Order, Coordination and Trust: Understanding Large-Scale Internet Public Goods

Coye Cheshire and Judd Antin

School of Information

University of California at Berkeley

Part I, Draft-in-Progress Copy

DO NOT CITE

Introduction

Since the early 1990's, many different applications and services have emerged on the Internet that allow individuals to communicate with text, images, video, and various combinations of multimedia information. In 2002 alone, the Internet contained over 500,000 terabytes of information (Lyman et al. 2003). In comparison, the entire print collection of the United States Library of Congress contains 'only' 10 terabytes of information. The fact that a massive and increasing amount of content is exchanged across the Internet is intriguing, but it also raises questions about the nature, composition, and purpose of all of that information. Why do individuals freely contribute information across the Internet? How do social incentives encourage contributions when direct monetary incentives are impossible or impractical? How do we develop 'trust' in the information systems with which we interact? These are all important questions that necessitate investigation, especially as individuals increasingly rely on the Internet for transmitting and receiving information on a daily basis.

Recent technological developments have facilitated the creation of new online exchange systems by making the hardware and software that supports those systems cheap and widely available. Millions of individuals who once simply passively *consumed* information provided by privileged gatekeepers now have the capacity to *produce* complex multimedia content and share it with a nearly unlimited audience at little or no cost. These new affordances have enabled both a wider variety of exchange systems that serve specific needs and a smaller number of extremely large systems.

Despite the increasing importance of large-scale participatory online exchange systems, there have been few efforts to classify and differentiate them. To date, most

studies discuss such systems only abstractly or in many cases lump various technologies into a single general category. Our attempt to classify online information exchange systems is not simply a convenience, however. Rather, we suggest that these systems' structural differences can profoundly influence conceptions of uncertainty, risk, and trust, as well as how these factors interact to create social incentives for exchange.

We begin this chapter by describing and categorizing information exchange systems on the Internet along two key dimensions: order and coordination. We argue that these dimensions are useful for understanding the levels of uncertainty that individuals face in different systems and, furthermore, that order and coordination can help us understand how individuals assess the trustworthiness of individuals within a system, and perhaps the systems themselves (as reliable agents for collecting or managing information). Finally, we suggest that order and coordination have an influence on how social incentives operate in the context of online information systems.

In the second part of the chapter, we focus on a specific Internet service, *Mycroft*, which falls within one quadrant in our classification system. As a highly ordered and coordinated system of information exchange, the *Mycroft* system attempts to build knowledge through large numbers of small information contributions. By examining data from *Mycroft's* emergent growth among new users in 2005-2006, we are able to investigate questions of trust and social incentives in a real-world setting.

Information Pools:

Public Goods and Productive Exchange on the Internet

The Nature of Information Goods

When information from many sources is collectively transmitted over a computer network so that it can be accessed by other individuals, it creates a type of public good. We call such systems *information pools*, or collective goods that are created from individual contributions of digital information in online environments. These digital information goods may include (but are not limited to) software, photographs, art, music, speeches/lectures, television shows, books, magazines, or movies.

Two particular characteristics of information help to reduce the costs of contributions to information pools, and differentiate the exchange of information on the Internet from the exchange of many types of physical goods. First, information can often be consumed by many individuals without affecting the consumption of anyone else. In economic terms, this is called *high jointness of supply*. Second, information can be transferred to an individual without the original owner losing her copy of the same information (i.e., the good can sometimes be perfectly replicated).[1] These two qualities are important for understanding the creation of information pools because they make it possible for an individual to keep the full value of the public good without impeding others' ability to consume it (Cheshire forthcoming).

[1] It is important to note that information *can* have the qualities of high jointness of supply and replication, but these are not necessary features for *all* information goods. Indeed, many types of information may have very low jointness of supply because each person's use of it reduces its availability to others. For example, a book can be checked-out from a library by only one person at a time. Yet, the same material in the book could be made available as an electronic file. In both cases (the physical book and the electronic book) the information could theoretically be replicated through photocopies or downloads, yet the jointness of supply may be much higher for the electronic download (since many can download the information at the same time) than for the physical manifestation of the book.

Digital information goods are also somewhat unique because the costs of sharing them are typically quite small (Kollock 1999). There are certainly some very small costs such as the time required to transmit information or the cost of a connection to the Internet. However, these costs are arguably much smaller than the *actual content value* of the information itself. Thus, the fact that the cost incurred by the contributor is very small compared to the content value of the good is another fundamental characteristic of information pools.

Public Goods, Group-Generalized Exchange and Order

When individuals contribute information to a collective good and receive benefits from the collective good, it creates a *group-generalized exchange* system. Following Ekeh's (1974) description of generalized exchange, Yamagishi and Cook (1993) define a group-generalized exchange system as one in which individuals contribute to a public good, yet only receive benefits from the contributions of *others*.[2] Such a situation creates an inherent social dilemma. (Yamagishi and Cook 1993) The public good can only be created if many individuals contribute to it, yet everyone is better off by not contributing (i.e., free-riding) and instead hoping that others will produce the public good (Yamagishi 1992). A group-generalized exchange system can therefore be viewed as a special case of the public goods problem.

In many public goods situations, all contributions are evenly redistributed back to *all* contributors from the public good. For example, individuals who work to reduce air pollution will enjoy cleaner air both as a result of their own contributions and those of

[2] Ekeh (1974) makes several other distinctions between types of group-generalized exchange systems. For example, some systems involve individuals receiving benefits from others (A← B, C, D; B ← A, C, D) and other systems involve individuals who take turns giving to all others (A→ B, C, D; B→ A, C, D)

others. On the other hand, consider the example of individuals who contribute recipes to a cookbook. In this case, everyone else benefits from the collection of recipes but an individual's own contribution(s) do not carry the same value because the contributor already owns the information. This distinction is important because it highlights the fact that there may be small costs associated with contributing digital information goods but rewards are *only* received through the contributions of others – one does not benefit from one's own contribution.

If individuals do not fully understand how or why the collective good may benefit them, there are few motivations to contribute to it. We do not help reduce air pollution without understanding, at least in an abstract sense, that we can breathe easier with cleaner air. Though we may not understand the causes or chemical elements of polluted air, we may understand its benefit instead based on the perceptions and beliefs of others who surround us. Furthermore, if individual contributions are not well organized, individuals may feel that their contributions are being wasted (or at least poorly managed). A cookbook with no table of contents and no index introduces significant barriers to utility, especially if it is large. We refer to the organization of contributions through specific means and towards a focused outcome as the problem of *order*. An *ordered system* is one in which the process of production and exchange is well defined, and the products are clearly specified. In an ordered system a scaffold is in place which determines when, how, and in what form exchanges and contributions are made to the collective good. Put simply, individuals know what they should do to contribute to a public good, and they know what their contribution will ultimately help to form.

For example, Wikipedia (www.wikipedia.org) is an ordered system because the product is clearly defined as, “a free encyclopedia that anyone can edit” (Wikipedia 2006). The process by which the online encyclopedia is produced is through ongoing textual revision (including simultaneous commentary in a different but connected forum). Thus, the collective good in this case is defined by the organization, and contributors can clearly see their impact on the public good as real-time additions to the encyclopedia. While some latitude is possible within this structure, deviating from it significantly would create a markedly different product – it would no longer be an encyclopedia.

On the other hand, consider the example of “blogging” sites. Blogs are journal-style websites on which individuals post text, multimedia, and links on any range of topics. A collection of blogging sites is a good example of an unordered system. Though we can loosely define blogs according to the top-level genre of production (journal-style entries in reverse chronological order) or by the software systems that support them, individual blogs are often idiosyncratic. The process through which blog content is produced varies widely. Indeed, the product that constitutes a “blog” is difficult to define because it takes so many potential forms. Even among collections of bloggers who contribute to a common theme, the definition of the public good in a given context may frequently change – for one group it may be the availability of alternative news and information, for another the ability to share personal stories and photos with friends and family. As we will describe below, this lack of order does not diminish blogging’s importance as a type of public good. However, it does influence issues such as the incentives for contributing content and the assessment of trustworthiness in blogs.

Productive Exchange and Coordination

Information pools can also be examples of *productive exchange*, or situations in which, "mutual interdependencies … are strong enough to make joint action or collaboration the most profitable option to individual actors" (Lawler, Thye and Yoon 2000). Individual contributions to an information pool are often interdependent because they combine to create a collective good that is greater than the sum of its parts. For example, individual contributions to Wikipedia are certainly valuable as standalone bits of information, but their true value comes from their collection into a single, well-defined public good (i.e., an encyclopedia) which no individual contributor could ever create alone.

An important component of productive exchange systems is the *coordination* of individual contributions. Specifically, "in productive exchange, there are strong incentives for exchange, and the main question is whether the actors can coordinate their actions to forge agreements" (Lawler, Thye and Yoon 2000). For example, if one individual has rice and the other fish, there will be no sushi unless each is aware of the other and there is a structure in place to allow them to exchange (see: Yamagishi 1992). Kollock (1999) argues that when individuals contribute digital information to produce public goods in online environments, the coordination and communication costs are often much lower than in other types of interaction situations (such as face-to-face interactions). The coordination costs may indeed be lower in online environments compared to offline systems, but the key point is that coordination costs are a crucial consideration in any system that produces a collective good—including online information pools.

We define a *coordinated system* as one in which the roles of the various actors within the system are clearly specified. Participants in the system undertake specific work-types, responsibilities, and interactions according to rules which are determined by the role(s) that they have chosen or to which they have been assigned. Many open source software communities, for example, use sophisticated software to distribute work to groups of software coders, testers, and debuggers. Within each of these categories there may still more specific roles. For instance, some contributors are assigned to debug specific parts of the program or add specific functionality (Moon and Sproull 2000). In doing so, participants help to further define existing roles and often develop specific competencies that maximize their efficiency.

Classifying Information Pools: Order and Coordination

In order to clarify the similarities and differences between various online information pools, we propose classifying them along the two key characteristics that we have identified: *order* and *coordination.* Of course, there are any number of other dimensions along which we might choose to classify exchange systems – work type, topic area, size, or geographic area, just to name a few. We concentrate on order and coordination because, 1) all information pools can be classified according to these two criteria, and 2) order and coordination are central to understanding how uncertainty, trust, and social incentives operate within different information exchange systems.

We must also offer a few caveats for this classification system. First, the structures of order and coordination must be known among the communities of actors. An exchange system with highly developed mechanisms for order and coordination about

which no one is aware is, in effect, an unordered and uncoordinated system. We should also point out that these are relative measures rather than absolutes. No system is truly unordered – the computer-mediated context itself puts some basic structure in place. Similarly we can expect that in any given system only a few actors will truly have no preconceptions about roles or how to act them out. The majority of actors may immediately understand some or all of the possible generalized roles within a system. Finally, it is essential to acknowledge that order and coordination each exist along a continuum, not as binary classifications.