

Social Transparency in Networked Information Exchange: A Framework and Research Question

H. Colleen Stuart¹, Laura Dabbish^{1,2}, Sara Kiesler¹, Peter Kinnaird¹, Ruogu Kang¹

Human Computer Interaction Institute¹

Heinz College²

Carnegie Mellon University

hcstuart@andrew.cmu.edu, dabbish@cmu.edu,

kiesler@cs.cmu.edu, kinnaird@cs.cmu.edu, ruogu1027@gmail.com

ABSTRACT

An emerging Internet trend is greater social transparency, such as the use of real names in social networking sites, feeds of friends' activities, traces of others' re-use of content, and visualizations of team interactions. Researchers lack a systematic way to conceptualize and evaluate social transparency. The purpose of this paper is to develop a framework for thinking about social transparency. This framework builds upon multiple streams of research, including prior work in CSCW on social translucence, awareness, and visual analytics, to describe three dimensions of online behavior that can be made transparent. Based on the framework, we consider the social inferences transparency supports and introduce a set of research questions about social transparency's implications for computer-supported collaborative work and information exchange.

ACM Classification Keywords

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

General Terms

Human Factors

Author Keywords

Social transparency; information exchange; social translucence; awareness; visualizations; collaboration; innovation; theory

INTRODUCTION

More information than ever before is being revealed about content, people, and their interactions online. Internet applications make visible individual's identities and actions as they access and share information. Facebook tracks user activities, and allows search engines and advertisers access to that information [73]. Twitter explicitly supports attribution of retweets and embedding of Tweets within a webpage [54]. It is technically possible to make almost any action on a piece of information visible to users within or

even across websites. Further, whereas early digital networks offered a comparatively impoverished social context for communication, today's information exchange may include text and audio conversations, faces, names, videos, and visualizations of social networks. Social transparency, which we define as the availability of social meta-data surrounding information exchange, is increasing at a rapid rate, and qualitatively changes the way we conduct work and interact with others online.

Only some aspects of social transparency have been conceptualized and studied in CSCW. Erickson and Kellogg [25] introduced the concept of social translucence and argued that making co-workers more visible, and letting them know when someone on the team acted on a joint project, would encourage participation and promote collaborative work. There also is a considerable history of research on awareness systems specific to a particular function, such as tracking the actions of a group on a shared artifact or the location of co-workers [1, 23, 32]. Advances in visual analytics have made online community activity more accessible as well. Suh et al. [74] and Viégas et al. [77] explored the benefits for Wikipedia projects of making edits and editors more understandable through visualizations.

Although this previous work illuminates the potential promise of increased levels of social transparency about co-worker actions for computer-supported collaborative work, we still lack a systematic way to conceptualize and evaluate social transparency. Technical advances are pushing the frontiers of social transparency beyond prior research frameworks and are starting to exceed some people's comfort levels. For example, the re-use of medical prescriptions in marketing led to a court challenge (e.g., U.S. Supreme Court, Sorrell v. IMS Health). Vigorous debates are being held on privacy, anonymity, and the use of personal identity data for communities and commerce (e.g., [47]).

The goals of this paper are fourfold: (1) we provide an overarching framework and consistent terminology for the concept of social transparency, showing how it extends from the basic model of information exchange. We build on previous work on social translucence and awareness, but capture a broader range of transparency in information exchange, including concepts already in use such as

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CSCW'12, February 11–15, 2012, Seattle, Washington, USA.
Copyright 2012 ACM 978-1-4503-1086-4/12/02...\$10.00.

authentication, anonymity, and provenance. (2) We then consider the inferences made by users that social transparency supports and their second order implications. Inferences stem from the immediate effects of changes in technology, such as making it easier to share information. Second order effects are indirect, longer term, and sometimes hard to anticipate effects. We argue that social transparency designs are making information visible without careful thought to the social inferences these designs support and the second order effects¹ [72] of those inferences. We focus on effects critical for collaborative work, such as the quality of information shared, coordination and control. (3) We point to some of the tradeoffs that are, or will, arise as the result of increasing social transparency. (4) Finally, we identify future areas of research related to social transparency to encourage CSCW researchers to study how these changes will affect collaborative work and information exchange online.

INFORMATION EXCHANGE

A basic sender-receiver communication model, derived from information theory [67], can describe information exchange between two individuals. A source (or sender), “Bob,” disseminates information to a receiver, “Alice,” and obtains confirmation from Alice in the form of verbal and non-verbal feedback that she received and understood the message (Figure 1). Information, defined as “knowledge [that] can be transmitted without loss of integrity ” ([41] p. 386), travels within a tunnel, whereby only Bob and Alice have access to the information, and know that the exchange occurred. Considerable research in communication, social psychology, and sociology has enriched this view of two-party exchanges, showing they are typically iterative and cooperative (e.g., [15]).

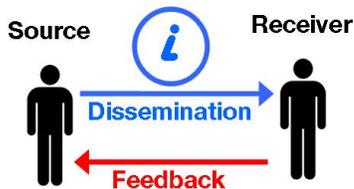


Figure 1. Dyadic information exchange

Today, dyadic information exchange is becoming less typical as people share their information on the Internet, purposely or not. Bob sends an email to Alice with copies to other people, posts comments in response to a blog, or posts a video of his family. Observers such as Lee, whom

¹ Studies of email provide many historical examples of first and second order effects [72]. For instance, email made it easy for organizations to create virtual teams, a first-order effect. However, the huge rise in popularity of virtual teams led to second-order effects, such as team coordination challenges [21] and employee stress [2].

Bob may not even know, might get forwarded Bob’s email or see his comments and video (Figure 2).²

The model in Figure 2 is as old as the Internet, but what is comparatively new is the increasing amount and quality of information that sources and receivers have about their information exchanges and the opportunity for observers like Lee to engage in these exchanges. Services within and across applications make visible three social dimensions of information exchange: the identities of those exchanging information—*identity transparency*, changes to content—*content transparency*, and actions taken during the interaction—*interaction transparency*.

IDENTITY TRANSPARENCY

Identity transparency refers to the visibility of the identity of the sender and/or receiver in an information exchange. For a source, identity transparency means information about the receiver is visible, whereas for the receiver, it means that information about the source is visible. Identity transparency also can be asymmetric; meaning only one party in the exchange has information about the identity of the other.

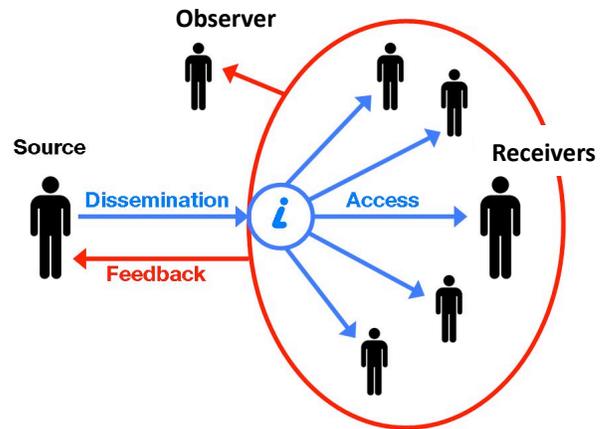


Figure 2. Networked information exchange: Multiple receivers access the same information and third parties can observe these actions on information.

Variations in Identity Transparency

Identity transparency overlaps with the security concept of message authenticity, which refers to the verifiability of the identity of a sender and/or receiver of a message. Where verifiability is typically all or none, however, identity transparency designs have many gradations and ranges from strong identity transparency, where much information about a person’s real world identity is persistently available across applications and settings, to complete anonymity,

² Observers could be individuals, as in our example here, or organizations. See Wicker and Schrader [80] for an in-depth discussion on the implications of surveillance for commercial purposes.

where there is no persistent identifier of any kind. These variations in identity transparency influence social inferences about similarity, reputation, relative status, and perceived credibility of an information source.

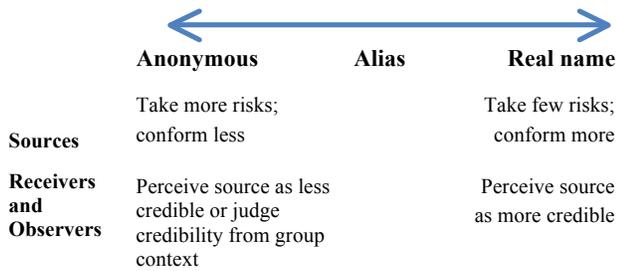


Figure 3. First order effects in the continuum of identity

Designs with strong identity transparency include a real name and/or information about source or receiver attributes, such as the person’s demographic characteristics and membership in groups (“profiles”), acts as a signal or cue to a source’s or receiver’s trust in others and willingness to be accountable for what he says and does [53, 62]. Identity information about the source or receiver can support social inferences about how similar or dissimilar a user is to others involved in an exchange. The concept of homophily suggests that similarity perceptions can influence the likelihood of information exchange [52] in that we are attracted to and prefer to interact with those who are similar to us. In situations where identity is transparent sources should be more likely to initiate new information exchanges with receivers perceived to be similar (or have similar views) to themselves [45]; receivers should be more likely to accept information from sources who are similar than from sources who are dissimilar [3, 43]. Source attributes such as the designation “superuser” in TurboTax’s Live Community and “barnstars” in Wikipedia, can also act as cues to social status, and increase the perceived credibility or the believability of a source and the information [53, 62].

Medium levels of identity transparency are created through designs that employ virtual identifiers such as unique user IDs and aliases allowing sources and receivers to be uniquely identified within an application and, if they choose, use the same identity across applications while avoiding using their real names or revealing demographic information about themselves. Virtual identifiers give sources control over their self-presentation in different communities [24], and receivers over contribution assessments within those communities. With repeated exposure and positive interactions, persistent aliases become more like real names in that they support trust, relationship development, and reputation building within sites and communities [61]. Persistent identifiers can also lead to reputational losses if people’s interactions are negative (e.g., in the website Slashdot as described by Lampe and Resnick [44]).

Anonymity, the lowest level of identity transparency, makes it harder for sources and receivers to interpret one

another’s credibility, and can signal untrustworthy or suspicious information. Accordingly, anonymity seems to function best within communities or groups where almost everyone is anonymous. Sources and receivers develop norms of information exchange, and pay attention to the social context to interpret others’ information [5, 17, 65]. Because anonymity reduces the personal risks associated with information sharing such as embarrassment, reputational losses, or social sanctions, anonymity in such communities and groups can facilitate the transmission of controversial, sensitive, critical, or novel information [38, 51].

To summarize, research suggests that different levels of identity transparency support different types of social inferences about sources and receivers. This research also suggests that sources and receivers change their behavior when we move along the continuum of weak versus strong identity transparency (Figure 3). When Bob uses his real name (or a persistent identity across many applications), he and his information are potentially more public, and he will likely take more care about the information he discloses or disseminates than when his identity is not known. Alice’s trust in information from Bob will increase. Strong identity transparency therefore is likely to increase sources’ sense of accountability and concern for how others perceive, value, and use their content online [38]. By contrast, when Bob is anonymous or uses a temporary alias, he will feel freer to say what he wishes, Alice is likely to take what he says less seriously, and she will need to look to the group context and norms to evaluate Bob’s information.

Second Order Effects of Identity Transparency

We argue that increasing identity transparency of sources and receivers will have second order effects in collaborative groups and communities, as a result of the inferences it supports. In particular, increased perceived accountability will result in greater accuracy of information shared and used, while lowering creativity by reducing the amount of unique or novel information shared.

Information Accuracy

One plausible effect of identity transparency is an increase in the accuracy of information. Identity transparency should result in higher reputational accountability of sources and allow receivers to better evaluate the accuracy and usefulness of information from sources. In turn, the ability to evaluate information and its source should increase trust in information and information sharing. Research has shown that employees often fail to share or use information from their own organization that are held in repositories or databases because they are anonymous in the system. As sources, they do not know how others will use their information, and as receivers, they cannot evaluate the usefulness of the information or contact the source for more details [6, 7]. When sources and receivers are identified, potentially useful information is more usable. In one test of this idea, contributions to MovieLens, an online movie recommender system, were both greater in number and higher quality when contributors knew that their reviews

would be checked by another MovieLens member [18]. This study began what could be a larger investigation into how online accountability, and designs for reinforcing it, affects the amount, accuracy, and usefulness of information that people create or disseminate to others within and across online communities. This research also could show how accountability changes, or fails to change, the psychology and detection of lying and false reporting online [59,79]. An open question concerns asymmetric identity transparency, when a source or receiver is identified but the other party is not. Figure 4 shows some examples of symmetric and asymmetric identity transparency, depending on whether the identity of the receiver and source are visible. The example of the downloaded paper will be familiar: Bob uploads his paper to his website. He will not know the identity of Alice, who downloads the paper and reads it. Applications such as Google Scholar offer post-publication citation services [10] and applications such as Turnitin.com offer plagiarism checking [22], but it is not as easy for Bob to know about Alice. Giving authors the ability to trace the identity of readers might be controversial, but would allow authors to have an early lead on potential collaborators or students, insight into the fields where they are having impact, and potential linkages to other relevant work.

Creativity

Within some communities and collaborations, an increase in accountability would lead to more conformity, a decline in risk taking, and less creativity. Because their identities are known, people in strongly transparent groups and communities are likely to work harder at being civil and doing what their group wants, but also increase their conformity to group ideas and norms. This consequence

Receiver's Perspective

		Source's Perspective	
		Receiver Not Visible	Receiver Visible
Source Not Visible	<i>Symmetric</i> Anonymous public post	<i>Asymmetric</i> Company tracks product views	
Source Visible	<i>Asymmetric</i> Downloaded paper	<i>Symmetric</i> Email between friends	

Figure 4. Symmetry and asymmetry in identity transparency

needs to be much better understood as it could affect the culture of creative collaborations. Suppose Alice, Bob, and Charles see that they have each accessed a particular paper; they may perceive they form a group [12]. As members of a group, they will be more likely to communicate with one another, and if they do, they will feel social pressure to adhere to each other's opinions. Members of a triad are less independent in their behavior than members of dyads [69,70]. Bob can be more confident that Charles will behave in a reasonably predictable and honorable way, because Bob and Alice can sanction Charles' inappropriate

behavior. This increased sense of security for Bob and Alice, and civility of the group as whole, also comes with a cost to the individual group members—each will have to take more care with what he says.

The foregoing discussion suggests that there may always be defensible reasons within some communities to remove identity transparency, and to support anonymous information exchange for individuals and groups. The argument raises several research questions. First, what are the various situations in which communities and collaborations want anonymity? Researchers have investigated why some users choose not to actively contribute to online communities, which in turn hides their identity (e.g., lurkers [55, 56]). To our knowledge, however, there are no studies of when people prefer anonymity. Second, are there ways to assure anonymity for some parts of collaborative work (e.g., brainstorming) but maintain strong identity transparency for most of it? Google's Chrome browser incorporated a so-called "incognito mode," designed to remove traces of users' activity from their own machines [16]. Is it possible to design a similarly easy-to-use mode for controlling personal identification in groups? Tor and other similar proxy services provide anonymity today, but these solutions rely on third parties and are awkward for spontaneous work in collaborations. Researchers might consider the technical requirements for retrofitting such that anonymity (or lack thereof) is guaranteed for certain tasks.

CONTENT TRANSPARENCY

Content transparency refers to visibility of the origin and history of actions taken on information. For example, Wikipedia keeps a complete history of all the edits that have been made to every page. Viégas, Wattenberg, and Dave created the HistoryFlow visualization of editorial changes to improve the accessibility of these revision logs [77]. The history flow visualization uses edit timestamps in Wikipedia to construct a visual representation of the persistence of different pieces of text within a Wikipedia article. By making editorial changes more visible, the visualization increases individuals' ability to see how a document evolved over time as a result of many editors' actions on the content. Designs that incorporate content transparency can support inferences about others' activities (activity awareness) as well as inferences about the credibility of the content of a piece of information.

Variations in Content Transparency

One way systems can provide content transparency is using designs that make provenance information visible. Provenance refers to the origin or earliest known history of some artifact or content. When interface designs make provenance visible to sources and recipients, they both know where information originated. For instance, when Bob uploads his original essay to the web, then the parts of Bob's essay that Alice uses in her teaching lectures can be traced back to Bob. Provenance allows for checking plagiarism and the authenticity of data. A lack of

provenance means that content has no clear ownership; hence it is easy to borrow, steal, repurpose, or fake.

Provenance can support inferences about others activities, leading to activity awareness, consciousness of other individuals' actions [23], because provenance is often determined by tracing back through prior actions on information. For example, Software developers become aware of others actions on code by looking at their commit logs [71]. Many researchers in CSCW have studied activity awareness but not usually because they are interested in determining the source or ownership of content. Instead, they have used the concept out of a larger concern for supporting coordination of collaborative work, especially distributed work [32, 66].

Revision control systems that require authentication, such as those commonly used in enterprise-scale software engineering, are a good example of content transparency used to support coordination. These systems provide detailed information about all changes to software. Engineers can make specific queries to see the changes. Visualizations of the system can help them create a mental model or narrative of the workflow [57].

Provenance can also support inferences about the characteristics of the individuals who have acted on the information over time. Visualizations of how people have edited or changed information over time not only aid coordination but also serve as an indicator of the expertise of those who created or changed information [40, 74, 77]. These visualizations can also support inferences about social structure within the set of individuals who have acted on the information, for instance, has anyone taken a leading role and are others following? Transparency of the development of content can reveal the roles of the individuals involved in its creation—what Gutwin, Greenberg, and Roseman referred to as structural awareness [33].

Second Order Effects of Content Transparency

Because content transparency lets people see what is happening to a project or a community issue, and when, where, and how sources and receivers have changed information, this meta-information can serve as a reminder for others to contribute or respond to the changes in content. We argue that these kinds of cues will produce broader second order effects.

Productivity

Activity awareness may drive productivity by reminding project members about interdependent projects, and through social facilitation processes. The reminding cues [36] that derive from strong content transparency should increase the productivity of collaborations and the self-corrective effects of crowd sourced information. Content transparency also can increase productivity through social facilitation processes; when sources know that their actions are visible, this awareness should prompt them to work harder. In one study, cashiers in a supermarket were more productive when they were watched by a highly productive peer [49].

Distributed teams are more likely than collocated teams to experience inattention to tasks, group conflict, neglect, and delays in their work [11, 19, 35]. We attribute some of the cause of these problems to weak content transparency, whereby members lack feedback and immediate knowledge of each others' activities, and fail to attend each others' progress [21]. Increased content transparency in the form of activity awareness can provide feedback as to whether the collaboration is proceeding as expected and keep collaborators on track, enhancing mutual knowledge and improving the accuracy of shared mental models [1, 23]. Documented work progress also provides collaborators with a common reference point for discussion, a shared understanding of the development of the project, what else needs to be done, and who might do it. For example, one IM prototype application shows activity in projects, with an icon indicating when any member of the project is changing a document [1, 23]. These kinds of cues may speed up content development by reducing lags between iterations of work, mitigating coordination problems, and reducing social loafing [37].

Stress

As with identity transparency, however, there are contexts in which strong content transparency is likely to have some negative consequences for collaborations and communities. Content transparency makes everyone's behavior more visible and therefore more open to being evaluated and second-guessed by others. Research on "evaluation apprehension" suggests that when people are worried about others' evaluations of their work, they sometimes make mistakes and learn less than when they are not watched (e.g., [75]).

Although there is much excitement about new visualizations supporting content transparency, it is not clear whether these designs can support detailed activity awareness without overloading and subsequently stressing users. Imagine that every project Bob is working on notifies him of Alice, Charles', Sam's (and many others') changes to content in each project. Are there upper boundaries to content transparency? Many visualizations of content changes provide people with vast meta-information about information. Examples of such visualizations include the history flow visualization mentioned above, and the code swarm visualization project [58]. Although these examples may excel for certain tasks, researchers have already identified a tension between visualizations of content changes and increased cognitive load from adding more resources to a person's desktop (see [75]).

In many cases, the volume of changes to pieces of shared information may far exceed the attentional capacity of individuals. For example, a study of awareness in software development suggested that engineers found the email messages generated to notify them of individual changes to code disruptive, and struggled to keep up with the information contained within often keeping changes private to avoid disturbing others [71]. Dashboards and feeds only partially alleviate these issues, and selecting the correct

level of detail remains a challenge [76]. Notifications about changes to content can generate constant interruptions to ongoing tasks, known to be disruptive to productivity [50]. Also, transparency in collaborations might increase the stress that many people feel as they feel social pressures to keep up with multiple projects and collaborators [2]. Thus an important research question is at what level of detail does content transparency need to be to mitigate overload while supporting collaborative function? And how frequently should individuals attend to it, to avoid distraction while maintaining awareness?

INTERACTION TRANSPARENCY

When information is transmitted online, interface designs can make the details of an information exchange more or less visible to third parties. We define interaction transparency as the ability of a third party to observe an information exchange between a source and receiver, and (assuming symmetry) the knowledge that source and receiver have about the presence and identity of these third parties. This information supports social inferences about normative behavior as well as the social structure within a community or closed group. These inferences, when aggregated to the group level, may foster herding behavior, foster access to new information and influence creativity.

Variations in Interaction Transparency

Many Web applications track information access behavior and display it to one or more third parties (observers who did not create the content or access it). For example, if Alice publicly shares her photo on Flickr.com, the website indicates to Alice and anyone who visits the site how many times her photo has been viewed by others. Flickr allows users to designate friends as contacts and to track their activities in real time [46].

Interaction transparency supports social inferences about what constitutes acceptable behavior by sources, receivers, and third parties themselves. This happens because meta-information about third parties who opt to view an interaction can act as a source attribute. For example, if Charles and his friends watch Bob exchange content with Alice, the Bob-Alice interaction seems more important to Sally, another third party who sees that Charles and his friends have an interest in Bob's information. Generally, meta-information about people's behavior influences observers' interpretations of content and responses to sources (e.g., [4]). People tend to assume that others' behavior is usual, correct and worth following (see [13] for a review). Thus, guests at a hotel are more likely to reuse their towels when they are informed that most of the other guests at the hotel participate in the same environmental conservation program [30]. They are less likely to litter when they see a lonely piece of trash on the sidewalk than if there is no litter at all; the single piece of litter reminds them that most people are not littering [14]. Third parties who see others' information exchanges also are more open to learning from the information. For example, newcomers to an online community who observe exchanges among more senior members are more open to adopting

community norms [42]. Interaction transparency may be beneficial to newcomers because they become acclimatized to the normative behaviors of the community [60].

Interaction transparency also conveys information about social structure to third parties. While knowledge of a social structure is an important resource (e.g., [8]), individuals are likely to make errors in their assessments of what that social structure looks like [39]. Having accurate information about others social relationships would allow individuals to make more accurate assessments of other's social capital and locate expertise within a network.

Third party meta-information also helps potential receivers filter through large amounts of information to bring the most valued content to the surface. When Kijiji, for example, lists the number of page views for each classified ad on that site (www.kijiji.com), and the New York Times lists its top articles based on recommendations and blog coverage, viewers get a sense of what is important to other users (www.nytimes.com). Suh et al. developed Wiki Dashboard to help users filter through Wikipedia articles [74].

Second Order Effects of Interaction Transparency

Interaction transparency reveals social information who is exchanging information with whom, who is dependent on whom for information, and which persons or groups are the primary conduits for information. We argue this information will have several second order effects.

Popularity and Herding

Interaction transparency increases people's knowledge about the popularity of information, sources, receivers, and third parties. Because popularity signals credibility, popular information and popular people have a huge advantage in gaining more popularity [64]. The result may be increased tendencies for herding behavior within communities and collaborations as interaction transparency increases.

Rating systems and public usage statistics exemplify information given to third parties about exchanges among sources and receivers. Someone coming to a site can see summary information at a glance about who has liked or disliked different content. Google HotPot and Stumbleupon curate content based on people's social connections [28]. Research suggests that this kind of rating information does cause herding behavior. In a study of music exchange markets, songs rated highly early in the rating process skyrocketed in popularity, irrespective of song quality [64]. Herding behavior could be expected to increase with more interaction transparency, and could affect the opinions and behavior of collaborations and communities.

Herding behavior is worrisome because it may reduce the variety of content that people access since they are being fed information that is consumed by others within their social network. Seeing what one's friends or coworkers are looking at could reduce access and use of new information, and increase segregation of information exchange based on ideology and group memberships [27].

New Information

Increased interaction transparency could make larger, more dispersed networks visible. For example, Microsoft Academic Search (academic.research.microsoft.com) is a visualization designed to display co-authorship networks that display, for instance, co-authors of co-authors. Making networks more visible could counteract the tendency for people to only interact within their own groups. It would show that social structure is not as rigid as might be assumed, and useful social capital [9]. Charles and Sue might be trustworthy and predictable collaborators but not the most useful contacts for each other in all situations. If Charles and Sue only share information with each other and their joint contacts, the information that Charles receives from Sue will tend to be redundant with the information he gets from his other existing contacts [31]. In short, we argue that revealing social networks could increase the uniqueness of information people seek and receive. This effect could have interesting implications for collaborative work and innovation.

Previous work has shown how attributes of a source influences the likelihood that new information is disseminated [63]. This previous work has not considered how interaction transparency will influence the spread and reuse of information. Reputation cues and collaboration networks should influence the information that users seek out and reuse. If a researcher learns through a revealed collaboration network about people who do similar work or have new ideas about similar topics, he might be more likely to seek out and reuse that content. Increased visibility would make a search for new information more efficient because users are pointed to individuals with similar interests who were previously unknown [34].

Interaction with Identity Transparency

Earlier, we noted that when receivers know the identity of sources, and this transparency is symmetric, sources may feel more accountable for their information and more careful about what they reveal. As well, knowing that third parties are observing them can make sources more cautious about their information exchange behavior, especially their use or reuse of sensitive information such as searches for health information. The influence of observers is especially influential in new, uncertain, or risky situations (e.g., [48]). Being observed may lead sources to alter the information they share such that it becomes more easily accessible to a broader audience, or they may omit certain content that would offend others or is proprietary or could be considered plagiarism. Third party transparency could affect the way receivers respond to information from sources. For example, receivers might confirm receipt of publically conveyed information more quickly than privately transmitted content, and provide more feedback, because they want to appear conscientious in the eyes of third

parties or the broader community. These speculations have not been studied but are important to understanding the conditions under which interaction transparency increases or decreases innovation.

Privacy has received considerable attention in the CSCW research literature (e.g., [20, 29]). Perceived privacy violations can be caused by asymmetric identity and interaction transparency, as when a third party knows the identity and actions of sources or recipients but the reverse is not true [43]. Third parties could have codified, reused, or disseminated information about the sources or recipients without their knowledge or permission. Researchers have found that feedback regarding who is observing a user's activities (that is, creating symmetry) allays people's privacy concerns [43], where in other situations asymmetry might be desired by users [78]. However, we still lack sufficient information about the effects and tradeoffs involved in asymmetric interaction transparency. For instance, threats to privacy with asymmetric interaction transparency exist in countries with governments that block or censor websites, and even arrest people for their content. Often the rules are not clear and it is impossible to trace the existence or identity of third party observers or to make all parties anonymous. Shklovski and Kotamraju [68] describe how people in Kazakhstan and Russia self-censor their behavior and avoid anonymization services to avoid appearing suspicious. Their study suggests that cross-national and cross-culture research is needed to understand better the implications of asymmetry in interaction transparency.

Another important question regarding identity and interaction transparency is whether and how sources can stop the flow of meta-information about their content. The Internet makes it easy for recipients and third parties to archive content and meta-information without the source's knowledge. Once content is archived by another user, the source loses control over the lifespan of that content. Are there technical means to add tracking devices or automatic degradation of content over time? Just as a newspaper passed around on a rainy day will slowly fall apart, can the creator of content specify certain types of file accesses or modifications that would partially or fully corrupt the content?

Perhaps it is possible to design DRM solutions, accessible to individuals, to enhance innovation through attribution and tracking of content's reuse. Any transfer to a network-enabled machine could cause the data to "phone home." Systems like Vanish attempt to give the source control over the lifespan of data by supporting a data self-destruct date [26], but this solution has not yet been widely adopted. Researchers need to consider more fully the technical requirements and user impact of transparency control.

Table 1. Some Effects of Social Transparency

Dimension of Transparency	Transparency Components	First Order Effects Cited in Prior Research	Hypothesized Second Order Effects
Identity Transparency	Real names	Accountability: Names and virtual identifiers aid in developing a reputation [44,61]	<p>↑ Information Accuracy More identifiably in a community or collaboration increases accountability, accuracy of information</p> <p>↓ Creativity More identifiably in a community or collaboration increases conformity</p>
	Virtual identifiers (e.g., IDs and aliases)	Image maintenance: Virtual identifiers provide control over people’s self presentation [24]	
Content Transparency	Profile information (e.g., demographic, historic data)	Trust: Profile information increases credibility [53, 62] Perceived similarity: Profile information attracts similar others [45]	<p>↑ Productivity Provenance and revision awareness increase productivity in collaborative projects</p> <p>↑ Attentional Demand Provenance and revision awareness increase stress from collaborative work</p>
	Anonymity	Trust: Anonymity applications are used to reduce risk [38] People use cues from the social context to judge sources and receivers [17,65]	
Interaction Transparency	Provenance and greater awareness of authorship	Activity awareness: Revision histories keep people in the loop (who is doing what) [74]	<p>↑ Herding Revealed popularity multiplies the impact of popularity</p> <p>↑ Innovation Revealing weak ties encourages access to new information and ideas</p> <p>↓ Privacy Asymmetries in identity and interaction transparency increase privacy violations</p>
	Revealed revision history	Credibility: increases trustworthiness of content [40,74]	
Interaction Transparency	Evidence of third party access (e.g., link referral history)	Social networks: Reveals new connections between users [34] Norms: Reveals ‘normal’ behavior through popularity of content [64] Self censorship: Due to being in the public eye [48]	<p>↓ Privacy Asymmetries in identity and interaction transparency increase privacy violations</p>
	Visualizations of social networks	Network perception: People don’t perceive social networks accurately [39]	

CONCLUSION

Social transparency, the ability to observe and monitor the interactions of others within and across applications on the Internet is transforming the way we collaborate and share information online. Although the field of CSCW has conceptualized social translucence and awareness within a small group setting, we lack a framework for researching social transparency across a larger scale. In this paper, we have presented such a framework for analyzing social transparency on the Internet. This framework extends the notion of social translucence and activity awareness, and suggests new research on the effects and design of socially transparent collaborative work systems. A summary of the dimensions of social transparency, their related inferences and second order effects can be found in Table 1.

We argue that designers of social systems need to consider the inferences that transparency supports and its implications for broader social outcomes. In the offline world, cues about others behavior serve as powerful signals that inform our own activities. In the online setting, designs

that incorporate social transparency are recreating many of the signals we rely on in our offline environment with potentially broader more far-reaching consequences. Informed by the existing research on social psychology, decision-making, and interaction in online communities, our framework suggests inferences individuals are likely to make based on social transparency cues. When aggregated across individuals, transparency related inferences in turn have implications for the social structures individuals are embedded in. We have considered here the influence of identity, content and interaction transparency on group and organizational-level outcomes such as information quality, creativity, productivity, stress, and herding (summarized in Table 1 above). The implications of social transparency for collaborative systems suggest a host of interesting research opportunities in CSCW. In particular, the second-order system level effects of social transparency are understudied. We insufficiently understand the social context for positive and negative effects of transparency--whether information quality will improve or creativity will decline, whether

collaborations will be more effective or more stressful, and whether communities will attend too much to popular content or will be exposed to new content they might not otherwise have seen.

ACKNOWLEDGMENTS

This work was supported by a National Science Foundation grant, CNS-1040801.

REFERENCES

1. Bardram, J.E. and Hansen, T.R. The AWARE architecture: Supporting context-mediated social awareness in mobile cooperation. *CSCW 04*, (2004), 192-201.
2. Barley, S.R., Meyerson, D.E., and Grodal, S. E-mail as a source and symbol of stress. *Organization Science* 22, 4 (2010), 887-906.
3. Barnlund, D.C. and Harland, C. Proximity and prestige as determinants of communication networks. *Sociometry* 26, 4 (1963), 467-479.
4. Bearden, W.O. and Etzel, M.J. Reference group influence on product and brand purchase decisions. *Journal of Consumer Research* 9, 2 (1982), 183-194.
5. Bernstein, M., Monroy-Hernandez, A., Harry, D., Andre, P., Panovich, K., and Vargas, G. 4chan and /b/: An analysis of anonymity and ephemerality in a large online community. *ICWSM*, Association for the Advancement of Artificial Intelligence (2011).
6. Boh, W. Mechanisms for sharing knowledge in project-based organizations. *Information and Organization* 17, 1 (2007), 27-58.
7. Boh, W.F. Reuse of knowledge assets from repositories: A mixed methods study. *Information & Management* 45, 6 (2008), 365-375.
8. Borgatti, S.P. and Halgin, D.S. On network theory. *Organization Science*, , 1-14.
9. Bourdieu, P. Le capital social: Notes provisoires. *Actes de la Recherche en Sciences Sociales* 31, (1980), 2-3.
10. Butler, D. Computing giants launch free science metrics. *Nature*, 2011. <http://www.nature.com/news/2011/110802/full/476018a.html>.
11. Cascio, W.F. Managing a virtual workplace. *Academy of Management Executive* 14, 3 (2000), 81-90.
12. Castano, E., Yzerbyt, V., Paladino, M.P., and Sacchi, S. I belong, therefore, I exist: Ingroup identification, ingroup entitativity, and ingroup bias. *Personality and Social Psychology Bulletin* 28, 2 (2002), 135-143.
13. Cialdini, R.B. and Goldstein, N.J. Social influence: Compliance and conformity. *Annual review of psychology* 55, 1974 (2004), 591-621.
14. Cialdini, R.B., Reno, R.R., and Kallgren, C.A. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology* 58, 6 (1990), 1015-1026.
15. Clark, H.H. and Brennan, S.E. Grounding in communication. In L.B. Resnick, J.M. Levine and S.D. Teasley, eds., *Perspectives on socially shared cognition*. American Psychological Association, 1991, 127-149.
16. Clifford, S. Will Google's Chrome help or hurt advertisers? *The New York Times*, 2008. <http://bits.blogs.nytimes.com/2008/09/03/will-googles-chrome-help-or-hurt-advertisers/?scp=6&sq=incognito chrome&st=cse>.
17. Connolly, T., Jessup, L.M., and Valacich, J.S. Effects of anonymity and evaluative tone on idea generation in computer-mediated groups. *Management Science* 36, 6 (1990), 689-703.
18. Cosley, D., Frankowski, D., Kiesler, S., Terveen, L., and Riedl, J. How oversight improves member-maintained communities. *CHI 05*, (2005), 11.
19. Cramton, C.D. The mutual knowledge problem and its consequences for dispersed collaboration. *Organization Science* 12, 3 (2001), 346-371.
20. Cranor, L. Internet privacy, a public concern. *netWorker: The Craft of Network Computing June/July*, (1998), 13-18.
21. Cummings, J. Geography is alive and well in virtual teams. *Communications of the ACM* 54, (2011), 24-26.
22. Dahl, S. Turnitin(R): The student perspective on using plagiarism detection software. *Active Learning in Higher Education* 8, 2 (2007), 173-191.
23. Dourish, P. and Bellotti, V. Awareness and coordination in shared workspaces. *CSCW 92* 3, November (1992), 107-114.
24. Ellison, N., Heino, R., and Gibbs, J. *Managing Impressions Online: Self-Presentation Processes in the Online Dating Environment*. Blackwell Publishing Inc, 2006.
25. Erickson, T. and Kellogg, W.A. Social translucence: An approach to designing systems that support social processes. *CHI 00* 7, (2000), 59-83.
26. Geambasu, R., Kohno, T., Levy, A.A., and Levy, H.M. Vanish: Increasing data privacy with self-destructing data. *Usenix Security Symposium*, USENIX Association (2009), 299-316.
27. Gentzkow, M. and Shapiro, J.M. Ideological segregation online and offline. *NBER Working Paper w15916*, 10 (2010).
28. Gilbertson, S. Google Hotpot smartens up local search, but it's no Yelp killer. *Wired*, 2010. <http://www.wired.com/epicenter/2010/11/google-hotpot-smartens-up-local-search-but-its-no-yelp-killer/>.
29. Godefroid, P., Herbsleb, J., Jagadeesany, L., and Li, D. Ensuring privacy in presence awareness: An automated verification approach. *CSCW 2000*, (2000), 59-68.
30. Goldstein, N.J., Cialdini, R.B., and Griskevicius, V. A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research* 35, 3 (2008), 472-482.
31. Granovetter, M.S. The strength of weak ties. *American Journal of Sociology* 78, 6 (1973), 1360-1380.
32. Gutwin, C. and Greenberg, S. A descriptive framework of workspace awareness for real-time groupware. *CSCW 02* 11, 3 (2002), 411-446.
33. Gutwin, C., Greenberg, S., and Roseman, M. Workspace awareness in real-time distributed groupware: Framework, widgets, and evaluation. *CHI 96*, (1996), 281-298.
34. Guy, I., Jacovi, M., Perer, A., Ronen, I., and Uziel, E. Same places, same things, same people? Mining user similarity on social media. *CSCW 10*, (2010), 41-50.
35. Hinds, P.J. and Mortensen, M. Understanding conflict in geographically distributed teams: The moderating effects of shared identity, shared context, and spontaneous communication. *Organization Science* 16, 3 (2005), 290-307.
36. Hintzman, D.L. Research strategy in the study of memory: Fads, fallacies, and the search for the "coordinates of truth." *Perspectives on Psychological Science* 6, 3 (2011), 253-271.
37. Karau, S.J. and Williams, K.D. Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology* 65, 4 (1993), 681.
38. Kiesler, S. and Sproull, L. Group decision making and communication technology. *Organizational Behavior and Human Decision Processes* 52, 1 (1992), 96-123.
39. Kilduff, M. and Krackhardt, D. Bringing the individual back in: A structural analysis of the internal market for reputation in organizations. *Academy of Management Journal* 37, 1 (1994), 87-108.
40. Kittur, A., Suh, B., Chi, E.H., and Alto, P. Can you ever trust a Wiki? Impacting perceived trustworthiness in Wikipedia. *CSCW 08*, (2008), 7-10.
41. Kogut, B. and Zander, U. Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science* 3, 3 (1992), 383-397.
42. Kraut, R.E., Burke, M., Riedl, J., and Resnick, P. Dealing with newcomers. In R.E. Kraut and P. Resnick, eds., *Evidence-based social design: Mining the social sciences to build online communities*. MIT Press.

43. Krishnamurthy, B., Naryshkin, K., and Wills, C. Privacy leakage vs. protection measures: The growing disconnect. *Web 2.0 Security and Privacy Workshop*, (2011).
44. Lampe, C. and Resnick, P. Slash (dot) and burn: Distributed moderation in a large online conversation space. *CHI 04*, (2004), 543-550.
45. Lampe, C., Ellison, N., and Steinfield, C. A familiar face(book): Profile elements as signals in an online social network. *CHI 07*, (2007), 435-444.
46. Lerman, K. and Jones, L. Social browsing on Flickr. *Proceedings of International Conference on Weblogs and Social Media*, arxiv.org (2006).
47. Madrigal, A. Why Facebook and Google's concept of 'real names' is revolutionary. *The Atlantic*, 2011. <http://m.theatlantic.com/technology/archive/2011/08/why-facebook-and-googles-concept-of-real-names-is-revolutionary/243171>.
48. Marwick, A.E. and boyd, D. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society* 13, 1 (2010), 114-133.
49. Mas, A. and Moretti, E. Peers at work. *American Economic Review* 99, 1 (2009), 112-145.
50. McFarlane, D. and Latorella, K. The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction* 17, 1 (2002), 1-61.
51. McKenna, K.Y.A. and Bargh, J.A. Coming out in the age of the Internet: Identity "demarginalization" through virtual group participation. *Journal of Personality and Social Psychology* 75, 3 (1998), 681-694.
52. McPherson, M., Smith-Lovin, L., and Cook, J.M. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27, 1 (2001), 415-444.
53. Metzger, M.J., Flanagin, A.J., Eyal, K., Lemus, D.R., and McCann, R.M. Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. In *Communication Yearbook*. Lawrence Erlbaum, 2003, 293-335.
54. Nagarajan, M., Purohit, H., and Sheth, A. A qualitative examination of topical tweet and retweet practices. *AAAI Conference on Weblogs and Social Media*, (2010), 295-298.
55. Nonnecke, B., East, K.S., and Preece, J. Why lurkers lurk. *Americas Conference on Information Systems*, (2001), 1-10.
56. Nonnecke, B., Preece, J., and Andrews, D. What lurkers and posters think of each other. *Proceedings of the International Conference on System Sciences*, (2004), 1-9.
57. Ogawa, M. and Ma, K.-L. Code Swarm: A design study in organic software visualization. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (2009), 1097-1104.
58. Ogawa, M. and Ma, K.-L. code_swarm: A design study in organic software visualization. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (2009), 1097-1104.
59. Ott, M., Choi, Y., Cardie, C., and Hancock, J.T. Finding deceptive opinion spam by any stretch of the imagination. *Computational Linguistics*, (2011), 309-319.
60. Ren, Y., Kraut, R., and Kiesler, S. Encouraging commitment in online communities. In R. Kraut and P. Resnick, eds., *Evidence-based social design: Mining the social sciences to build online communities*. MIT Press, Cambridge, MA.
61. Resnick, P., Kuwabara, K., Zeckhauser, R., and Friedman, E. Reputation systems. *Communications of the ACM* 43, 12 (2000), 45-48.
62. Rieh, S.Y. and Danielson, D.R. Credibility: A multidisciplinary framework. *Annual Review of Information Science and Technology* 41, 1 (2007), 307-364.
63. Rogers, E. *Diffusion of innovations*. The Free Press, New York, 1995.
64. Salganik, M.J., Dodds, P.S., and Watts, D.J. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311, 5762 (2006), 854-856.
65. Sassenberg, K. and Postmes, T. Cognitive and strategic processes in small groups: Effects of anonymity of the self and anonymity of the group on social influence. *British Journal of Social Psychology* 41, (2002), 463-80.
66. Scupelli, P., Kiesler, S., and Fussell, S.R. Project view IM: A tool for juggling multiple projects and teams. *CHI 05*, (2005), 1773.
67. Shannon, C.E. A mathematical theory of communication. *Bell System Technical Journal* 27, July 1928 (1948), 379-423.
68. Shklovski, I. and Kotamraju, N. Online contribution practices in countries that engage in Internet blocking and censorship. *CHI 11*, (2011), 1109-1118.
69. Simmel, G. Individual and society. In K. Wolff, ed., *The sociology of Georg Simmel*. Free Press, New York, 1950, 145-169.
70. Solano, C.H. and Dunnam, M. Two's company: Self-disclosure in triads versus dyads. *Social Psychology Quarterly* 48, 2 (1985), 183-187.
71. Souza, C.R.B. De, Redmiles, D., and Dourish, P. "Breaking the code", moving between private and public work in collaborative software development. *Proceedings of the International ACM SIGGROUP conference on Supporting group work*, ACM Press (2003), 105-114.
72. Sproull, L. and Kiesler, S. *Connections: New ways of working in the networked organization*. MIT Press, Cambridge, MA, 1992.
73. Steel, E. and Fowler, G. Facebook in privacy breach. *Wall Street Journal*, 2010. <http://online.wsj.com/article/SB10001424052702304772804575558484075236968.html>.
74. Suh, B., Chi, E.H., Kittur, A., and Pendleton, B.A. Lifting the veil: Improving accountability and social transparency in Wikipedia with wikidashboard. *CHI 08*, (2008), 1037-1040.
75. Thompson, L.F., Sebastianelli, J.D., and Murray, N.P. Monitoring online training behaviors: Awareness of electronic surveillance hinders e-learners. *Journal of Applied Social Psychology* 39, 9 (2009), 2191-2212.
76. Treude, C. and Storey, M.-A. Awareness 2.0: Staying aware of projects, developers and tasks using dashboards and feeds. *Work*, ACM (2010), 365-374.
77. Viégas, F.B., Wattenberg, M., and Dave, K. Studying cooperation and conflict between authors with history flow visualizations. *CHI 04* 6, (2004), 575-582.
78. Volda, A., Volda, S., Greenberg, S., and He, H.A. Asymmetry in media spaces. *CSCW 08*, (2008), 313-322.
79. Vrij, A., Granhag, P.A., and Porter, S. Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest* 11, 3 (2011), 89-121.
80. Wicker, S., and Schrader, D. Privacy-aware design principles for information networks. *Proceedings of the IEE 99*, 2 (2011), 330-350.