Design Considerations for Collaborative Visual Analytics

Jeffrey Heer, Maneesh Agrawala
University of California, Berkeley

ABSTRACT

Information visualization leverages the human visual system to support the process of sensemaking, in which information is collected, organized, and analyzed to generate knowledge and inform action. Though most research to date assumes a single-user focus on perceptual and cognitive processes, in practice, sensemaking is often a social process involving parallelization of effort, discussion, and consensus building. This suggests that to fully support sensemaking, interactive visualization should also support social interaction. However, the most appropriate collaboration mechanisms for supporting this interaction are not immediately clear. In this article, we present design considerations for asynchronous collaboration in visual analysis environments, highlighting issues of work parallelization, communication, and social organization. These considerations provide a guide for the design and evaluation of collaborative visualization systems.

CR Categories and Subject Descriptors: H.5.2. User Interfaces, H.5.3 Group and Organization Interfaces
Additional Keywords: visualization, analysis, collaboration, design, computer-supported cooperative work

1 INTRODUCTION

Information visualization technologies leverage the human visual system to support analysis and communication of large amounts of information. However, visual analysis and decision making often involve not only perceptual and cognitive processes, but social processes. People may disagree on how to interpret data and contribute contextual knowledge that deepens understanding. As participants build consensus or make decisions they learn from their peers. Furthermore, some data sets are so large that thorough exploration by a single person is unlikely. Such scenarios arise in business intelligence [36], intelligence analysis [37, 47], and public data consumption [18]. In this spirit, a recent report [47] names the design of collaborative visualization tools as a grand challenge for visualization research.

While existing visualization research has explored techniques for collocated collaboration (*e.g.*, large displays and shared workspaces) and synchronous distance work (*e.g.*, real-time networked displays), little research attention has been paid to asynchronous collaboration around visualizations [48]. By partitioning work across both time and space, asynchronous collaboration offers greater scalability for group-oriented analysis. There is evidence that, due in part to a greater division of labor, asynchronous decision making can result in higher-quality outcomes—broader discussions, more complete reports, and longer solutions—than face-to-face collaboration [1].

One challenge to achieving such benefits is determining the appropriate design decisions and technical mechanisms to enable

E-Mail: jheer@cs.berkeley.edu, maneesh@cs.berkeley.edu
Address: Soda Hall, UC Berkeley, Berkeley, CA 94720-1776

IEEE Symposium on Visual Analytics Science and Technology 2007
October 30 - November 1, Sacramento, CA, USA
978-1-4244-1659-2/07/$25.00 ©2007 IEEE

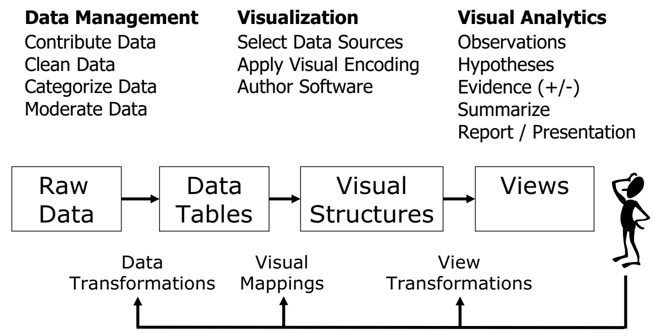

Figure 1. The Information Visualization Reference Model. Source data is mapped into data tables which are visually encoded and presented in interactive views [9, 28]. Collaboration may occur at the level of data management, visualization, or analysis.

and catalyze social data analysis around visual media. Previously, we began exploring this space by building and evaluating *sense.us*, a system for asynchronous collaborative visualization [29]. Our observations of usage have provided numerous examples of group sensemaking in action: cycles of observation, question, and hypothesis; social navigation to interesting or controversial data; and identification of problematic or incorrect data values. We wish to better support these observed behaviors by grounding our design decisions in both theoretical and practical knowledge of group interaction. Furthermore, additional systems have recently been introduced to support collaborative analysis around both statistical and geographic data (see Figure 2); each supports simple text comments and view sharing through bookmarking. A theoretically-grounded design framework can be applied to contrast these existing offerings and guide the future research and development of social visual analysis systems.

Creating effective mediated collaboration environments raises a number of design questions. How should collaboration be structured, and what shared artifacts can be used to coordinate contributions? What are the most effective communication mechanisms? Based upon our experiences to date and a survey of research in analytics, social psychology, sociology, organizational studies, and computer-supported cooperative work, we identify a set of design considerations to inform the development of asynchronous collaborative information visualization systems. We have grouped our design considerations into seven topical areas: Division and allocation of work; Common ground and awareness; Reference and deixis; Incentives and engagement; Identity, trust, and reputation; Group dynamics; and Consensus and decision making. In each of these areas, we discuss the underlying *accomplishments* that enable effective collaboration, and suggest specific *mechanisms* by which they could be achieved.

As a thorough treatment of these subjects would warrant multiple volumes, we attempt only to identify key issues to guide work in collaborative visualization. After discussing each topical area in turn, we conclude by summarizing the various design considerations presented and suggesting avenues for future research and development in collaborative visual analytics.

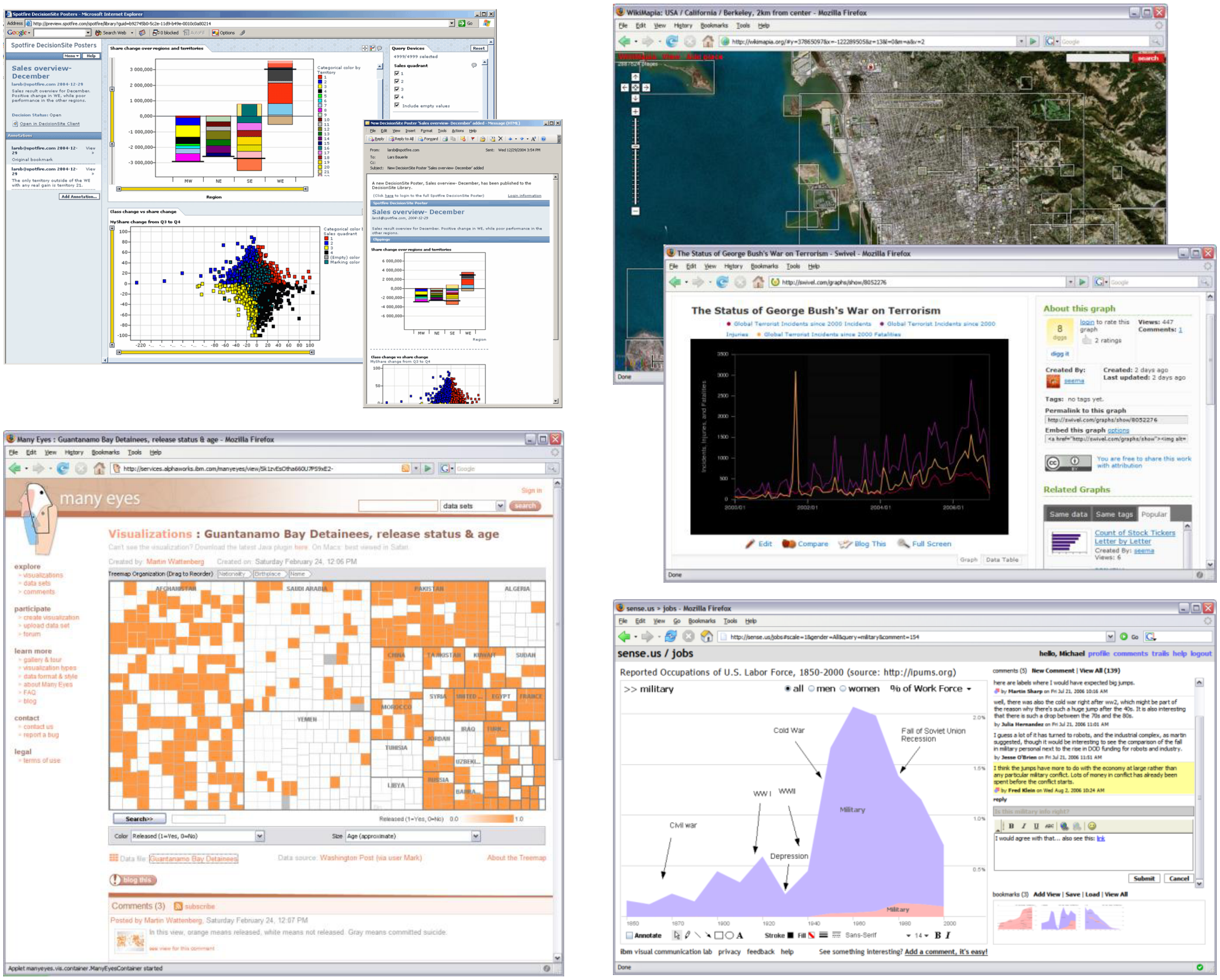

Figure 2. Asynchronous Collaborative Visualization Systems. Clockwise from top-left, Spotfire Decision Site Posters [43], Wikimapia [50], Swivel [46], Sense.us [29], and Many Eyes [33]. These systems support varied levels of sharing, discussion, and annotation of visualized data.

2 Division and Allocation of Work

A fundamental aspect of successful collaboration is an effective division of labor among participants. This involves both the segmentation of effort into proper units of work and the allocation of individuals to tasks in a manner that best matches their skills and disposition. Primary concerns are how to split work among multiple participants and meaningfully aggregate the results.

Benkler [2] describes the role of modularity, granularity, and cost of integration in the peer production of information goods, drawing on examples such as online discussions, open source software, and Wikipedia. *Modularity* refers to how work is segmented into atomic units, parallelizing work into independent tasks. The *granularity* of a module is a measure of the cost or effort involved in performing the task. The optimal granularity of modules is closely tied to the incentives for performing the work. For example, in online scenarios where the incentives tend to be small and non-monetary, a small granularity is needed to facilitate work, encouraging people to participate in part due to the ease of contributing. A variety of granularities enables different classes of contribution to emerge.

The third aspect of Benkler's model is the *cost of integration*: what effort is required to usefully synthesize the contributions of each individual module? Collaborative work will only be effective if the cost of integration is low enough to warrant the overhead of modularization while enforcing adequate quality control. There are a number of mutually inclusive approaches to handling integration: automation (automatically integrating work through technological means), peer production (casting integration as an additional collaborative task given to trusted participants), social norms (using social pressures to reduce vandalistic behavior), and hierarchical control (exercising explicit moderation).

Questions for collaborative visualization include how to facilitate the modularization of work. The first step is determining the modules of work and their granularity. Existing frameworks for aiding this task include structural models of visualization design and sensemaking processes. Once modules have been identified, one can then attempt designs which reduce the cost structure for these tasks. Another important concern is the proscription of particular task types or roles—what aspects should be formally inscribed in the system and what should be left open to negotiation and definition by work groups themselves?

2.1 The Information Visualization Reference Model

One model for identifying modules of contribution is the *information visualization reference model* [9, 28], a general pattern for describing visualization applications (Figure 1). The model decomposes the visualization process into data acquisition and representation, visual encoding of data, and display and interaction. Each phase of this model provides an entry point for collaborative activity. Contributions involving data include uploading data sets, cleaning or reformatting data, moderating contributed data (e.g., to safeguard copyright or privacy concerns), and affixing metadata (e.g., providing keyword tags). Additional contributions of varying granularity lie in the application of visual encodings. Examples include matching data sets with existing visualization components, editing visual mappings to form more effective visualizations, and authoring visualization software components. Both Many Eyes [33] and Swivel [46] enable contribution of data sets and visual mappings. The primary focus of this paper, however, is at the level of interaction, where we consider how collaborative visual analysis and exploration can effectively be conducted.

2.2 THE SENSEMAKING MODEL

To better understand analytic contributions, we consult the sensemaking model [9, 39], which grounds the use of information visualization in a theory of how people search for, organize, and create new knowledge from source information. Social issues accrue at each phase of the model: how do people communicate, how do they judge others' contributions, how are groups formed, and what motivates contributions? Each of these issues is addressed in subsequent sections. As indicated by the numerous interconnections in Figure 3, the sensemaking process has a much higher degree of coupling than the information visualization reference model, carrying implications for the granularity and integration of contributions.

Intelligence analysis provides examples of both cooperative and competitive models of work [47]. In cooperative scenarios, modules may be of fine granularity and pooled such that collaborators can immediately benefit from the work of others. Examples include identifying relevant information sources, connections between sources, and positing hypotheses. Such work may involve tightly coupled collaboration, requiring awareness and communication among participants. In competitive scenarios, modules are larger and work is not integrated until a later stage of sensemaking, such as detailed, evidence-backed hypotheses or recommended actions. While lacking the benefits of resource pooling, this approach encourages individual assessment and can reduce groupthink bias. Accordingly, it may benefit collaborative visualization systems to support both fine-grained and coarse-grained work parallelization.

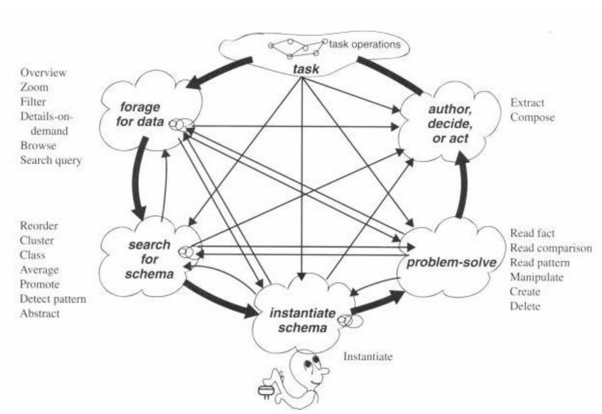

Figure 3. The Sensemaking Cycle. The diagram depicts the various phases and loops of the sensemaking process, annotated with common tasks. The image is taken from Card et al [9].

If adopting a competitive model, the main concern is with integrating the end results of the sensemaking process. How can analytic conclusions or suggested actions be presented, compared, and evaluated? This gives rise to a consensus and decision making problem of its own, an issue discussed later. If cooperative models are used, either across all collaborators or within teams, we should consider social issues affecting each phase of sensemaking.

2.2.1 INFORMATION FORAGING

The first such phase is information foraging [36]. Given the underlying metaphor of foraging for food, an activity often performed by social packs of animals, social information foraging [35] seems a natural extension. This argues for collaborators to pool findings, such as discovery of relevant information, and to support notification updates and information retrieval. Challenges include formalizing contributions, such as identifying trends or outliers of interest and positing explanatory hypotheses, and providing retrieval mechanisms by which others can access them. Additional possibilities lie in analyzing and displaying activity traces to facilitate social navigation [20], metaphorically similar to the scent trails left by ants foraging for food. In this form, general usage itself can be treated as an implicit module of work, a possibility discussed further in section 3.2.

2.2.2 INFORMATION SCHEMATIZATION

The next phases of sensemaking concern the construction and population of information schemata. This could be conducted in a general form by enabling discussion amongst collaborators. One challenge is to synthesize the results of discussion into more accessible forms, such as summaries of arguments and evidence. The cost structure of these tasks could be further reduced, and the integration of contributions facilitated by, providing additional shared artifacts or external representations [53] for structuring group work. For example, the analytic sandbox of [51] provides a visual environment for spatially organizing hypotheses and positive and negative evidence, while [3] describes a system for collaborative use of analytic evidence matrices.

2.2.3 PROBLEM-SOLVING, DECISION-MAKING, AND ACTION

The final phases of sensemaking involve problem-solving and action. This may or may not take place within the collaborative analysis environment. Findings gained from analysis may serve as input to collaboration in other media, suggesting the need to both facilitate external access to the contents of the visual analysis environment and extracting content for use in other systems. If problem-solving and decision making are conducted within the system, aforementioned issues regarding communication, discussion, and consensus must be addressed.

3 COMMON GROUND AND AWARENESS

Inspired by linguistics, social psychologists have investigated fundamental prerequisites for successful communication. Clark and Brennan describe the concept of *common ground* [15], the shared understanding between conversational participants enabling communication. Through shared experience and discussion, people constantly monitor their mutual understanding. For example, facial expressions, body language, and backchannel utterances such as "uh-huh" and "hmm?" provide *grounding* cues of a participant's current level of understanding. Both positive evidence of convergence of understanding and negative evidence of misunderstanding are used to establish common ground.

Interestingly, an imperfect shared understanding is often sufficient. *The principle of least collaborative effort* states that conversational participants will exert just enough effort to achieve successful communication [13]. Collaborative effort may be applied during both a *planning* stage, in which a participant formulates their next utterance, and an *acceptance* stage, in which a participant ascertains if partners have understood the utterance. This principle serves as an evaluation guide for collaboration mechanisms based on their effect upon the cost structure of interaction. For example, multiple studies have shown that the media of communication affects the cost structure of collaborative effort [4,21]: views of a shared visual environment minimize the need to verbally confirm actions that can be assessed visually. However, media effects such as latency can hamper the efficiency benefits of such cues [21].

At both general and detailed levels, grounding theory provides a useful guide for design decisions. When collaborating around visualizations, participants must be able to see the same visual environment in order to ground each others' actions and comments, suggesting the need for mechanisms for bookmarking or sharing specific states of the visualization. This includes both sharing within the visualization environment itself and across other media. For example, the results of visual analysis might effectively be shared embedded in an external web page, where common ground is better established within a dedicated, familiar readership. At minimum, the ability to easily pass around pointers (e.g., URLs) to specific views is indispensable. This entails that collaborative visualizations be able to explicitly represent and export their internal state space [29,48].

3.1 Discussion Models

Given the ability to access a shared viewpoint, one must still determine the forms of discussion and annotation around that view. For example, one could use visualization bookmarks within a standard discussion forum, interspersing links to desired views within the text. This form of *independent discussion* is unidirectional, linking from text to the visualization. Independent, unthreaded comments are used by both Decision Site Posters [43] and Many Eyes [33]. Another approach is *embedded discussion*, placing conversational markers directly within the visualization, such as comments over annotated geographic regions in Wikimapia [50]. This approach provides unidirectional links that point from the visualization to text.

Grounding might be further facilitated by more deeply tying independent discussion to the visualization state space. *Doubly-linked* commentary [29] allows comments to link to specific views as in independent discussion, while also enabling all such discussions to be retrieved in situ as visualization views are visited. Our hypothesis is that directly associating commentary with specific states of the visualization will facilitate grounding by disambiguating the context of discussion, while also enabling serendipitous discovery of relevant discussion during exploration. Evidence for this hypothesis could take the form of simplified referential utterances or facilitation of reader comprehension.

3.2 Awareness

Another important source of grounding comes from *awareness* of others' activities, allowing collaborators to gauge what work has been done and where to allocate effort next [10,19]. Within asynchronous contexts, participants require awareness of the timing and content of past actions. This suggests that designs should include both *history* and *notification* mechanisms (e.g., [6]) for following actions performed on a given artifact or by specific individuals or groups. Browseable histories of past action are one viable mechanism, as are subscription and notification technologies such as RSS (Really Simple Syndication) and Atom.

User activity can also be aggregated and abstracted to provide additional forms of awareness. *Social navigation* [20] involves the use of activity traces to provide additional navigation options, allowing users to purposefully navigate to past states of high interest or explore less-visited regions (the "anti-social navigation" of [49]). For example, navigation cues may be added to links to views with low visitation rates or to action items such as unanswered questions and unassessed hypotheses.

4 Reference and Deixis

A vital aspect of grounding is successfully referring to artifacts, people, places, or other items. As discussed by both Clark [12] and Brennan [4], reference can take on many different forms; we focus on reference in spatial contexts. When collaborating around visual media, it is common to refer to specific objects, groups, or regions visible to participants. Such references may be *general* (e.g., "north by northwest"), *definite* (e.g., named entities), *detailed* (e.g., described by attributes, such as the "blue ball"), or *deictic* (e.g., pointing to an object and saying "that one", also referred to as *indexical* reference). Once the referent has been successfully established and grounding has been achieved between participants, collaboration can move forward.

Clark [12] surveys various forms of spatial indexical reference, grouping them into the categories of *pointing* and *placing*. Pointing behaviors use some form of vectorial reference to direct attention to an object, group, or region of interest, such as pointing a finger or directing one's gaze. Placing behaviors involve moving an object to a region of space that has a shared, conventional meaning. Examples include placing groceries on a counter to indicate items for purchase and standing across from the teller to indicate that you will be the purchaser. In addition to directing attention, indexical reference allows patterns of speech and text to change. Participants can use deictic terms like "that" and "there" to invoke indexical referents, simplifying the production of utterances along the principle of least collaborative effort.

4.1 Pointing

Hill and Hollan [30] further discuss the various roles that deictic pointing gestures can play, often communicating intents more complicated than simply "look here". For example, different hand gestures can communicate angle (oriented flat hand), height (horizontal flat hand), intervals (thumb and index finger in "C" shape), groupings (lasso'ing a region), and forces (accelerating fist). They go on to state that successfully supporting deixis is key to the future of visualization applications.

While other forms of reference are often most easily achieved through speech or written text, deictic reference in particular offers important interface design challenges for collaborative visualization. Our hypothesis is that methods for performing nuanced pointing behaviors can improve collaboration by favorably altering its cost structure. Hill and Hollan make this claim explicitly, arguing for "generally applicable techniques that realize complex pointing intentions" by engaging "pre-attentive vision in the service of cognitive tasks".

An additional concern is ambiguity of reference. Clark et al [14] demonstrate how people's common ground can affect ambiguity resolution. Thus, two people with greater familiarity might successfully communicate using ambiguous references, while a third participant remains confused. By providing interaction techniques for pointing that facilitate unambiguous references, designers might not only aid human communication, but allow for machine-readable forms of pointing or annotation, supporting a navigable index of references. For example, this could allow users to search for all commentary or visualizations that reference a particular data item.

Another design consideration is how various forms of reference may be applied in tandem. For example, one might deictically refer to a particular object, but formulate a broader selection by abstracting from the properties of that object (e.g., "select all items that are blue like this one"). The implicit interplay between gesture and text, often segmented in time and interpreted subconsciously in synchronous interactions, may need to be more concretely reified in asynchronous contexts. For example, a text comment involving multiple deictic terms may need to link those terms explicitly to visual annotations, as the gestural cues used in face-to-face communication are not available for disambiguation.

5 Incentives and Engagement

If collaborators are professionals working within a particular context (e.g., financial analysts or research scientists) there may be existing incentives, both financial and professional, for conducting collaborative work. In a public goods scenario, incentives such as social visibility or sense of contribution may be the motivating factors. Incorporating incentives into the design process may increase the quantity and/or quality of contributions, and could even provide additional motivation in those situations that already have well established incentive systems.

Benkler [2] posits an incentive structure for collaborative work consisting of monetary incentives, hedonic incentives, and social-psychological incentives. *Monetary* incentives refer to material compensation such as a salary or cash reward. *Hedonic* incentives refer to well-being or engagement experienced intrinsically in the work. *Social-psychological* incentives involve perceived benefits such as increased status or social capital.

5.1 Personal Relevance

A number of observations of social use of visualization have noted an affinity of visualization users for data which they find personally relevant [27,48,49]. For example, collaborative visual analysis of the occupations of American workers [29] often started by searching for their own profession and those of their friends and family, similar to how people searched for names in the popular NameVoyager visualization [49]. The hypothesis is that by selecting data sets or designing their presentation such that the data is seen as personally relevant, usage rates will rise due to increased hedonic incentive. For example, geographic visualizations can facilitate navigation to personally relevant locations through typing in specific zip codes or city names.

5.2 Social-Psychological Incentives

In the case of social-psychological incentives, the visibility of contributions can be manipulated for social effects. Ling et al [32] found that users contributed more if reminded of the uniqueness of their contribution or if given specific challenges, but not under other theoretically-motivated conditions. Cheshire [11] ran a controlled experiment finding that, even in small doses, positive social feedback on a contribution greatly increases contributions. He also found that visibility of high levels of cooperative behavior across the community increases contributions in the short term, but has only moderate impact in the long term. These studies suggest that social-psychological incentives can improve contribution rates, but that the forms of social visibility applied have varying returns. One such incentive for visual analysis is to prominently display new discoveries or successful responses to open questions. Mechanisms for positive feedback, such as voting for interesting comments, might also foster more contributions.

5.3 Game Play

Finally, it is worth considering game play as an additional framework for increasing incentives. In contrast to environments such as spreadsheets, many visualizations already enjoy game-like properties, being highly visual, highly interactive, and often animated. Heer [27] discusses various examples in which playful activity contributes to analysis, applying insights from an existing theory of playful behavior [8] that analyzes the competitive, visceral, and teamwork building aspects of play. For example, scoring mechanisms could be applied to create competitive social-psychological incentives. Game design might also be used to allocate attention, for example, by creating a team-oriented "scavenger hunt" analysis game focused on a particular subject matter. Salen and Zimmerman [40] provide a thorough resource for the further study of game design concepts.

6 Identity, Trust, and Reputation

Aspects of identity, reputation, and trust all influence the way people interact with each other. Other things being equal, a hypothesis suggested by a more trusted or reputable person will have a higher probability of being accepted as part of the group consensus [34]. For social sensemaking in a computer-mediated environment, design challenges accrue around the various markers of identity and past action that might be transmitted through the system. For example, Donath [17] describes how even a cue as simple as one's e-mail address can lead to a number of inferences about identity and status.

6.1 Identity Presentation

Many theorists try to understand interpersonal perception via the signals available for interpretation by others. Goffman [23] distinguishes between *expressions given* and *expressions given off* to indicate those parts of our presentation of self that are consciously planned (e.g., the content of our speech) or unconsciously generated (e.g., a wavering of voice indicating nervousness), each of which is interpreted to form opinions of a person. Donath [17] classifies such signals into *conventional signals*—low cost signals that are easy to fake (e.g., talking about going to the gym)—and *assessment signals*—more reliable signals that are difficult to fabricate (e.g., having large muscles).

When considering the implications of identity assessment for designing collaborative visualization systems, it is important to take into account the context of deployment. If collaborators are already familiar to each other, there may be little that needs to be done to support identity and reputation formation, as there are existing channels through which this can be conducted. It may be enough to simply identify collaborators' individual contributions with recognizable names. Many organizations maintain online personnel directories to aid awareness and collaboration; visual analysis systems should be able to leverage such existing artifacts.

On the other hand, if collaborators begin as strangers, mechanisms for self-presentation and reputation formation need to be included in the system design. Possible mechanisms include identity markers, such as screen names, demographic profiles, social networks, and group memberships. Considerations include the type of personal information germane to the context of visual analysis; for example, is a playful or professional environment desired? Attributes such as age, geographic location, interests, and skills might help assess a collaborator's background knowledge, affecting the confidence one places in hypotheses. Of course, this picture is complicated if such measures are self-reported conventional signals subject to fabrication. This raises the challenge of crafting assessment signals of identity and reputation.

6.2 Reputation Formation

Considering how interpersonal assessment develops over time gives rise to questions of reputation and trust formation. In the case where participants only interact through the system itself, means of gauging a user's past actions or contributions are needed to not only aid awareness (cf. §3) but to facilitate reputation formation. Observations of past actions provide *implicit* means of reputation formation, allowing collaborators to make interpersonal judgments grounded in past activity. One challenge for design is to consider what pieces of information are most informative for reputation formation.

Some systems instead provide *explicit* reputation mechanisms, such as seller ratings in online markets such as eBay [38]. In a visual analysis environment, collaborators might rate each other's contributions according to their interestingness or accuracy. This may help surface contributions with higher relevance, provide a reputation metric for contributors, and provide a social-psychological incentive for high quality contributions.

7 Group Dynamics

The makeup of collaborative groups is another aspect important to social sensemaking. Many scenarios, such as business and research, may involve work groups that are already well established. In such cases, standard group management and communication features common to many collaborative applications may be sufficient. However, when organizing effort in public goods scenarios, explicit mechanisms for assisting group formation may aid collaborative visualization efforts.

7.1 Group Management

At a basic level, formal *group management* mechanisms present useful means for addressing issues of scalability and privacy. Group management mechanisms can support the coordination of a work group on a specific task within a larger collaborative environment, providing notification and awareness features at the group level. Groups also provide a means of filtering contributions, improving tractability and reducing information overload for participants who may not be interested in the contributions of strangers. Finally, groups provide a means of limiting contribution visibility, providing one mechanism for individual privacy within large-scale online scenarios.

7.2 Group Size

One challenge for group management is the choice of group size. Larger groups may be able to achieve more through a larger labor pool, but can incur social and organizational costs. For example, larger groups are more likely to suffer from the *free rider* problem due to diluted accountability [25]. Pirolli [35] describes a mathematical model of social information foraging that measures the benefit of including additional collaborators in information gathering tasks. His analysis finds that beyond certain sizes, additional foragers provide decreasing benefits, suggesting that an optimal group size exists, dependent on the parameters of the foraging task. A useful future experiment would be to apply Pirolli's framework to real visual analysis teams.

7.3 Group Diversity

Another issue facing group formation is the diversity of group members. In this case diversity can include the distribution of domain-specific knowledge among potential participants and other differences such as geographical location, culture, and gender. Organizational studies [16, 42] find that increased group diversity can lead to greater coverage of information and improved decision making. However, diversity can also lead to increased discord and longer decision times.

Various measurements of diversity may be applied to suggest a set of group members to gain adequate coverage for an analysis task. Such measurements might come from analyzing differences between user profiles and structural features of participants' social networks [7]. Such networks may be explicitly articulated or inferred from communication patterns, such as the co-occurrence of commenters across discussion threads. Wu et al's [52] study of organizational information flow found that information spreads efficiently among homophilous group members but not across community boundaries, further suggesting the value of identifying structural holes and directing bridging individuals in the social network towards particular findings.

8 Consensus and Decision Making

The need to establish group consensus arises in many phases of the sensemaking cycle. Examples include agreement about the data to collect, how to organize and interpret data, and making decisions based upon the data. Consensus may arise through discussion or may involve the aggregation of individual decisions.

8.1 Consensus and Discussion

Mohammed [34] combines various contributions in social psychology and organizational studies to posit a model for cognitive consensus in group-decision making. This model takes into account the assumptions, category labels, content domains, and causal models possessed by each participant, and how they can evolve through discussion. One tangible recommendation that comes from this work is to clearly identify the points of dissent, creating focal points for further discussion and negotiation. From a design perspective, this suggests the need for communication mechanisms that allow such items to be labeled and addressed. Collaborative tagging [24] is one potential candidate.

Scheff [41] notes that consensus requires more than participants simply sharing a belief; participants must believe that their beliefs are the same, and achieve realization that others understand one's position. This implies the need for feedback loops for gauging mutual understanding. Along these lines, it may be useful to consider the effects of multiple communication channels on decision processes. Collaborative visualization environments that provide messaging, in either synchronous or asynchronous forms, might provide backchannels for negotiation and non-public discussion. The integration of instant messaging into the GMail e-mail service provides an example of how different communication channels can be weaved together in a single system.

The value of different forms of consensus can vary based on the task at hand. Hastie [26] found that group discussion improved accuracy when decision tasks had demonstrably correct solutions, allowing groups to evaluate their output. When task outcomes are open-ended, consensus through discussion is harder to evaluate. In a simulated graduate admissions task, Gigone and Hastie [22] found little value in discussion, as group decisions were well-matched by simply averaging prior individual decisions.

One design implication that again arises is the use of *voting* or *ranking* systems. Mechanisms for expressing support or disdain for hypotheses could aid data interpretation and further identify controversial points. For example, Wikimapia users can vote on the accuracy of labeled geographic regions and Swivel supports ratings of interestingness. A game-like variation on this approach is the creation of *prediction markets* [45]: individuals can be given a limited amount of "points" or "currency" that they can use to vote for hypotheses they find most promising. Hypotheses that are later validated could reap payback rewards for their supporters.

Design Consideration	Description	Sections
Modularity and Granularity	Identify appropriately-scoped units of work that form basic analytic contributions.	§2
Cost of Integration	Synthesize work while attempting to lower integration costs and maintain quality.	§2
Shared Artifacts	Structure collaboration through shared, editable representations.	§2, 3
Artifact Histories	Provide histories of actions performed on artifacts.	§3
View Sharing / Bookmarking	Enable lightweight sharing of views across media with bookmarks.	§3
Content Export	Support embedding of annotated views in external media (e.g., email, blogs, reports)	§3
Discussion	Support commentary; consider implications of discussion model on common ground.	§3, 8
Notification	Support notification subscriptions for views, artifacts, people, and groups.	§3
Action Flags	Mark needed future actions: unanswered questions, need for evidence, etc.	§3, 8
Social Navigation	Make activity patterns visible, determine popular and neglected data regions.	§2, 3
Recommendation	Suggest related views, comments, and data to current points of interest.	§3
Pointing Techniques	Support nuanced pointing through selection techniques and visual effects.	§4
Personal Relevance	Increase engagement by increasing personal relevance of data sets.	§5
Social-Psychological Incentives	Increase engagement by surfacing individual contributions.	§5
Game Play	Game design elements can provide incentives and be used to direct effort.	§2, 5
Identity Markers	Enable identification of collaborators in a contextually-appropriate manner.	§6
User Profiles	Support awareness of others' backgrounds and skills.	§6
Activity Histories	Personal action histories allow past contributions to be assessed.	§3, 6
Activity Summaries	Activity indicators or summaries aid reputation and visibility of contributions.	§3, 5, 6
Group Management	Group creation and management mechanisms address issues of scale and privacy.	§7
Group Size	Optimal group size determination can improve efficiency of analysis.	§2, 7
Group Diversity	Appropriate within-group diversity can result in more complete results.	§2, 7
Voting and Ranking	Quantitative measures can be used for consensus and to lower integration costs.	§2, 6, 8
Presentation	Support creation and export of presentations for telling analysis stories.	§3, 4, 8

Table 1. Selected Design Considerations for Collaborative Visual Analytics. The table lists many of the individual design considerations visited in this article, providing a brief description and noting the relevant sections that discuss the issue in detail.

8.2 Information Distribution and Presentation

An important dimension of group consensus is the distribution of information across group members. Both Stasser [44] and Gigone and Hastie [22] find that groups generally do not successfully pool information, biasing decision-making in the direction of the initial information distribution. This may be due to the inertia of individual decisions made prior to discussion or due to already-shared information providing common ground for discussion, biasing conversation against the unshared information. A potential benefit of collaborative analysis is better information pooling, providing a record of findings and opinions that can be surveyed prior to decision-making and discussion. Improving collective information foraging may help inform group decision-making by changing the information distribution.

Common forms of information exchange in group sensemaking are reports and presentations. Narrative presentation of analysis "stories" is a natural and often effective way to communicate analytic findings, and has been observed as a primary use of Spotfire's Decision Site Posters [43]. The challenge to collaborative visualization is to provide mechanisms to aid the creation and distribution of presentations. For example, the sense.us system [29] allows users to construct and share trails of related views to create tours spanning multiple visualizations. This approach could be further improved with support to build presentations semi-automatically using interaction histories, export such presentations into external media, and apply previously discussed pointing techniques. Bookmarking can also enable recipients of a presentation to backtrack to the original visualization to conduct more analysis or verification.

9 Conclusion And Future Directions

This article presents design considerations for collaborative visual analytics, attempting to identify accomplishments which facilitate collaboration and suggest mechanisms for achieving them. Highlights include a list of collaborative visualization tasks, techniques to improve shared context and awareness, and suggestions for increasing engagement and allocating effort. Many of these considerations are summarized in Table 1. The overarching goal is to effectively parallelize work, facilitate mutual understanding, and reduce the costs of collaborative tasks.

Visiting these considerations also provides an agenda for future research in collaborative visual analytics, surfacing hypotheses in need of study and suggesting new technical mechanisms:

- What is the effect of different discussion models (e.g., independent, embedded, doubly-linked) on participation and the establishment of common ground?
- Beyond textual discussion, what external representations will effectively support collaborative analysis? How do such artifacts affect grounding and the cost of integration?
- How can the synthesis of individual contributions be improved? Can (semi-)automatic summarization or merging of separately developed data views (e.g., [5]) be used to form aggregated contributions?
- How should selection and visual emphasis techniques be designed to provide nuanced pointing behaviors? Can referenced objects be unambiguously recognized by both human and machine collaborators?
- How can pointing and graphical annotation gracefully handle dynamic visualizations and changing data sets?
- How should social navigation cues be effectively added to visual analysis tools to unobtrusively improve awareness?
- Can automated techniques be used to help allocate effort? For example, mining past contributions, user profiles, and inferred social networks may enable systems to direct collaborators to tasks in need of attention.
- How can the fruits of collaborative visual analysis be more effectively exported, shared and embedded in external media such as web pages, e-mail, and presentations?

These and other challenges present exciting opportunities for advancing visual analytics research.

ACKNOWLEDGEMENTS

We thank Martin Wattenberg, Fernanda Viégas, Stu Card, Coye Cheshire, and Andrew Fiore for their insightful comments.

REFERENCES

[1] Benbunan-Fich, R., Hiltz, S.R., Turoff, M. (2003). A comparative content analysis of face-to-face vs. asynchronous group decision making. *Decision Support Systems* archive, 34(4), 457-469.

[2] Benkler, Y. (2002). Coase's Penguin, or, Linux and the Nature of the Firm. *Yale Law Journal*, 112(369).

[3] Billman, D., Convertino, G., Pirolli, P., Massar, J. P., Shrager, J. (2006). Collaborative intelligence analysis with CACHE: bias reduction and information coverage. PARC UIR Tech Report.

[4] Brennan, S.E. (2005). How conversation is shaped by visual and spoken evidence. In *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Trueswell and Tanenhaus (eds.), MIT Press, Cambridge, MA, 95-129.

[5] Brennan, S.E., Mueller, K., Zelinsky, G., Ramakrishnan, I.V., Warren, D.S., Kaufman, A. (2006). Toward a Multi-Analyst, Collaborative Framework for Visual Analytics. *Proc. IEEE VAST 2006*.

[6] Brush, A.J., Bargeron, D., Grudin, J., Gupta, A. (2002). Notification for shared annotation of digital documents. *Proc. ACM CHI 2002*.

[7] Burt, R. S. (2004). *Structural holes and good ideas*. American Journal of Sociology, 110(2), 349–399.

[8] Caillois, R. (1961). *Man, Play, and Games*. Free Press of Glencoe.

[9] Card, S.K., Mackinlay, J.D., Shneiderman, B. (1999). *Readings in Information Visualization: Using Vision To Think*. Morgan-Kaufmann.

[10] Carroll, J., Rosson, M.B., Convertino, G., Ganoe, C.H. (2005). Awareness and teamwork in computer-supported collaborations. *Interacting with Computers*, 18(1):21-46.

[11] Cheshire, C. (2006, under review.) Social Psychological Selective Incentives and the Emergence of Generalized Information Exchange.

[12] Clark, H. H. (2003). Pointing and placing. In S. Kita (Ed.), *Pointing. Where language, culture, and cognition meet* (pp. 243-268). Erlbaum.

[13] Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1-39.

[14] Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior* , 22, 245- 258.

[15] Clark, H.H. and Brennan, S.E. (1991). Grounding in Communication. In Perspectives on socially shared cognition, L. B. Resnick, R. M. Levine, and S. D. Teasley, eds. *American Psychological Association*, p. 127–149.

[16] Cummings, J. (2004). Work groups, structural diversity, and knowledge sharing in a global organization. *Management Science*, 50(3), 352–364.

[17] Donath, J.S. (1998). Identity and Deception in the Virtual Community. In M. Smith and P. Kollock (eds.) *Communities in Cyberspace*. Routledge.

[18] Dorling, D., Barford, A., Newman, A. (2006). Worldmapper: The World as You've Never Seen It Before. *IEEE Transactions on Visualization and Computer Graphics.* 12(5). Sep/Oct 2006.

[19] Dourish, P., Belotti, V. (1992). Awareness and coordination in shared workspaces. *Proc. ACM CSCW 1992*. Toronto, Ontario. p. 107-114.

[20] Dourish, P., Chalmers, M. (1994). Running Out of Space: Models of Information Navigation. *Proc. Human Computer Interaction (HCI'94)*.

[21] Gergle, D., Kraut, R.E., Fussell, S.R. (2004). Language efficiency and visual technology: Minimizing collaborative effort with visual information. *Journal of Language & Social Psychology*, 23(4):491-517.

[22] Gigone, D., Hastie, R. (1993). The Common Knowledge Effect: Information Sharing and Group Judgment. *Journal of Personality and Social Psychology*. Vol. 65, pp. 959-974.

[23] Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Anchor Books, New York, NY.

[24] Golder, S.A., Huberman, B.A. (2006). The Structure of Collaborative Tagging Systems. *Journal of Information Science* 32(2). April 2006.

[25] Hardin, R. (2003). The Free Rider Problem. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2003 Edition).

[26] Hastie, R. (1986). Experimental Evidence on Group Accuracy. In B. Grofman & G. Owen (eds.), *Decision Research*, 2:129-157. JAI Press.

[27] Heer, J. (2006). Socializing Visualization. In *CHI 2006 Workshop on Social Visualization*.

[28] Heer, J., Agrawala, M. (2006). Software Design Patterns for Information Visualization. *IEEE Transactions on Visualization and Computer Graphics.* 12(5). Sep/Oct 2006.

[29] Heer, J., Viégas, F., Wattenberg, M. (2007). Voyagers and Voyeurs: Supporting Asynchronous Collaborative Information Visualization. *Proc. ACM CHI 2007*. San Jose, CA.

[30] Hill, W. C., Hollan, J. D., Wroblewski, D., McCandless, T. (1992). Edit wear and read wear. *Proc. ACM CHI 1992*. pp. 3–9.

[31] Hill, W.C., Hollan, J.D. (1991). Deixis and the Future of Visualization Excellence. *Proc. of IEEE Visualization*. pp. 314–319.

[32] Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Cosley, D., Frankowski, D., Terveen, L., Rashid, A. M., Resnick, P., and Kraut, R. (2005). Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10(4).

[33] Many Eyes. (2007). http://many-eyes.com

[34] Mohammed, S. (2001). Toward an Understanding of Cognitive Consensus in a Group Decision-Making Context. *The Journal of Applied Behavioral Science*, Vol. 37, No. 4, (Dec., 2001), pp. 408-425.

[35] Pirolli, P. (2006, in progress). Social Information Foraging. Chapter 18, *Information Foraging*.

[36] Pirolli, P., Card, S. K. (1999). Information Foraging. *Psychological Review* 106(4): 643-675.

[37] Pirolli, P., Card, S.K. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. *Proc. of International Conference on Intelligence Analysis*.

[38] Resnick, P., Zeckhauser, R., Swanson, J., Lockwood, K. (2006). The Value of Reputation on eBay: A Controlled Experiment. *Experimental Economics*, 9(2):79-101.

[39] Russell, D.M., Stefik, M.J., Pirolli, P., Card, S.K. (1993). The Cost Structure of Sensemaking. *Proc. ACM CHI 1993*. Amsterdam, NL.

[40] Salen, K., Zimmerman, E. (2003). *Rules of Play: Fundamentals of Game Design*. MIT Press.

[41] Scheff, T.J. (1967). Toward a Sociological Model of Consensus. *American Sociological Review*, 32(1). (Feb., 1967), pp. 32-46.

[42] Schultz-Hart, S., Frey, D., Lüthgens, C. & Moscovici, S. (2000). Biased Information Search in Group Decision Making. *Journal of Personality and Social Psychology*, 78(4), 655-669.

[43] Spotfire (2006). http://spotfire.com/products/decisionsite_posters.cfm

[44] Stasser, G., Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, 57, 67-78.

[45] Surowiecki, J. (2004). *The Wisdom of Crowds*. Random House.

[46] Swivel. (2007). http://www.swivel.com

[47] Thomas, J.J. and Cook, K.A., eds. (2005). Illuminating the Path: The Research and Development Agenda for Visual Analytics. IEEE Press.

[48] Viégas, F.B. and Wattenberg, M. (2006). Communication-Minded Visualization: A Call to Action. *IBM Systems Journal*. 45(4).

[49] Wattenberg, M., Kriss, J. (2006). Designing for Social Data Analysis. *IEEE Trans. on Visualization and Computer Graphics*. 12(4): 549–557.

[50] Wikimapia. (2007). http://wikimapia.org

[51] Wright, W., Schroh, D., Proulx, P., Skaburskis, A., Cort, B. (2006). The sandbox for analysis: concepts and evaluation. *Proc. ACM CHI 2006*.

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

[53] Zhang, J., Norman, D.A. (1994). Representations in Distributed Cognitive Tasks. *Cognitive Science*, 18(1):87-122.