A Multilevel Science of Social Information Foraging and Sensemaking

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

ABSTRACT

The time is ripe for a new paradigm of research on social information foraging and sensemaking. Such science would provide the theoretical foundation for advances in information systems. An efflorescence of new social and behavioral phenomena has emerged on the Web and the Internet creating new opportunities for both basic and applied science. Human-computer interaction and information retrieval has largely focused on solitary individuals working with applications and content. There is an opportunity now to develop predictive, quantitative, formal models of social information foraging and sensemaking.

INTRODUCTION

The cooperative production, sharing, and use of information has provided the machinery for scientific progress for several centuries. Cooperative information sharing and information analysis is often proposed as a means to avoid catastrophic failures such as 9/11 or the NASA Columbia explosion. In other words, social information foraging and sensemaking is generally thought to improve situation awareness, problem solving, decision making, innovation, and discovery. Now, we are witnessing the effects of the wide distribution of technology (e.g., Web 2.0, mobile phones, ubiquitous computation) to support cooperative information networks among a broader swath of organizations and the general public. Scientific understanding of social information foraging and sensemaking in such networks can provide the foundation for predicting socio-technical systems that increase discovery, innovation, situation awareness, and human intelligence at both the aggregate and individual level.

There appears to be a number of intuitions about why cooperation might improve information foraging and sensemaking. (1) Like over-the-horizon radar, an individual information forager may receive information otherwise unseen because of the information flowing to him or her from a social network of collaborators. (2) Collectively, by arranging the spotlights of attention of individual sensemakers to insure maximum coverage of information, one can diminish the chances of failing to bring to light some crucial data that might otherwise be missed. (3) Coordinated assemblies of "content experts" may be exploited for their specialized skills and knowledge. (4) A diversity of viewpoints can be brought to bear to provide a broader collective wisdom and to provide mutually corrective forces to overcome the cognitive heuristics and biases that often create blindness to unconsidered possibilities.

However, there is also a considerable body of evidence from social psychology and elsewhere that indicates that collaborative information processing often results in worse outcomes than working alone. Further, it is common knowledge that cooperation involves overhead costs that, for the individual, act as disincentives for collaborative activities. Despite the potentially negative features of collaboration**,** many tools for social information foraging have emerged as disrupting Web 2.0 technologies.

The rich socio-technical network of information producers and consumers has resulted in a rich ecology in which new phenomena, and new problems, have arisen. There is a need to develop theories, methods, and measurement techniques that advance our understanding and foster continued technical innovation.

INITIAL STEPS TOWARDS A THEORY

As an example, I summarize an attempt to integrate a theory in this area. It is not the only theory or even the only kind of theory that could be developed, but it is the one that is most familiar to me.

Framework

Following the framework developed by Allen Newell (see [11]), one can view human behavior as the emergent consequence of a hierarchy of systems (see Table 1). The basic time-scale of operation of each level increases by approximately a factor of 10 as one moves up the hierarchy. Social and behavioral phenomena can be organized into different bands at different time scales, with each band being dominated by different factors. The *biological band* is dominated by biological factors (e.g., neural firing rates), the *psychological band* by human information processing mechanisms, the rational band by goals, feasible actions, constraints, and adaptive tendencies, and the *social band* by social and aggregative mechanisms. Although dominated by different kinds of factors, there remains the assumption that each level is realizable (explainable) by levels below. This has a number of consequences, but of particular interest are the dual notions that (1) phenomena at one level can be analyzed often with only minimal assumptions about the lower level but (2) sometimes effects at a lower level percolate upwards to have dramatic impact at higher levels.

An example of near-independence of the levels is the success of *rational analyses* (at the rational band of Table 1) in information foraging theory [11] which were able to make many predictions about human-information interaction with (for instance) the Web, using a kind of optimization analysis of Web tasks and systems. On the other hand, models of human cognition were developed to show how human psychology (at the psychological band in Table 1) approximates the rational models. However, it can also be shown that small perturbations at the psychological band have large effects at the rational band. For instance, small changes in the perception of Web links in the psychological model of Web navigation can be shown to cause a radical transition in navigation cost predictions in the rational model.

Table 1. Time scales of human behavioral phenomena.

Scale	Time Unit	Band
10^7 s	Months	Social
10^6 s	Weeks	
10^5 s	Days	
10^4 s	Hours	Rational
10^3 s	10 min	
10^2 s	Minutes	
10^1 s	10 seconds	Psychological
10^0 s	Seconds	
10^{-1} s	100 msec	
10^{-2} s	10 msec	Biological

Can the social phenomena we are witnessing with things like Web 2,0 be successfully explained at the multiple levels in Table 1? Can we take emergent phenomena, such as the long tail distributions that characterize content production and usage frequencies over users and explain them as a consequence of networked interactions among rational agents? Can we predict how a small change to a Wiki interface or a recommendation interface will cause changes to the psychology of human-information interaction that "percolate upwards" to produce greater participation rates or better quality results? Or, which will produce polarized communities of biased individuals? These are the kinds of issues that have now become possible to address.

Sketch of a Theory

Building on previous studies of individual information foragers and sensemakers [12-15] my colleagues and I have started to work towards a theory of social information foraging and sensemaking [11]. The theory draws upon models from optimal foraging theory [5], computational ecologies [6, 9, 10], library science [17], and anthropological studies of scholars [16].

Figure 1 summarizes key predictions of the theory. Figure 1a shows how collaboration with a diverse group improves the rate of return to the individual information forager. Being embedded in a cooperative social network of sensemakers provides the individual with the ability to explore more of the space of information more rapidly than could be done alone—like an over-the-horizon radar [4]. For instance, scientists typically receive a substantial amount of relevant information about their core fields from peers (e.g., in the form of preprints, personal recommendations, personal communications, etc., see Sandstrom [16]).

Figure 1b shows how the theory predicts how cooperation among a set of sensemakers improves the probability of making important (or difficult) discoveries. This means that unseen patterns, connections, inferences that are latent in the raw data may come to light as one increases diverse but cooperative sensemakers [17]. The particular lognormal distribution in Figure 1b is often observed in communities of practice [9, 10].

Figure 1c illustrates that the effective size of cooperating groups is a combination of benefits and interference costs to the individual that determine the effective size of a group. People typically join a group only if the benefits (to the individual) outweigh the costs of cooperation [5]. Figure 1c also illustrates how a reduction in the costs of cooperation is predicted to increase the effective size of the cooperating group. Lowering the costs of participation leads naturally to larger groups of cooperating sensemakers. We have all had the experience of participating in work-oriented groups (e.g., this workshop) and often increases in the size of the group increases the cost of participation and has a point of diminishing returns to the individual.

Figure 1d illustrates the effects of diverse heuristics and biases in a cooperative group on the mitigation of confirmation bias [8, 18]. Diversified groups show greater mitigation of confirmation biases than homogenous groups [2]. What is known in the popular literature as "group think" can be mitigated by structuring the makeup and roles of the individuals participating in the collective.

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 is evident, for instance, in the study of social information foraging among scholars [16]. 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 [3], 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,

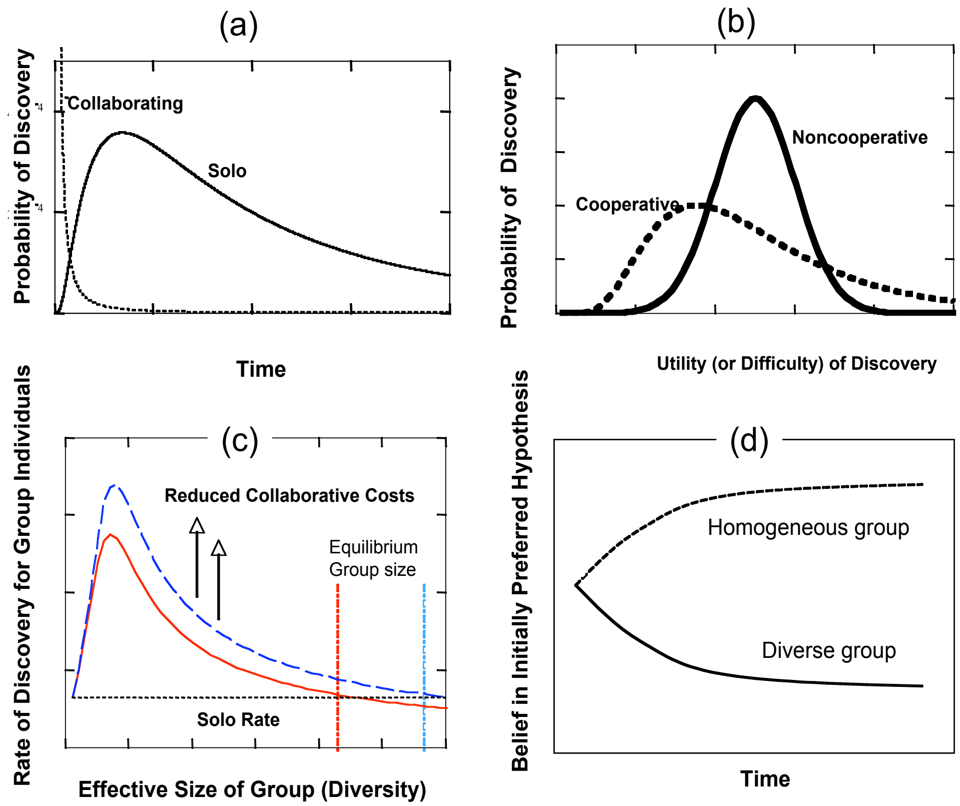

Figure 1. Key predictions of a Basic Social Information Foraging theory [11].

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 [19].

SUMMARY

There has been a recent explosion of technologies that exploit or enhance social information foraging and sensemaking. These include the Web 2.0, blogs, email, internet groups, collaborative tagging, recommender and collaborative filtering systems [e.g., 7], and social data mining systems [e.g., 1]. In modern society, people interact with information technology that more or less helps them find and use the right knowledge at the right time. Increasing the rate at which people can find, make sense of, and use valuable information improves the human capacity to behave intelligently. Given the increased ease with which it is possible to study social networks and information flow in the electronic world, we have great opportunity to develop a science capable predicting the effects of technologies on social information foraging and sensemaking.

REFERENCES

1. Amento, B., L.Terveen, Hill, W., Hix, D. and Schulman, R. Experiments in social data mining: The TopicShop system. *ACM Transactions on Computer.-Human Interaction*, *10* 1(2003). 54-85.
2. Billman, D., Convertino, G., Shrager, J., Massar, J.P. and Pirolli, P. Collaborative intelligence analysis with CACHE and its effects on information gathering and cognitive bias *Human Computer Interaction Consortium*, Boulder, CO, 2006.
3. Burt, R.S. Structural holes and good ideas. *American Journal of Sociology*, *110* 2(2004). 349-399.
4. Burt, R.S. Structural holes and good ideas. *American Journal of Sociology*, *110*(in press).
5. Clark, C.W. and Mangel, M. The evolutionary advantages of group foraging. *Theoretical Population Biology*, *30* 1(1986). 45-75.
6. 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.
7. Herlocker, J.L., Konstan, J.A., Terveen, L.G. and Riedl, J. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, *22* 1(2004). 5-53.
8. Heuer, R.J. *Psychology of Intelligence Analysis*. Center for the Study of Intelligence, Washington, D.C., 1999.
9. Huberman, B.A. The performance of cooperative processes. *Physica D*, *42*(1990). 38-47.

10. Huberman, B.A. and Hogg, T. Communities of practice, performance and evolution. *Computational and Mathematical Organizational Theory*, *1*(1995). 73-92.

11. Pirolli, P. *Information foraging: A theory of adaptive interaction with information*. Oxford University Press, New York, 2007.

12. Pirolli, P. Rational analyses of information foraging on the Web. *Cognitive Science*, *29* 3(2005). 343-373.

13. Pirolli, P. and Fu, W. SNIF-ACT: A model of information foraging on the World Wide Web. in Brusilovsky, P., Corbett, A. and de Rosis, F. eds. *User Modeling 2003, 9th International Conference, UM 2003*, Springer-Verlag, Johnstown, PA, 2003, 45-54.

14. Pirolli, P., Fu, W., Chi, E. and Farahat, A., Information scent and Web navigation: Theory, models, and automated usability evaluation. in *Human-Computer Interaction International Conference*, (Las Vegas, NV, 2005), Lawrence Erlbaum Associates.

15. Pirolli, P., Lee, T. and Card, S.K. Leverage points for analyst technology identified through cognitive task analysis, PARC, Palo Alto, CA, 2004.

16. Sandstrom, P.E. Scholarly communication as a socioecological system. *Scientometrics*, *51* 3(2001). 573-605.

17. Swanson, D.R. Undiscovered public knowledge. *The Library Quarterly*, *56* 2(1986). 103-118.

18. Tversky, A. and Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science*, *185*(1974). 1124-1131.

19. Wu, F., Huberman, B.A., Adamic, L.A. and Tyler, J.R. Information flow in social groups. *Physica A: Statistical and Theoretical Physics*, *337* 1-2(2004). 327.