Collaborative Synthesis of Visual Analytic Results

Anthony C. Robinson

GeoVISTA Center, Department of Geography, The Pennsylvania State University

ABSTRACT

Visual analytic tools allow analysts to generate large collections of useful analytical results. We anticipate that analysts in most real world situations will draw from these collections when working together to solve complicated problems. This indicates a need to understand how users synthesize multiple collections of results. This paper reports the results of collaborative synthesis experiments conducted with expert geographers and disease biologists. Ten participants were worked in pairs to complete a simulated real-world synthesis task using artifacts printed on cards on a large, paper-covered workspace. Experiment results indicate that groups use a number of different approaches to collaborative synthesis, and that they employ a variety of organizational metaphors to structure their information. It is further evident that establishing common ground and role assignment are critical aspects of collaborative synthesis. We conclude with a set of general design guidelines for collaborative synthesis support tools.

KEYWORDS: Synthesis, collaboration, visual analytics, user-centered design.

INDEX TERMS: H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces

1 INTRODUCTION

Visual analytic tools promise to supply analysts with the means necessary to tackle complex and dynamic problems. Most of the time we can expect analysts to work in teams, blending together expertise from a variety of related domains to address a multifaceted problem. Supporting collaboration among analysts requires attention not only to direct analytical support, but also to the challenge of organizing and making sense out of collections of analytical results. This paper reports on experimental results with experts from geography and disease biology to explore this latter challenge.

The central interest of this research is to characterize and design for collaborative result synthesis. Synthesis can be defined in a variety of ways, but for the purpose of this research we define it as *the stage of an analytic process in which analysts organize and combine individual analytical results into coherent groups that are used to assign meaning and/or encapsulate complex ideas[1].*

Our focus on synthesis builds upon prior work in geographic visualization, which describes a research process that starts with exploration, moves to confirmation, transitions to synthesis, and ends with presentation [2, 3]. This work also complements current research in visual analytics focused on supporting collaboration. In terms of the sense making process developed by Pirolli and Card [4], this work focuses on characterizing how users

Email: arobinson@psu.edu

schematize and hypothesize about collections of information. This work also answers a call by Heer and Agrawala for research to explore how users merge their work into collaborative collections [5]. Prior work in this area includes research by Isenberg and Carpendale that characterized how users work with interactive visualization artifacts on collaborative tabletop displays [6]. In a more recent study, Isenberg et al. explore the analytical processes that users employ when collaborating with a collection of visualization artifacts printed on paper [7].

Tackling the problems posed by synthesis is a priority because it is important to shape the direction of new synthesis tools that are currently in development. Examples of tools that are currently being used to collect, organize, and add meaning to analysis artifacts include Oculus Info nSpace [8] and i2 Analyst's Notebook [9]. Developers of these tools and others like them stand to benefit from research that explores the process of synthesis to suggest designs for future synthesis support tools.

2 SYNTHESIS EXPERIMENTS

To observe how synthesis takes place an experiment was designed for participants to simulate the real-world task of determining the source of an avian influenza outbreak in the Pacific Northwest. The following sections describe how the experiment was designed and how its results were analyzed.

2.1 Study Participants

Five geography experts from the Penn State GeoVISTA Center and five infectious disease experts from the Penn State Center for Infectious Disease Dynamics (CIDD) were recruited to take part in individual and collaborative synthesis experiments. These experts are postdoctoral research associates and senior PhD candidates in their respective laboratories. They were recruited with email solicitations and were provided with a $50 stipend.

The strategy guiding participant selection was to explore synthesis as conducted by those who might become analysts in the future. It is important to know not only what is needed to support current analysts, but also to know how synthesis support tools should function once people currently receiving education and training enter the analyst workforce in years to come. Mixing participants from different backgrounds was also done to ensure that the collaborative synthesis setting accurately reflected how collaboration was likely to occur in real-world situations, where analysts from a variety of backgrounds will work together on a multi-faceted problem.

The work reported here is part of a larger project that also involves 8 analysts at Pacific Northwest National Laboratory (PNNL). Analysts at PNNL completed individual synthesis experiments in July of 2007. Results from this work and individual PSU experiments are currently in preparation.

2.2 Experiment Details

Participants in this study completed two synthesis experiments in October of 2007. First, they completed an individual experiment in which they had to develop hypotheses for the source of the outbreak using a set of analytical artifacts. Second,

they worked in pairs to rank their hypotheses in a collaborative session. The focus of this paper is on the latter experiment. Details about how the individual experiments were conducted are provided in the following section to provide necessary context for understanding the collaborative experiments.

2.2.1   Individual Synthesis Experiments

The basic experiment design features a synthesis activity in which participants organize and annotate a set of physical artifacts (3.5" x 5" laminated cards) on a 36" x 36" paper-covered workspace. Participants were also provided with markers, pens, adhesive tags, and post-it notes of multiple sizes and colors to modify the workspace as desired.

The use of physical artifacts and tools was intended to explore how synthesis occurs without the constraints imposed by current software tools, which typically constrain the types of organizational metaphors one can use. Sellen and Harper [10] suggest that real materials can be useful for eliciting interface design guidelines by revealing how participants make use of their affordances to complete tasks. Recent work in visual analytics by Isenberg et al. [7] uses a paper-based experimental approach to explore and characterize how participants analyze visualizations.

Participants were instructed that an avian influenza outbreak had occurred in the Pacific Northwest and that their task was to develop hypotheses for the source of the outbreak using the artifacts and tools they had been provided. During the experiment, participants were asked to provide a verbal protocol to state what they were doing. In addition, participants were instructed to inform the lead investigator whenever they had an emergent hypothesis, and to then briefly describe this hypothesis so that it could be recorded. At the conclusion of each experiment, the participant was asked to describe how they had organized their information and what they had done to indicate their hypotheses on the workspace.

2.2.2   Collaborative Synthesis Experiments

Five collaborative experiments (referred to as PSU 1C, 2C, etc.) were conducted in October of 2007 with pairs of analysts from PSU. One analyst in the pair was an expert geographer from the GeoVISTA Center, while the other was an expert disease biologist from CIDD. After completing individual experiments in separate sessions, they worked together with a new 36" x 36" workspace situated between their individual workspaces to develop a ranked set of plausible hypotheses to provide to a decision maker. All three experiments were conducted back to back on the same day. It was not possible to conduct parallel experiment sessions, so the GeoVISTA participant completed their individual experiment first, the CIDD participant went next while the GeoVISTA participant waited outside, and finally both worked together on the collaborative task.

In the preceding individual experiments, the artifact sets were selected so that each participant in a paired group had 36 artifacts, including information that the other was missing – creating a situation where participants had partially overlapping information, a condition likely to occur in a real-world analysis situation. 24 of the artifacts were common to both participants, and each was given 12 that were unique. For the collaborative experiments a duplicate set of artifacts was provided for use with the blank collaborative workspace.

Each group was given one hour to complete the collaborative synthesis task. A video camera recorded the experiment to allow for later analysis of actions and verbal reports. A short debriefing session was conducted at the end of the session to have participants describe how they had organized their information and ranked hypotheses.

2.3   Analysis Artifacts

We developed a set of 48 analytical result artifacts (Figure 1) for use in synthesis experiments. The design of these artifacts was based in part on the *Sign Of The Crescent* training activity data. The *Sign Of The Crescent* training activity [11] was developed by the Joint Military Intelligence College. It contains a set of text reports that include source information and time stamps. The goal of the activity is for analysts to determine the most likely hypothetical outcomes of a terrorism scenario based on the information they can gather from these reports. The *Sign Of The Crescent* is commonly used as a training exercise for intelligence and crime analysts [12]. Because there is little guidance for the development of analytical artifacts for synthesis experiments, this activity was used as a model to suggest the types and amounts of information that should be included in each artifact.
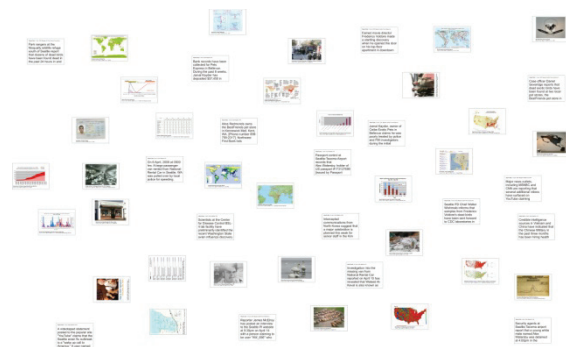

Figure 1. Synthesis artifacts developed for experiments

The artifacts used in the our synthesis experiments differ from those in the *Sign Of The Crescent* activity in that they include maps, photos, and other graphics as well as text reports. This was done to more accurately simulate the many types of analytical artifacts that analysts are likely to generate using visual analytic tools. A 2-to-1 ratio of graphic artifacts to text artifacts was used to develop the complete set. This choice reflects our assumption that visual analytic tools typically generate graphical rather than textual results. The complete set of artifacts is provided in high-resolution format in Supplement A.

2.3.1   Embedded Hypotheses

The artifacts were designed to weave a multi-threaded story regarding an avian influenza outbreak in the Pacific Northwest. Each artifact features source information and a timestamp, and there are photographs, video captures, maps, data graphics, and text reports included in the set. The content and source information for these materials was designed to impart varying levels of credibility to the evidence to simulate real-world information complexity. Based on the information provided by the artifacts, there are at least five potential sources for the outbreak, and there are many more permutations possible given combinations of those potential sources. The five threads devised for the experiment include: a natural occurrence based on bird migration, a person named Alex Watersby who intentionally spread the flu to wild birds and through pet stores, an Al-Qaeda operative named Waleed Al-Keval who infected wild birds, an unintentional outbreak caused by illegal pet trade activity by local pet stores, and a plot by North Korea and China to spread avian influenza to disrupt poultry commerce in America.

2.3.2 Randomization

The artifacts were assigned a number from 1 to 48, then they were sorted for each of the five collaborative groups using a random permutation provided by the website www.randomization.com. This was done in order to avoid potential order effects as much as possible. The randomly sorted artifacts were then split into overlapping groups of 36, so that each participant had 12 unique and 24 common artifacts. The randomized artifacts were placed on the workspace in a stack at the beginning of each experiment.

2.4   Coding and Analysis

The collaborative synthesis experiments generated five video recordings. The goal was to extract useful information from these recordings in a systematic manner that would enable the development of specific software design guidelines. To that end, a coding scheme was developed to describe the low-level events that users initiated to complete the synthesis experiment. The focus here was on defining which software tools and functions would be necessary to support what the user was doing with the artifacts during the experiment. The decision was made that coding would apply only to actions that were separable and obvious, using the verbal protocol of the user as well as the context of the action (for example, the work immediately prior to the action in question) to help guide choices. The problem of deciding what is a separable action is this case is non-trivial, and a conservative approach was taken here to identify actions that involved the actual use of artifacts, tools, or the workspace. More subtle actions like gestures or verbal declarations were not set aside for coding in this study.

The first coding scheme was developed by the lead author after watching sample videos and noting what software tools would be required to support what participants were doing with artifacts, tools, and the workspace to complete the synthesis task. This scheme was then further refined by a group of seven interface and software developers from the GeoVISTA Center during a meeting in which the experiment and a sample of the videos were reviewed. A final round of refinement then took place after the lead author completed sample coding of two of the five videos. The final coding scheme is presented here, and definitions for each code can be found in Supplement B:

- Annotate (text)
- Annotate (drawing)
- Group Artifacts (hypothesis)
- Group Artifacts (category)
- Group Artifacts (type)
- Group Artifacts (time)
- Group Artifacts (read or un-read)
- Group Artifacts (unknown)
- Collapse Group of Artifacts
- Expand Group of Artifacts
- Link Artifacts (hypothesis)
- Link Artifacts (network)
- Sort (category)
- Sort (type)
- Sort (time)
- Search (category)
- Search (time)
- Search (for read or un-read artifacts)
- Search (keyword)
- Tag (hypothesis)

- Tag (category)
- Tag (time)
- Tag (network)
- Tag (certainty)
- Tag (follow-up)
- Tag (place)
- Zoom (single item)
- Zoom (group of items)

Videos were coded using Transana [13], a software tool designed for qualitative analysis of audio and video data. In Transana, timestamps can be added to transcripts that serve as live links back to that moment of video or audio. During coding, events were timestamped, and then assigned a letter and number code from the coding scheme. This allowed the time of the event to be recorded along with its description, and these timestamps were used later during code reliability evaluation.

The larger research project that this paper is part of includes individual synthesis experiments with 8 analysts at PNNL, generating eight videos. Ten videos resulted from the individual portion of the PSU experiments, and five from the collaborative portion of the PSU work (results from these five are the focus of this paper). Thus, we coded a total of 23 videos.

A common method of assessing the credibility of coded data is to evaluate code reliability [14]. Ten sample videos (4 from PNNL, 4 from individual PSU experiments, and 2 from collaborative PSU experiments) were selected to test coding reliability. Project funding did not allow for a full recode. Two graduate student coders were recruited from the GeoVISTA Center who had not participated in the experiment. Coders were each given five videos that included event markers but were stripped of the event code that the lead author had assigned.

Code reliability was assessed using a percent agreement measure, a commonly used method of inter-rater reliability [14]. Percent agreement simply measures the percent of coded entries that match between coders. It is an imperfect measure, as it does not account for the chance that a coder will guess when assigning a code and be correct. However, in this case there are 28 codes, so the chance of guessing a matching code at random is small. The percent agreement between the independent coders and the lead author was 88.4%.

Coded results were converted into Microsoft Excel tables that were then manipulated using Tableau desktop visualization software [15] to create time series graphs for each user and summary graphs for each user group. These materials were further edited using CorelDRAW to enhance readability, to assign an appropriate color scheme, and to develop layouts for presentation.

3   SYNTHESIS EXPERIMENT RESULTS

The following sections describe cumulative coding results, results for particular pairs of participants, common patterns of activity across multiple videos, and the organizational metaphors that were used by participants. Where relevant, differences between groups from GeoVISTA and CIDD are highlighted.

The graphical results presented here can be interpreted using the legend provided in Figure 2. Choosing colors to indicate 28 different codes required the use of two glyphs for each coded event. The primary glyph shows the color of the major category it is associated with (Annotate, Group, Sort, etc...), and the smaller glyph attached to its bottom indicates which subtype of that category was assigned. A qualitative color scheme was used from www.colorbrewer.org, and a grayscale ramp was used to fill the subtype glyphs. Although the codes are qualitatively different,

visual clarity was not attainable using qualitative schemes for both higher level categories as well as subtypes.

Figure 2. Color legend for coded experimental results

3.1 Cumulative Results

Cumulative summaries of the coded events (Figure 3) from collaborative synthesis experiments provide several important insights. If all coded events are summarized irrespective of who initiated them, the five most common events are zoom (single item), annotate (text), group (category), group (timeline), and tag (hypothesis). When code results are broken down according to who initiated them, GeoVISTA and CIDD participants differ in a few ways. GeoVISTA participants annotate more frequently, and the 3rd most common event for GeoVISTA participants is tagging hypotheses – a code that appears substantially lower down the list for CIDD participants. A small number events were coded as collaborative actions, where both participants worked together to complete an action. The most common collaborative actions are search (read/unread), and zoom (single).
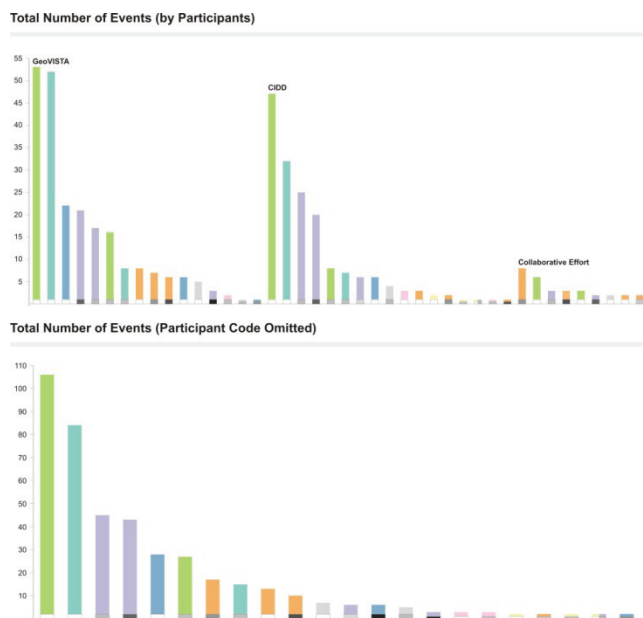

Figure 3. Cumulative results from collaborative experiments

3.1.1 Zooming

It comes as no surprise that the most common coded event from collaborative synthesis experiments is zooming a single item. The collaborative experiment followed shorter individual experiments where participants worked independently with an incomplete set of artifacts to develop hypotheses. Instructions for the collaborative experiment asked participants to rank the hypotheses they had found independently. Therefore the collaborative setting encouraged participants to revisit their work and to explore their colleagues' work for additional evidence to support or refute their hypotheses – requiring individual artifacts to be closely examined.

3.1.2 Annotation

Annotation in the form of text on the workspace was another common collaborative synthesis event. In many cases one of the

participants took on the role of being the "reporter" and would write down hypotheses, key facts, and other information while the session progressed. The coding results indicate that GeoVISTA participants were responsible for more text annotations than CIDD participants, although that does not necessarily indicate differences in quality or relevance to the task. Less frequent were graphical annotations. These were typically in the form of regions drawn around groups or arrows drawn between artifacts.

3.1.3 Grouping

As participants continued to refine and develop hypotheses, a substantial amount of artifact grouping took place - usually in the form of category groups or timeline groups. Typically participants would develop these new groups with parts of existing groups from their workspaces that had been organized prior to the collaborative experiment. These results indicate that grouping is not only important as an initial stage of synthesis when artifacts are first evaluated, but that groupings will evolve over time in collaborative settings.

3.1.4 Tagging

Of the most frequent coded events, tagging was much more common by GeoVISTA participants than CIDD participants. Three groups; PSU 1C, 3C, and 4C used tagging on workspaces and artifacts. GeoVISTA participants primarily used tags to identify hypotheses. These tags were typically used in two ways. Large post-it notes were annotated with short descriptions of hypotheses. Small post-its or arrow tags were placed on specific artifacts to indicate their presence in particular hypotheses.

3.1.5 Collaborative Events

Although a relatively small portion of events were identified as purely collaborative (meaning both participants completed an event together), the character of these actions is noteworthy. Participants would often choose to work together to search for artifacts. Most of the time these searches sought to identify artifacts that they had in common versus those that only one of them had. Searches for artifacts related to particular categories or keywords were also conducted as hypotheses were refined. Usually one participant would recall seeing artifacts from a category or containing a specific keyword, and then both participants would work together to try and find those artifacts.

3.2 Individual Session Results

Cumulative coding results outlined in the previous section describe general findings based on the relative frequency of events. This section focuses on some of the unique strategies that participant groups used to conduct collaborative synthesis. It also describes the organizational methods that participants used to complete the task of ranking hypotheses.

Coded results for the collaborative synthesis experiments are provided in Figure 5. Each session is represented with a three-part track, split according to who initiated events. Events in the A and B tracks refer to GeoVISTA and CIDD participants respectively, and C events are those that both participants conducted together. The legend in Figure 2 applies to these results as well. Supplement C provides detailed results for each pair of participants, where each set of coding results is shown with photographs of the final workspaces, a graph that summarizes events, and a short written summary.

3.2.1 Establishing Common Ground

Collaboration often involves an initial period in which collaborators must develop an understanding of each other's work

[16]. Participants came into the collaborative experiment having just completed their own individual work with the artifacts. They knew they would be participating in a collaborative activity, but they were not told that they would be using their individual results, so participants found out at the beginning of the experiment when the instructions were explained that they would need to evaluate each other's previous findings. Participants were not instructed to establish common ground in a particular way, but each group chose to do so in the same way. After the start of the session, the GeoVISTA participant would begin explaining how they had organized their workspace. After they were finished, the CIDD participant would do the same. As shown in Figure 5, groups differed in terms of how much time was spent on this activity. PSU 2C and 5C took only a few minutes to establish common ground, while PSU 3C took the longest with almost fifteen minutes of explanation. During these explanations there were occasional zooms and other actions as particular artifacts were pointed out by one participant to the other.

A key aspect of establishing common ground for participants in these experiments was to assess which information they had seen before and which they had not. Participants quickly discovered during their personal workspace explanations that they had partially overlapping information. Groups varied from one to the next in terms of how systematically they identified the pieces they had in common. All groups discussed this issue, but only PSU 2C and 3C conducted sustained searches for uncommon artifacts.

3.2.2 Collaborative Synthesis Strategies

The coded results for each group are largely unique from one another. It is difficult to discern patterns in the sequence of events that may be considered a clear strategy. The most obvious pattern is evident when examining which participant initiated which sorts of actions. This can be seen by looking at the cumulative graphs of coding results for each participant group (see Supplement C). Participants in PSU 1C, 2C, and 4C chose roles to complete the experiment, assigning one member the responsibility of making annotations (Figure 4). Role assignment has been noted in previous research as an important aid to collaboration [17]. In two instances, the GeoVISTA participant was in charge of annotation, while in the remaining instance the CIDD participant took that role.
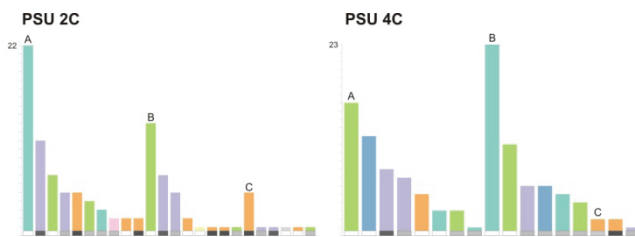

Figure 4. Evidence of role assignment in collaborative synthesis

Code results for PSU 5C show a similar pattern with different types of events figuring more heavily for each participant, but in this case it was clear from video evidence that they worked separately as a result of conflict rather than collaboration.

The strategies that groups pursued to complete the collaborative task differed in terms of how they were begun. There were three observed strategies for approaching the task: one group chose to address the timeline of events first, two groups chose to identify common and uncommon information first, and two groups decided to lay out their hypotheses first. See Figure 5 in reference to the remainder of this section.

Participants in PSU 1C approached the collaborative task by devoting particular attention to the temporal dimension of the evidence. After establishing common ground, participant B began constructing a timeline while participant A searched for artifacts based on time of occurrence. They situated their timeline around the idea that hypothesis plausibility depended on whether or not the events made sense compared to the time it takes for H5N1 to build up in a host and begin shedding (at which point it becomes contagious to others).

The PSU 2C and 3C groups approached the task by first determining which information they shared and which they did not. After establishing common ground, participants in PSU 2C went to each other's original workspace to identify which artifacts they had not seen before. From there, they focused their attention on modifying participant B's workspace to complete the task.

PSU 4C and 5C focused on hypotheses first. After explaining their prior work, participants in PSU 4C began by writing hypotheses on post-it notes and placing them on the collaborative workspace. From there, they constructed hypothesis groups and timelines to refine their work. Participants in PSU 5C also began by focusing on their hypotheses, choosing to discuss them without manipulating any artifacts on the workspace.

3.2.3 Ranking Hypotheses

The task for participants was to rank the hypotheses they had developed according to which were the most plausible, assuming they would later communicate their results to decision makers. Participants were not instructed to use any particular type of ranking, as we wanted to see what groups would do unguided.
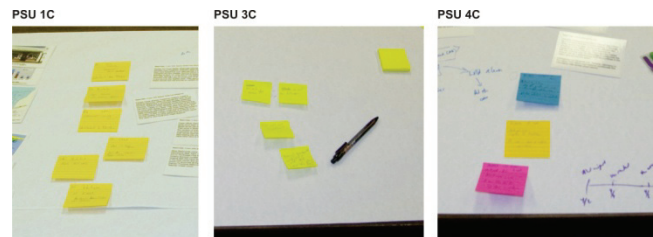

Figure 5. Post-its used to record hypotheses

Groups chose two ways to rank their final hypotheses. Three groups: PSU 1C, 3C, and 4C used post-it notes to record hypotheses (Figure 6). Then, each participant had an opportunity to rank them according to preference. Those groups then finished by negotiating a final ranking. Participants stated that this method was particularly effective because the post-it notes could be moved around easily to develop the final ranking. Two groups: PSU 3C and 5C, did not record a set of final ranked hypotheses on the workspace. Their ranking was based only on discussion.

3.2.4 Collaborative Deadlock

One group's results stands out as an outlier. PSU 5C initiated the fewest events out of all of the groups, and failed to develop a collaboratively-ranked set of hypotheses. The participants in this group did not agree with each other about which information was most important and which hypotheses were most plausible. They spent the majority of the session period discussing their differences. A risk associated with conducting any type of collaborative research is that the participants may not work together well, as happened in this case. In the last twenty minutes of their session, the participants of PSU 5C began working separately from one another to try and satisfy the requirement of the task to develop a ranked set of hypotheses.
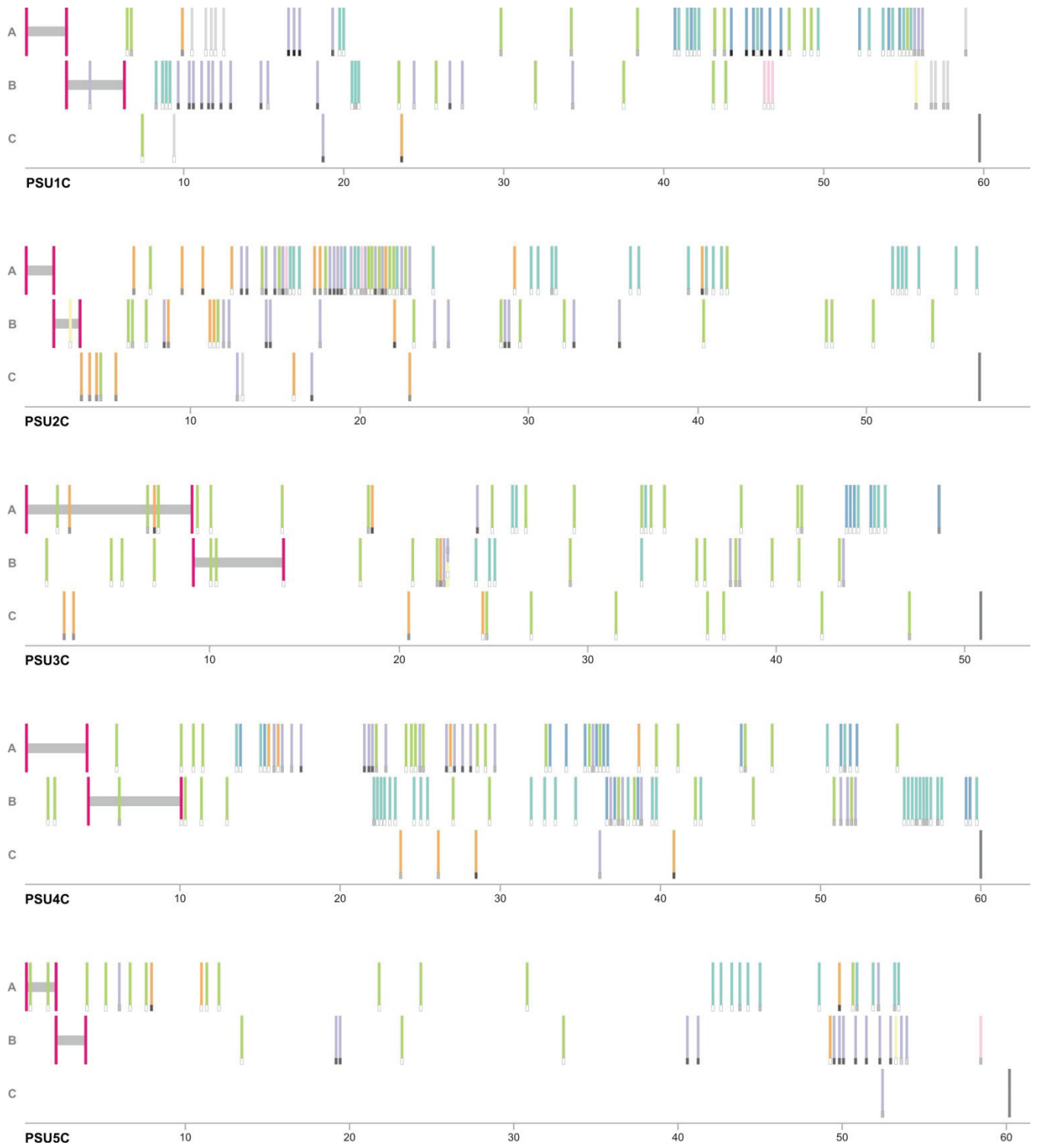
Figure 6. Detailed experiment results. Prefix A indicates actions by the GeoVISTA participant, B indicates actions by the CIDD participant, and C indicates collaborative actions. A gray bar with red ends is used to highlight the time participants spent explaining their prior work.

They did not discuss this as a solution to their conflict; rather the transition to this behavior was unstated. Participant A worked on a concept-map style of organization for the primary hypotheses they felt was most likely, while Participant B created groups on another portion of the workspace separate from A's work (Figure 7). Following an extended period of independent work, the participants then talked about the relative merits of their respective hypotheses, but did not reach a consensus about a ranking. The root of their disagreements appeared to be A's insistence on a single hypothesis that involved assumptions about data that was not explicitly present in the artifacts. B was not willing to commit to hypotheses that were based on those assumptions. B strongly advocated sticking to what she knew as a basis for hypotheses, not what might be true.
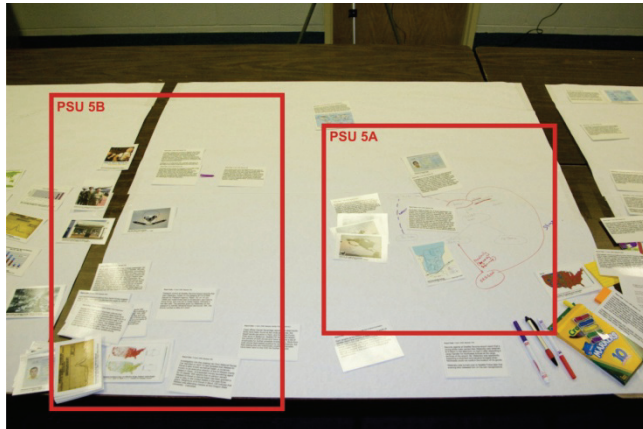

Figure 7. Participants work separately due to disagreement

3.2.5   Organizational Metaphors

One reason for conducting synthesis experiments with physical artifacts on a blank workspace is to see what types of organizational metaphors participants employ in the absence of a pre-determined method. Results from our experiments indicate that a multitude of methods are often mixed together.
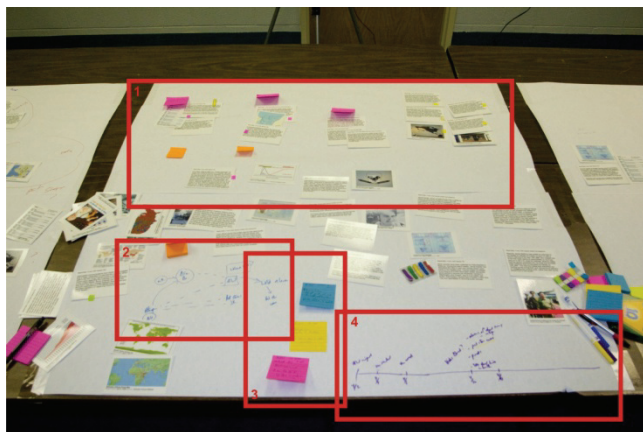

Figure 8. Collaborative workspace for 4C, showing the use of multiple organizational methods

The participants in PSU 4C mixed together four metaphors to develop their collaborative workspace (Figure 8). Along the top of the workspace (1) they created hypothesis groups that were tagged to indicate relevant times, places and other attributes. At bottom left (2) a social network was drawn to determine which people

were related to each other in some way. At bottom center (3) the set of final hypotheses were summarized onto separate post-it notes and ranked. At bottom right (4) the participants drew a graphical timeline and plotted key events along it to try and discern the validity of a particular hypothesis.

PSU 1C (Figure 9) used post-it notes at the top left (1) of the workspace to represent hypotheses. At top right (2), post-it notes were annotated and arranged with drawn links between them to develop a representation of a social network. Across the bottom (3) 1B drew a large, simplified version of an artifact about H5N1 virus growth to use as a timeline reference for artifacts in their hypotheses.
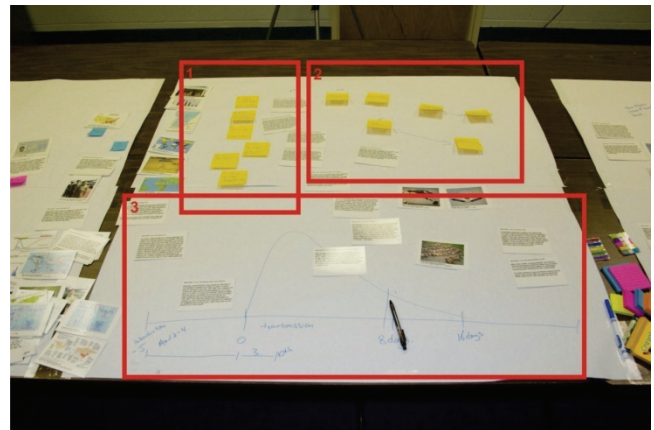

Figure 9. Collaborative workspace for 1C, showing the use of multiple organizational methods

3.2.6   Preserving prior work

An important aspect of collaborative synthesis is the use of pre-existing artifact collections. In the experiment, participants were provided with a complete duplicate set of artifacts and a blank workspace to use if they so desired to complete the hypothesis ranking task. All groups except for PSU 3C chose to use all or part of the duplicate artifact set to create new artifact groups on the blank collaborative workspace. PSU 3C participants used the collaborative space only to write down a timeline of events and to rank post-it notes that represented hypotheses.

Groups PSU 1C, 3C, and 4C verbalized their desire to maintain their individual work in its original state to make it easier for them to recall their findings. PSU 3C deviated from this goal over time, moving artifacts from participant A's workspace over to B's workspace, where they worked from during most of the experiment. In the debriefing where participants were asked to discuss their work, groups that altered their original work stated it was an unwise decision, as it became difficult to recall their work once their personal space had been substantially altered.

4   DESIGN IMPLICATIONS FOR FUTURE SYNTHESIS TOOLS

Results from collaborative experiment evidence suggest that supporting collaborative synthesis will require flexible tools that begin by helping users establish common ground. Users should be able to walk others through their prior work, narrating how they have structured their information, and helping them quickly identify which pieces they have in common.

Collaborative synthesis tools need to support a wide range of organizational types (timeline, network, category groups, etc…) and allow for open forms of annotation and tagging. Experimental evidence did not show that participants gravitated toward single, global organizational strategies like those that are commonly

implemented in software. Participants frequently mixed together multiple organizational metaphors on the same workspace. This finding supports the call for flexibility that Isenberg et al. note regarding the process of analysis with visualization artifacts. Their study suggests that collaborative visual analytics tools should allow users to change their analysis strategy at will, and that a flexible range of workflow processes must be supported [7]. Our work here indicates that flexibility should also extend to the ways in which users are allowed to organize their information.

Developers of synthesis support tools should consider that groups will choose different strategies for approaching collaborative tasks. Some will begin with an emphasis on previously developed hypotheses, others might focus attention on which information overlaps and which does not, and still more may begin by reframing all of the information by some measure such as time, source, or certainty.

It is also important to support collaborative role assignment, as participants in experiments often chose to split tasks among them to complete their work. It is reasonable to expect that analysts from different backgrounds in more complex collaborative settings will also seek to adopt roles that suit their specific skills and expertise. Synthesis support tools need to provide users with the ability to customize interfaces accordingly.

The experimental setting we used allowed users to freely move and manipulate artifacts on the workspace, and to augment them with a variety of common office tools. Most users availed themselves of these tools, but they applied them in a variety of ways. Experimental evidence suggests that the goal for synthesis support software should be to mimic the simple flexibility of real objects while enabling the search/sort efficiency and multiple views that software can provide.

5   Conclusions and Future Research

This study provides experimental evidence to characterize collaborative synthesis and suggest design directions for future synthesis support tools. We have presented a new experimental approach for examining this important topic which uses a simulated real-world scenario to provide a variety of insights into how users collaborate to synthesize information.

While some of the results discussed here indicate the influence of the choices associated with designing the experiment (zooming and grouping items, for example), we can assume that real-world collaborative synthesis will involve similar conditions. Users will bring their prior results which have already been organized, and this will include items that other collaborators do not have.

Many important areas related to synthesis remain for investigation. For example, our work has imposed a single format for analytical artifacts. Little research has been done to evaluate different ways of representing analytical results.

Also, we do not know if what we have presented here will hold true when analysts are faced with much larger numbers of artifacts. Our experiment focuses on a tactical situation, in which participants have a short amount of time to tackle a problem. In other important scenarios, analysts will spend months or years working on a topic and may develop more elaborate collections.

Finally, our experiment design featured pairs of participants, and in many situations we can anticipate that much larger groups of analysts may be asked to work together to synthesize information. Future research is needed to characterize how collaborative synthesis takes place in larger group settings.

This research is part of a larger project to characterize and develop design guidelines for synthesis support. Our aim is to combine results from individual and collaborative synthesis experiments to develop a general design framework for use in the development of future visual analytic synthesis support tools. If

visual analytic tools can provide analysts the answers they seek, they must also help analysts collect, organize, and share their results in useful ways.

References

[1].   Robinson, A.C. *Synthesizing Geovisual Analytic Results*. in *IEEE Symposium on Visual Analytics Science and Technology*. 2007. Sacramento, CA: IEEE Computer Society.

[2].   DiBiase, D., *Visualization in the Earth Sciences*. Earth and Mineral Sciences, Bulletin of the College of Earth and Mineral Sciences, The Pennsylvania State University, 1990. **59**(2): p. 13-18.

[3].   MacEachren, A.M. and D.R.F. Taylor, *Visualization in Modern Cartography*. 1994, Oxford: Pergamon Press.

[4].   Pirolli, P. and S. Card. *The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis*. in *International Conference on Intelligence Analysis*. 2005. McLean, VA.

[5].   Heer, J. and M. Agrawala. *Design considerations for collaborative visual analytics*. in *IEEE Symposium on Visual Analytics Science and Technology*. 2007. Sacramento, CA. IEEE Computer Society.

[6].   Isenberg, P. and S. Carpendale, *Interactive tree comparison for co-located collaborative information visualization*. IEEE Transactions on Visualization and Computer Graphics, 2007. **13**(6): p. 1232-1239.

[7].   Isenberg, P., A. Tang, and S. Carpendale. *An exploratory study of visual information analysis*. in *ACM Conference on Human Factors in Computing Systems*. 2008. Florence, Italy.

[8].   Wright, W., et al. *The sandbox for analysis - concepts and methods*. in *ACM Conference on Human Factors in Computing Systems*. 2006. Montreal, Canada: ACM.

[9].   i2, *Analyst's Notebook*. 2007, i2 Incorporated: McLean, VA.

[10].   Sellen, A. and R. Harper. *Paper as an analytic resource for the design of new technologies*. in *ACM Conference on Human Factors in Computing Systems*. 1997. Atlanta, GA. ACM.

[11].   Hughes, F. and D. Schum, *Discovery-Proof-Choice, The Art and Science of the Process of Intelligence Analysis - Preparing for the Future of Intelligence Analysis*. 2003, Washington, DC: Joint Military Intelligence College.

[12].   Booker, J., et al. *High-resolution displays enhancing geo-temporal data visualizations*. in *ACM Southeast Regional Conference*. 2007. Winston-Salem, NC: ACM.

[13].   Woods, D.K. and C. Fassnacht, *Transana*. 2007, Wisconsin Center for Education Research: Madison, WI.

[14].   Yawn, B.P. and P. Wollan, *Interrater reliability: completing the methods description in medical records review studies*. American Journal of Epidemiology, 2005. **161**: p. 974-977.

[15].   Mackinlay, J.D., P. Hanrahan, and C. Stolte, *Show me: automatic presentation for visual analysis*. IEEE Transactions on Visualization and Computer Graphics, 2007. **13**(6): p. 1137-1144.

[16].   Chuah, M.C. and S.F. Roth. *Visualizing common ground*. in *IEEE Seventh International Conference on Information Visualization*. 2003. London, UK. IEEE Computer Society.

[17].   Zhu, H. *Some issues of role-based collaboration*. in *IEEE Canadian Conference on Electrical and Computer Engineering*. 2003. Montreal, Canada: IEEE Computer Society.