer scientists to realize aspects of Bush's vision. The problem that inspired Bush was intellectual overspecialization. To solve this problem, Bush envisioned a device he called the Memex that would allow scholars to forage through personal stores of multimedia documents, and to save *traces* of paths through content that could then be *shared* with other scholars as a way of communicating new findings. The Memex was envisioned as a tool that would increase the capacity of individuals to attend to greater spans of emerging knowledge, and would increase the cooperative information sharing that Bush viewed as necessary to improvements in scientific discovery, which he expected to result in increased benefits to society. Bush's vision was not only to improve the information foraging ability of the individual user, but to also improve communication and collaboration.

With the rise of the Internet, Web, Web 2.0 and mobile communication we are witnessing the emergence of a decentralized network of knowledge production, sharing, and use that realizes the spirit of Bush's dream, if not the specifics. This paper is an attempt to sketch out a theoretical foundation for some phenomena that arise as the field of computer-human interaction shifts from concerns focused mainly on the solo user to concerns about the social phenomena that arise from many interacting users. This paper is an attempt to extend Information Foraging Theory [23] to predict the effects of diversity and social brokerage, the standing-on-the-shoulders-of-giants effect, the effects of social interference, and the role of user interface interaction costs.

Information foraging theory [23] has mainly focused on information seeking by the solitary user. The discovery of new knowledge, innovations, or inventions, however, is almost universally the result of collective action. This paper (based on a chapter in [23]) presents a somewhat idiosyncratic review of models and findings in various fields that may provide the basis for the development of theories of foraging by collectives, whether constituted by formal organizational structures or informal networks. The theoretical sketch draws upon research in optimal foraging theory (especially [14]), library science, computational ecology, management science, and sociometrics. Across these disciplines one finds general results concerning the costs and benefits of cooperative information foraging, the effects of group diversity, and patterns of social structuring that are correlated with innovative discovery.

This paper presents mathematical models that capture some basic elements of these results. The goal of the models, at this point, is not to capture all the details of social information foraging in all of their messy splendor. Rather, the hope is that the models, which are surely wrong, provide some insights to the main factors and trade-offs that structure the signature phenomena of social information foraging. To paraphrase R.M. May [21]: Here mathematics is seen in its quintessence, no more, but no less, than a way of thinking clearly about the consequences of basic assumptions in comparison to key empirical facts.

THE POWER OF COOPERATION

Specialization is a natural consequence of too much public knowledge for the individual mind to comprehend. Social networks involved in knowledge discovery, such as scientific communities, typically self-organize into a cognitive division of labor, with divisions based on the deliberate exclusion of possibly relevant information [31]. Some knowledge discovery organizations, such as the U.S. intelligence agencies are formally (and technologically) organized into specialty areas. The worry is that knowledge specialization leads to situations in which all the information required to make an important discovery is in the available record somewhere, but it is distributed across specialization boundaries with no single set of eyes in a position to see it all and make sense of it.

It is unlikely that we can estimate the number of discoveries that are latent in the public domain because of overspecialization. This unrealized potential, however, has received considerable attention in the information retrieval and library sciences, where it is known as the *undiscovered public knowledge problem* [29, 30]. We may characterize public knowledge [30] by making use of the knowledge level perspective [22]. Public knowledge is that which is directly recorded in publications, plus the implications of that knowledge (i.e., the implicative closure of recorded knowledge). The problem is that some of the implied knowledge may be undiscovered. These implications may include hidden refutations, hidden cumulative strength of individually weak studies, or other hidden links in the logic of discovery.

Overspecialization may lead to a failure for any one mind to grasp and connect all the dots. Information sharing is usually recognized as a strategy for extending the grasp of the solitary mind across specializations, to reduce the risk of failing to make discoveries implicit in the existing literature. To the extent that individual members of a core specialty can devote some effort to exploring related peripheral specialties, and sharing possible leads with others, then one might expect the group to perform more effectively.

Pirolli and Card [24] describe a business intelligence agency whose analysts were tasked to write monthly newsletters about core areas such as computer science or materials science. The main purpose of those newsletters was to identify new important science and technology trends. The organization received about 600 magazines, trade publications, and journals each month, and each analyst was responsible for scanning about 50 of these publications (an estimated 500 articles per month). In addition to culling material for their own newsletters, analysts would also notice articles pertinent to the specialties of other analysts, and would have such articles copied and routed to the appropriate specialist. An analyst would typically receive about 6 – 12 relevant articles per month from other analysts, at very little cost. The general belief of the analysts was that such cooperation enhanced the individuals' search capabilities, and reduced the risk of missing something relevant to a specialty area that had emerged in a non-specialty publication. Below, I discuss a more thorough study by Sandstrom [26] concerning information foraging by an informal network of scientists that exhibits a similar, though more intricate, pattern of information sharing.

SOCIAL CAPITAL OF DIVERSITY AND BROKERAGE

Cooperation may yield more benefits than simply making information search more parallel and making it less prone to failure. Membership in a group provides actual or potential resources that can be utilized or mobilized to achieve individual goals. This is known as *social capital* [2, 25], and much research has focused on determining what aspects of social structure provide such capital. Exposure to a greater diversity of knowledge, hence more novel ideas as a function of the time cost invested in information foraging, is another potential benefit of cooperation. Below, I summarize research on the effects of group diversity on cooperative information foraging, as well as the theory that people who provide brokerage of ideas across social clusters are often in position to make valuable novel discoveries. Since scientists (and many others who do knowledge discovery) are rewarded according to the novelty and value of their discoveries, we might expect individual group members to try to arrange themselves to be in positions that broker knowledge from peripheral fields to the group's core specialty field. One question that arises is why groups do not grow infinitely large (if cooperation and group size produces positive rewards). Below I discuss forces that may be involved in producing groups of stable sizes.

INFORMATION FORAGING BY NETWORKS OF SCHOLARS

Sandstrom [26] has studied information foraging and scholarly communication among a group of scientists in the field of behavioral ecology. Sandstrom's research combined a bibliometric approach with structured interviews. Sandstrom [26] first developed a spatial representation of scholarly publications in behavioral ecology, then had individual researchers in the community identify what they considered to be their core specialty vs peripheral specialties. Additional data were used to distinguish foraging strategies associated with the core vs periphery, and to suggest the costs and benefits of cooperation and the spanning of specialty literatures.

Sandstom [26] performed an author cocitation analysis to understand the intellectual structure and scholarly communication patterns of 63 authors selected from the

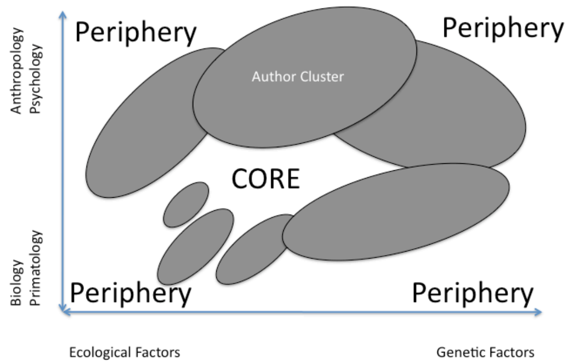

Figure 1. Schematic summary of the map of the field of behavioral ecology produced by a 2-D multidimensional scaling solution of an author cocitation analysis in Sandstrom [26]. Clusters of related authors were found to be arranged around the semantic core of the field, and many had publications in peripheral fields.

behavioral ecology literature for 1988-1995. A multidimensional scaling analysis provided a visualization that could be interpreted as different subfields of authors arranged spatially in a 2D semantic pace, which is presented schematically in Figure 1. Sandstorm identified the center of this semantic space as the *core* of the field, with other, related fields arranged around the periphery. A sizable number (27 out of 55) authors contributed to more than one research area. As discussed below, these authors could be viewed as brokers of knowledge from one field to another.

Sandstrom [26] also asked five experts to answer questions about their information foraging strategies [using the information seeking strategies defined in, 13] for literature in the own core field and peripheral fields. Socially mediated discovery tended to be the source of core literature for the experts, whereas solitary foraging tended to be the source of peripheral literature. Specifically, recommendations from colleagues, papers sent for prepublication reviews, and reprints sent by other authors and editors accounted for 30% of experts' referenced items. Of these items, 69% were identified as belonging to their core field. Solitary foraging, involving reading, following references (citation chaining), browsing, monitoring, or deliberate search accounted for 48% of experts referenced items. Of these, 61% were identified as belonging to peripheral fields. Communication in the core zone of this scientific field tends to occur through more social means, and more towards the prepublication stages, whereas foraging interactions in the peripheral zones tends to involve more solitary or formal mechanisms, and more towards the postpublication stages. Low-cost information foraging behaviors are associated with core zones, and high-cost information foraging behaviors are associated with peripheral zones.

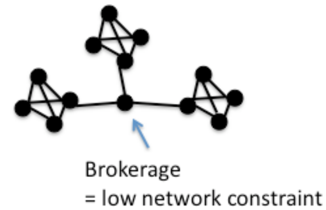

Figure 2. Schematic summary of Burt's [4] notion of network constraint. Sociometric analysis of social networks can be used to reveal individuals in brokerage positions between densely connected network clusters.

EFFECTS OF DIVERSITY AND THE BROKERAGE OF STRUCTURAL HOLES IN SOCIAL NETWORKS

Homogeneity of opinion, viewpoint, and information resources among a group of information foragers is likely to produce redundancy in what they find and how they interpret those findings. We might expect that groups of cooperative information foragers will be more effective if constituted by individuals with some degree of diversity. Individual foragers, who are positioned in social networks such that they broker information and ideas across groups, might be exposed to a greater diversity of information themselves, and be a conduit to greater diversity for their colleagues.

Organization and management studies [12] suggest that effective work groups are ones that share information and know-how with external members, and that effectiveness is improved by *structural diversity* of the group. Structural diversity is variability in features of the group that expose members to different sources of task information, know-how, and feedback. Such features include geographic locations, functional assignments, number of mangers to whom members report, and number of business units associated with the group. Cummins [12] studied 182 work groups in a Fortune 500 telecommunications firm and found that work group performance (as rated by senior executives) was significantly correlated with an interaction of structural diversity factors with knowledge sharing factors.

The findings of Cummings [12] are consistent with the theory of social structural holes (*structural holes theory*) proposed by Burt [4]. Structural hole theory is grounded in the analysis of social networks as revealed, for instance, by sociograms such as Figure 2 that capture information flow. The nodes in Figure 2 represent people or aggregate groups of people and the links represent information flow. Typically, such social networks of information flow will contain densely connected clusters. The sparse linkages between such clusters constitute *structural holes*. People who bridge such structural holes have an advantage of exposure to greater diversity of information and know-how, and *brokerage* across structural holes becomes a form of *social capital* that translates into the discovery of greater amounts of useful, productive knowledge. The node at the center of Figure 2 represents an individual X who belong to three separate densely connected clusters of individuals.

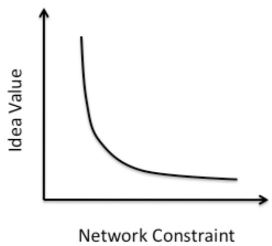

Figure 3. Schematic summary of Burt's [4] analysis of the value of ideas generated by managers as a function of their network constraint measure.

Individual X is better positioned to gain from the social capital of brokerage across structural holes between those clusters of individual and is positioned to introduce greater diversity into groups. As Burt [4] summarized:

Given greater homogeneity within than between groups, people whose networks bridge the structural holes between groups have earlier access to a broader diversity of information and have experience in translating information across groups. This is the social capital of brokerage…People whose networks bridge the structural holes between groups have an advantage in detecting and developing rewarding opportunities. Information arbitrage is their advantage. They are able to see early, see more broadly, and translate information across groups. Like over-the-horizon radar in an airplane, or an MRI in a medical procedure, brokerage across the structural holes between groups provides a vision of options otherwise unseen.

One of the exciting prospects for the study of social information foraging is the explosion of online data relevant to finding and measuring social networks using on-line resources. For instance, it appears that e-mail flow and Web links among personal home pages provide data that can be used to accurately construct social networks [18] and to study information flow.

A TEST OF THE RELATIONSHIP OF BROKERAGE TO INNOVATION AND INCENTIVES

Two important empirical questions arising from structural holes theory concern whether in fact (a) good ideas arise from the social capital of brokerage and (b) whether individuals are incented to work their way into brokerage positions. To answer these questions, Burt [4] studied 673 managers in the supply chain of a large American electronics company. Burt constructed a social network using a standard survey method.

To assess the quality of ideas, Burt conducted a survey in which managers were asked to generate ideas that would improve the company's supply chain management. These ideas were submitted to two senior managers who were asked to evaluate the value of the generated ideas. The evaluation. Burt also measured the *network constraint* of individuals, which is an index of the investments (e.g., time, effort, resources) of a person in direct and indirect social relationships. Low network constraint is associated with brokerage across structural holes (as in Figure 2). The hypothesis was that individuals in brokerage positions (low network constraint) would generate more valuable ideas. Burt [4] found a relationship between network constraint (brokerage) of individuals the value of ideas they generated found a relationship that is captured graphically in Figure 3. Burt [4] also found that network constraint predicted extra salary rewards for individuals: Overall, managers who discussed issues with managers in other groups were not only better paid, but were also likely to receive more positive job evaluations and to be promoted.

Burt's results suggest that brokerage across social network clusters is associated with higher valued ideas and greater rewards (Burt presents many other data analyses that support these conclusions). One may wonder, however, why social networks do not evolve such that the network constraint is uniform throughout, for all people, given that brokerage (low network constraint) appears to be individually rewarding. One possibility is that extra-group cooperation may be substantially more resource intensive and risky than intra-group cooperation, and people vary in their ability to create and maintain extra-group cooperation.

The notion that brokerage across groups is important to success is echoed in many other domains. It seems plausible that the scientists studied by Sandstrom [26], in which each scientist spans the core literature of the field in addition to idiosyncratic peripheral areas, might arise from an incentive structure that rewards brokerage of structural holes in the flow of information, know-how, and ideas.

A BASIC MODEL OF SOCIAL INFORMATION FORAGING

I have presented evidence that indicates that information foragers, typified by scientists, engage in social exchanges of information, and appear to arrange themselves such that they bridge across content areas and informal social networks. Such arrangements may be expected to expose the individual to a greater diversity of hints about where to focus their foraging and sense making efforts. Research in sociology and management science indicates that the exposure to diversity that arises from bridging social structural holes is associated with innovation and greater individual rewards. One key assumption made here, is that many of these phenomena will be observed again in the emergent mobile and Web 2.0 worlds.

In this section, I draw upon work in optimal foraging theory and computational ecology to develop a very simple basic model of the costs and benefits of cooperative information foraging. Diversity among information foragers is a critical variable in this model. At the end of this section I discuss the issue of group size.

The basic social information foraging model (*basic SIF model*) derives from the quantitative theory of cooperative problem solving developed by Clearwater, Hogg, and Huberman [10]. Many extensions of this model have been developed and tested in computational ecology [e.g., 11, 15, 19], so it is likely that the basic SIF model can be refined to meet many alternative constraints and assumptions. The work of Clearwater et al. [11] focused mostly on the analysis of the benefits of cooperative search processes. The basic SIF model incorporates some simple general assumptions about the nature of interference costs that arise in cooperation based on the computational ecology studies of group foraging in Seth [27]. Finally, the basic SIF model is cast in the form of the group foraging models developed in Clark and Mangel [9], which can be used to understand the relation between the size of a group and the individual rewards of cooperation, and which can also be used to understand why the expected size of groups will tend to be larger than optimal

Basic Search Assumptions

The basic SIF model assumes a heuristic process of search for useful knowledge in a space of discrete patches of information. It is assumed that a patch of information will yield some amount of utility for one or more foragers. To relate this heuristic search process to time, t, it is assumed that the number of processing steps required to find useful patches of information is large, the processing steps occur as a Poisson process, with each step occurring at rate λ_s steps per unit of time. The information environment can be characterized by the expected number of steps, T, required to find the next useful patch of information by random search process (i.e., with no heuristic involved and no cooperation). For this unguided, non-cooperative search process, the probability, p, of encountering a valuable information patch is,

$$p = 1/T, \tag{1}$$

and, because of the Poisson process assumptions, the probability density function for encountering a valuable information patch as a function of time is

$$P_{Find}(t) = \lambda_s \, p \, e^{-\lambda_s \, p \, t} .. \tag{2}$$

The expected time to find a patch will be

$$t_{Patch} = \int_0^\infty t \, P_{Find}(t) \, dt, \tag{3}$$

which is

$$\begin{aligned} t_{Patch} &= \frac{1}{\lambda_s \, p} \\ &= \frac{T}{\lambda_s} \end{aligned}. \tag{4}$$

Heuristics and Hints

The search heuristic of the individual information forager, i, can be characterized by the proportion, h_i, of remaining search steps that are eliminated [10, 17]. A heuristic of h_i = 0 is perfect and a heuristic of h_i = 1 moves the forager no closer or farther from finding a useful information patch. The number of steps required to find a useful information patch will be $h_i \, T$. The average time to find a patch for the heuristically guided, non-cooperating information forager will be,

$$t_{Patch} = \frac{h_i \, T}{\lambda_s}. \tag{5}$$

The basic SIF model assumes that heuristic *hints* are exchanged in cooperative information foraging regarding the likely location of useful information patches (for instance, as was observed to occur among analysts in the business intelligence agency discussed above). Another example of hints are the social tags in systems such as delicious.com. Shared tags provide navigation paths and ontological organization to available content and it seems to be assumed that sharing tags improves individual sensemaking and foraging.

Hints from cooperating information foragers may be characterized by the proportion, h_{ji}, of remaining search steps that are eliminated by the j^{th} distinct hint received by information forager i. Hints may vary in the validity of the search information conveyed, may vary in how they are interpreted by the information forager who receives them, and may vary in effectiveness depending on when they are exchanged in the search process. For instance, a good hint received late or not utilized will have a smaller effect than the same hint utilized early in search process. Similarly, to the extent that hints may contain redundant (correlated) search information, the effectiveness of hints will depend on what hints have already been processed. The h_{ji} should be interpreted as the *distinct* or *independent* heuristic effectiveness of a given hint given these conditions. We might also expect that as hints continue to arrive, they eventually repeat earlier information, and consequently yield no additional heuristic value in further reducing the search space. This is modeled simply by assuming that there is some maximum number of distinct effective hints, H. The expected number of steps required to find a useful information patch is defined to be,

$$k = h_i \prod_{j=1}^{H} h_{ji} T . \tag{6}$$

The average time to find a patch for a heuristically guided cooperating information forager will be,

$$t_{Patch} = \frac{h_i \prod_{j=1}^{H} h_{ji} T}{\lambda_s}. \tag{7}$$

Huberman [17] presents a derivation of the law—that I repeat here—that relates the diversity of effective hints to the distribution of number of steps required to successfully complete a search. Taking the logarithms of the effective values of the hints of a forager one gets,

$$\ln\left(\prod_{j=1}^{H} h_{ji}\right) = \ln(h_{1i}) + \ln(h_{2i}) + \cdots + \ln(h_{Hi}). \tag{8}$$

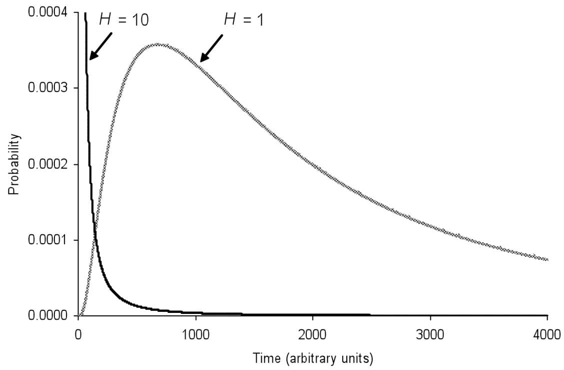

Figure 4. Probability density function for finding useful information at time *t*. As diversity increases from *H* = 1 to *H* = 10 it becomes more likely that useful information will be found sooner. The illustration assumes $T = 10000$, $\lambda_s = 1$, $h_i = .5$, $\mu = -1$, and $\sigma = 1$.

If the individual distributions of each of terms on the right side of Equation 8 have finite variance, and the number of hints is large, then the Central Limit Theorem applies and the logarithms of the hints, $\ln(h_{ji})$, will be normally distributed with mean μ and variance σ^2. Therefore, the distribution of the h_{ji} themselves will follow a lognormal distribution, which has a probability density function

$$\Lambda(\mu,\sigma,x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-(\ln x-\mu)^2}{2\sigma^2}}, \tag{9}$$

and an expected value

$$E\left[h_{ji} \mid \mu,\sigma\right] = e^{\mu+\frac{\sigma^2}{2}} \tag{10}$$

with variance

$$\mathrm{var}\left[h_{ji} \mid \mu,\sigma\right] = \left(e^{\sigma^2} - 1\right)^2 e^{2\mu+\sigma^2}. \tag{11}$$

The properties of the normal distribution imply that a sample of size H of the logarithms of the independent hint values in Equation 8 will have a mean $H\mu$ and variance $H\sigma^2$.

The probability density function for finding valuable information patch can be characterized as a lognormal distribution,

$$P_{Find}(t) \approx \Lambda\left(H\mu + \ln\frac{h_i T}{\lambda_s}, \sqrt{H}\sigma, t\right), \tag{12}$$

and the rate of finding valuable information patches can be characterized as a function, $\lambda(H)$, of the diversity of hints,

$$\lambda(H) \approx \frac{1}{E\left(H\mu + \ln\frac{h_i T}{\lambda_s}, \sqrt{H}\sigma\right)}. \tag{13}$$

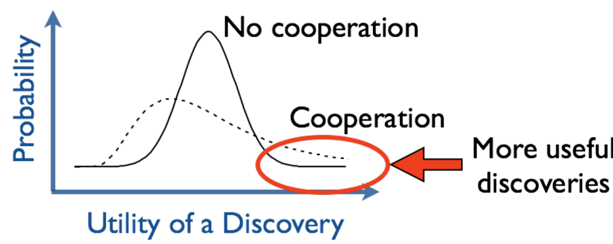

Figure 5. Increasing cooperation alters the probability distribution of achieving search results of different values.

The average time to find a valuable information patch is

$$t_{Patch}(H) \approx \frac{1}{\lambda(H)}. \tag{14}$$

Diversity Increases the Likelihood of Discovery

Figure 4 illustrates the prediction of the model with respect to the impact of cooperative information processing on increasing the likelihood that more useful information will be discovered sooner. One can see a basic lognormal distribution of performance times that shifts downwards as more useful hints are exchanged among participants in a cooperative.

Cooperation Increases the Likelihood of High-Value Discoveries

As discussed in greater detail in Huberman [17], the lognormal distribution of performance times makes interesting predictions about productivity of a cooperative group with respect to high-utility search results (see Figure 5). If one assumes that the various states of a search space have a binomial distribution of utilities, then, assuming a mildly effective search heuristic, the search performed by non-cooperating searchers will return a distribution of result values shown in Figure 5. Increasing cooperation will shift that distribution to a lognormal, and will especially increase the likelihood of search results at the higher end of the utility spectrum. In a sense this accounts for the "standing-on-the-shoulders-of-giants" effect that is frequently observed where an individual with an average amount of smarts benefits from the cooperation of others in making a better-than-average discovery.

Optimal Group Size, Interference Effects, and Equilibrium Group Size

Many species, besides humans, forage in groups. Flocks of birds are perhaps the most obvious example of group foraging. One general explanation for the evolutionary advantage of group foraging is that it may lead to improved use of information about food sources in scarce, patchy environments [8, 9]. Although there may be positive effects of foraging in a group, foraging groups do not become arbitrarily large, suggesting that there may be some form of interference costs (e.g., intra-group competition) that at some point outweighs the advantages of further increments in the size of groups.

It has been found empirically [16] and in computational modeling [27] that there is often a power-law relationship between the number of foragers, n, in a patch and the rate of

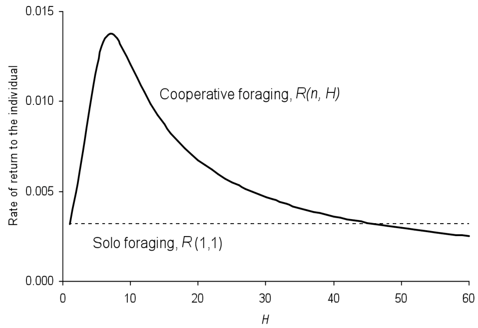

Figure 6. The individual rate of gain, $R(n, H)$ may have a peaked form when interference costs are included. It is assumed that $G = 10$, $\tau(n) = 100\ n^{.9}$, $n = H$. The dashed horizontal line indicates the rate of return for the solitary forager $R(1,1)$. The optimum size of the group is $n^* = 7$, whereas the equilibrium size of the group is $\tilde{n} = 45$.

consumption intake by each forager. To capture this mathematically one might assume that the individual forager's time to process an information patch in a group of n foragers is

$$\tau(n) = a\ n^c \qquad (15)$$

where $0 < c < 1$ is a rate parameter, and a is the time to forage for a patch when $n = 1$.

We may assume a model in which an information patch has some finite total amount of value, G, where the expected gain for each of the n agents in the patch is G/n. This might characterize the case in which there are a finite number of discoveries to be made in a domain, and once a particular discovery is made it is of has no additional value to the next forager (as happens in scientific publication, where authors gain no reputation for repeating the discoveries of others). This is admittedly a strong assumption that does not apply to all group foraging situations. The expected time for n agents to find a valuable information patch will be $t_{Patch}(H)/n$, or equivalently $1/[n\ \lambda(H)]$. When n agents forage simultaneously, the patch will be exhausted in $\tau(n)/n$ time units. We may now cast the basic the basic SIF model as a variation of a conventional foraging models [see also, 9]. The rate of gain, for the individual member of the group, is

$$R(n,H) = \frac{\lambda(n)G}{1 + \lambda(H) + \tau(n)} \qquad (16)$$

Clark and Mangel [9] discuss the relationship of interference effects to optimal and equilibrium group size. The solitary forager should choose to join a group if the expected returns for group foraging are greater than foraging alone, i.e.,

$$R(n,H) > R(1,1). \qquad (17)$$

If there is an interference cost function of the form in Equation 15, then we may obtain peaked rate of gain functions such as the one illustrated in Figure 6.

Figure 6 can also be used to discuss why the equilibrium group size, $\tilde{n}$, may be greater than the optimum group size, n^*. Suppose solitary foragers have joined a group until it has the optimum size n^*. Solitary foragers should continue to join the group so long as the rate of return for group foraging is still above the rate of return for solitary foraging, as stated in Equation 17. Members of the group may see their individual rates of return diminish from the optimum as new members join the group, but remaining in the group is still better than solo foraging. Consequently, individuals will join the group until the addition of new members makes the individual rate of return less than solitary foraging. Consequently, when $R(n, H)$ is peaked, as in Figure 6, we may expect the equilibrium size to be $\tilde{n} > n^*$.

One can raise the question as to the plausibility of cooperative interference effects in information foraging. One example of this [7] is the apparent decrease in social tag effectiveness that seems to occur as the number of tags (and presumably taggers) grows. More specifically, it appears that a general phenomena across social tagging sites (see Figure 7 for the delicious.com data) that the mutual information between tags and documents decreases over time, as participation and tagging increase. This means that each individual tag is less effective (has less information) over time.

Another illustration of the growing interference appears to occur in Wikipedia [20]. Figure 8 shows how the proportion of effort devoted to actually producing article content has declined from about 95% to less than 65% over time.

One can ask whether these cooperative interference effects have any impact on participation rates, and such direct studies remain to be done. However, it is interesting that the number of editors actively contributing to Wikipedia has

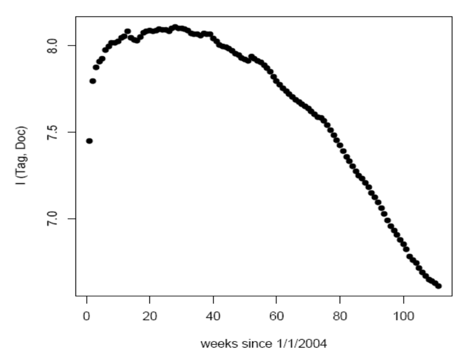

Figure 7. The mutual information $I(tag, doc)$ has been observed to decrease over time at delicious.com. From [7].

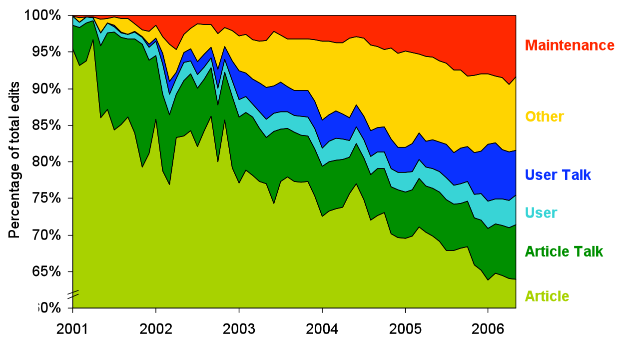

Figure 8. The proportion of total editing effort going to article pages has declined over time, as the Wikipedia user base increased. From [20].

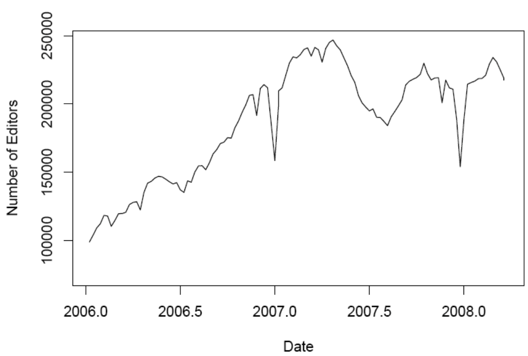

Figure 9. The number of editors contributing per week to Wikipedia.

apparently plateaued (see Figure 9), and this effect can be observed for pages created recently as well as those created some time ago.

Reducing Interference Costs is Predicted to Increase Participation

One implication of the model presented graphically in Figure 6 is that any changes to the technology or policies that reduce the impact of having to deal with others should increase the overall participation rates. This effect is summarized in Figure 10. Reducing the costs of cooperation extends the tail of the rate of returns curve, which also extends the point at which it crosses the solo foraging threshold. Consequently the equilibrium group size is predicted to increase.

Decreasing the Cost-of-Effort of a User Interface is Predicted to Increase Production and Participation

In many social Web applications, such as social tagging or Wikipedia, there are interaction costs associated with producing and sharing information products (e.g., tags; Wikipedia content). In many situations, information foraging theory predicts that productivity rates should increase, and time allocations decrease, as the cost-of-effort associated with producing knowledge with the user interface is decreased.

The information patch model [23] can extended to model tag production. One can assume a simple characterization of the tag producer's task as involving a trade-off between reading + interaction time vs tag production time. In the case of the tagging model, one can assume that the user's tagging activity around an individual article constitutes a "patch" of productive activity of some value to the user.

Imagine an idealized user who navigates the Web and reads articles. This idealized user iteratively navigates to a page, reads it, and moves onto the next. Now assume that this idealized user is also engaged in tagging the articles that were read. For each Web page, the user engages in a set of micro-tasks around the addition of tags (e.g., generating the tag from memory or from the just-read text and somehow entering it into a tagging system). So this idealized user's time can be divided into time devoted to (a) interaction (navigation) plus reading and (b) tag generation. On each article, the user spends some amount of time, on average, engaged in tag generation activities, and this is called the average *within-patch* time, t_W. The user also spends some amount of time, on average, navigating to the article and reading it, and this is called the average *between-patch* time, t_B.

This tagging-patch model assumes that tag production on a particular article produces diminishing returns as a function of time. In other words, on average, as time progresses, the user generates tags at ever-diminishing rate. The cumulative production of tags on an average article may be characterized by a gain function, $g(t_w)$.The overall average rate of gain from tag production is

$$R = \frac{G}{T_B + T_W} \quad (18)$$

where G is the sum of all the gains from all tagging, T_B is the total between-patch time (reading and interaction time in the current case) and T_W is the total within-patch time (tagging in this case). Under some strong but relatively

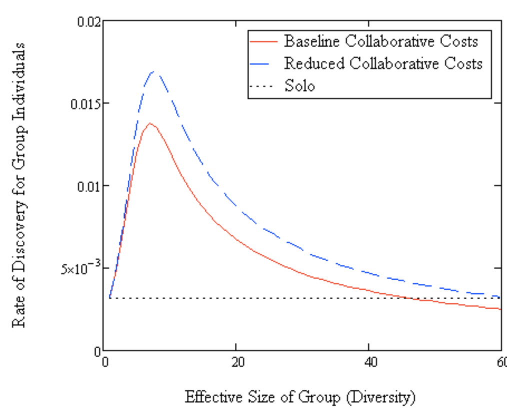

Figure 10. Reducing the costs of cooperation extends the tail of the rate of returns curve, which also extends the point at which it crosses the solo foraging threshold. Consequently the equilibrium group size is predicted to increase.

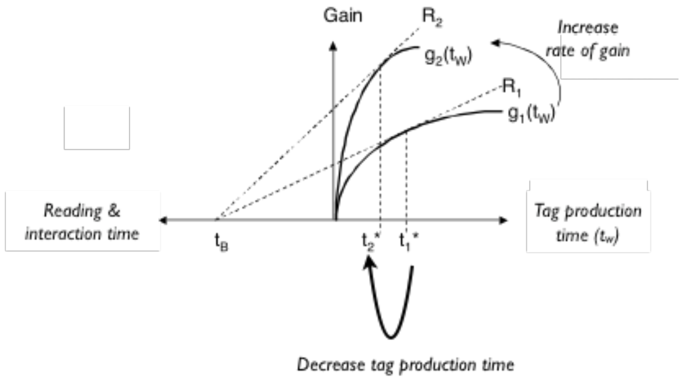

Figure 11. An information patch model. Charnov's Marginal Value Theorem states that the rate-maximizing time to spend in patch, t*, occurs when the slope of the within-patch gain function g is equal to the average rate of gain, which is the slope of the tangent line R, The average rate of gain, R, increases with improvements in the gain function, while simultaneously decreasing the optimal time to allocate to tag production.

general assumptions [6, 23] the optimal time allocation to spend tagging is t_W^*:

$$R = g'(t_W^*) \tag{19}$$

where g' is the marginal rate of (within-patch) tag production. Equation 18 captures the rule: Continue tag production until the marginal rate of gain for continued tagging drops below the overall rate of gain R.

Figure 11 presents a graphical representation of this model familiar in optimal foraging theory [28]. Between-patch time is plotted horizontally from the origin to the left and within-patch time is plotted form the origin to the right. The curve g_1 represents a hypothetical diminishing returns function for tag production. A line plotted from the intercept t_B to a point tangent to g_1 will have a slope equal to the overall average rate of gain from tag production R, and the point of tangency to g_1 will be $g_1(t_W^*)$, thus giving us the optimal average time t_W^* to allocate to tagging.

Figure 11 also includes another gain function, g_2 that represents the effects of lower time cost associated with producing tags. Going through the same graphical solution of plotting a tangent line to g_2, one can see that the optimal time allocation to tagging is reduced while increasing the overall rate of gain R and increasing the number of tags produced.

The information patch model predicts that lowering the time cost of tag production will increase the number of tags produced per document by individuals while decreasing the amount of time spent tagging. This is also a basic assumption of Benkler [1].

Figure 12 shows an application of the information patch model to data collected in an experiment conducted in our lab contrasting a lower cost-of-effort tagging technique called Click2Tag [3] against a more generally used type-to-tag technique found in such systems as delicious.com.

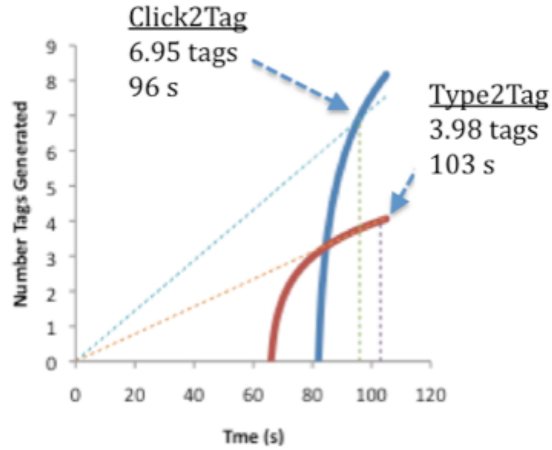

Figure 12. Fit of information patch model to observed average reading+tagging times and average number of tags produced assuming logarithmic gain functions.

Although not a perfect instantiation of the model in Figure 11, it is a very close approximation to the observed relations.

GENERAL DISCUSSION

Like the conventional foraging models presented in Pirolli [23] the basic SIF model is surely wrong. However, it serves as a tool to reason generally about several aspects of the power of cooperation and the social capital that is relevant to finding information. The model suggests that so long as the diversity of agents increases with group size, then the size of a group increases the overall power of cooperative discovery. As individual foragers increase the diversity of their cooperating contacts they will improve in performance. This provides a mathematical rationale for the idea that brokerage positions in social networks provide social capital. The model also provides a rationale for the observed lognormal distribution of innovative discoveries.

In communities of practice that depend on foraging in overly rich information environments, there appears to be pressure to self-organize into a balance of some division of labor, plus some degree of cooperation. This was evident in the study of social information foraging among scholars. The division of labor is necessary because of the limits of human attention, but some investment in cooperation can lead to increased returns and less risk of missing something important. The power of cooperation is related to the amount of diversity of the information foragers. Greater diversity leads to greater returns for the group and the individual. This is related to the notion that brokerage (diverse social contacts) provides social capital, and there is evidence that brokers in the flow of information are more likely to be sources of innovative discoveries. Although there are benefits to cooperation, those benefits trade against interference effects that ultimately seem to limit the size of groups. In addition, because of the diversity of individuals, and because of the way people associate with like-minded people, information is typically likely to flow to small finite sized groups.

A variety of technologies have emerged to exploit or enhance social information foraging. Web, blogs, email, internet groups, collaborative tagging, wikis, recommender systems, and other technologies are all aimed at supporting cooperative information sharing and their success implies their effectiveness. Given the increased ease with which it is possible to study social networks and information flow in the electronic world, it is likely that there will be more studies of the effects of technologies on social structure and social capital, hence a need for a suitable theoretical framework.

ACKNOWLEDGEMENTS

This work was supported in part by Office of Naval Research Contract No. N00014-08-C-0029 to Peter Pirolli

REFERENCES

1. Benkler, Y. The wealth of networks: How social production transforms markets and freedom. Yale University Press, New Haven, CT, 2005.
2. Bourdieu, P. The forms of capital. in Richardson, J.G. ed. Handbook of theory and research in the sociology of education, Greenwald Press, New York, 1986.
3. Budiu, R., Pirolli, P. and Hong, L., Remembrance of things tagged: How tagging effort affects tag production and human memory. in CHI 2009 Conference on Human Factors in Computing Systems, (Boston, MA, 2009), ACM.
4. Burt, R.S. Structural holes and good ideas. American Journal of Sociology, 110 (2). 349-399.
5. Bush, V. As we may think. Atlantic Monthly, 176. 101-108.
6. Charnov, E.L. Optimal foraging: The marginal value theorem. Theoretical Population Biology, 9. 129-136.
7. Chi, E.H. and Mytkowicz, T. Understanding the efficiency of social tagging systems using information theory Proceedings of the nineteenth ACM conference on Hypertext and hypermedia, ACM, Pittsburgh, PA, USA, 2008.
8. Clark, C.W. and Mangel, M. Foraging and flocking strategies: Information in an uncertain environment. American Naturalist, 123 (5). 626-641.
9. Clark, C.W. and Mangel, M. The evolutionary advantages of group foraging. Theoretical Population Biology, 30 (1). 45-75.
10. Clearwater, S.H., Hogg, T. and Huberman, B.A. Cooperative problem solving. in Huberman, B.A. ed. Computation: The micro and macro view, World Scientific, Singapore, 1992, 33-70.
11. Clearwater, S.H., Huberman, B.A. and Hogg, T. Cooperative solution of constraint satisfaction problems. Science, 254. 1181-1181.
12. Cummings, J.N. Work groups, structural diversity, and knowledge sharing in a global organization. Management Science, 50 (3). 352-364.
13. Ellis, D. A behavioral approach to information retrieval system design. Journal of Documentation, 45. 171-212.
14. Giraldeau, L.-A. and Caraco, T. Social foraging theory. Princeton University Press, Princeton, NJ, 2000.
15. Glance, N.S. and Huberman, B.A. Dynamics of social dilemmas Scientific American, 1994, 58-63.
16. Hassell, M. and Varley, G. New inductive population model for insect parasites and its bearing on biological control. Nature, 223. 1133-1136.
17. Huberman, B.A. The performance of cooperative processes. Physica D, 42. 38-47.
18. Huberman, B.A. and Adamic, L.A. Information dynamics in a networked world. in Ben-Naim, E., Frauenfelder, H. and Toroczkai, Z. eds. Complex networks, Springer-Verlag, Berlin, 2004.
19. Huberman, B.A. and Hogg, T. Communities of practice, performance and evolution. Computational and Mathematical Organizational Theory, 1. 73-92.
20. Kittur, A., Suh, B., Pendleton, B.A. and Chi, E.H. He says, she says: conflict and coordination in Wikipedia Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, San Jose, California, USA, 2007.
21. May, R.M. Uses and abuses of mathematics in biology. Science, 303. 790-793.
22. Newell, A. The knowledge level. Artificial Intelligence, 18. 87-127.
23. Pirolli, P. Information foraging: A theory of adaptive interaction with information. Oxford University Press, New York, 2007.
24. Pirolli, P. and Card, S.K. Information foraging. Psychological Review, 106. 643-675.
25. Putnam, R. Bowling alone: The collapse and revival of american community. Simon and Schuster, New York, 2000.
26. Sandstrom, P.E. Scholarly communication as a socioecological system. Scientometrics, 51 (3). 573-605.
27. Seth, A.K. Modeling group foraging: Individual suboptimaliy, interference, and a kind of matching. Adaptive Behavior, 9 (2). 67-90.
28. Stephens, D.W. and Krebs, J.R. Foraging theory. Princeton University Press, Princeton, NJ, 1986.
29. Swanson, D.R. Fish oil, Raynoud's Syndrome, and undiscovered public knowledge. Perspectives in Biology and Medicine, 20 (1). 7-19.
30. Swanson, D.R. Undiscovered public knowledge. The Library Quarterly, 56 (2). 103-118.
31. Wilson, P. Unused relevant information in research and development. Journal of the American Society for Information Science, 45 (2). 192-203.