SANDIA REPORT

SAND2016-10412
Unlimited Release
Printed October 2016

Information Theoretic Measures for Visual Analytics: The Silver Ticket?: A Summary of a 2016 Exploratory Express LDRD Idea and Research Activity

Laura A. McNamara, Travis L. Bauer, Michael J. Haass and Laura E. Matzen

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-mission laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: http://www.ntis.gov/search

SAND2016-10412
Unlimited Release
Printed October 2016

Information Theoretic Measures for Visual Analytics: The Silver Ticket?

Laura A. McNamara, ORG-05346
Travis L. Bauer, ORG-05635
Michael J. Haass, ORG-01461
Laura E. Matzen, ORG-01463

P.O. Box 5800
Albuquerque, New Mexico 87185-MS0519

Abstract

In the context of text-based analysis workflows, we propose that an effective analytic tool facilitates triage by a) enabling users to identify and set aside irrelevant content (i.e., reduce the complexity of information in a dataset) and b) develop a working mental model of which items are most relevant to the question at hand. This LDRD funded research developed a dataset that is enabling this team to evaluate propose *normalized compression distance (NCD)* as a task, user, and context-insensitive measure of categorization outcomes (Shannon entropy is reduced as order is imposed). Effective analytics tools help people impose order, reducing complexity in measurable ways. *Our concept and research was documented in a paper accepted to the ACM conference Beyond Time and Error: Novel Methods in Information Visualization Evaluation*, part of the IEEE VisWeek Conference, Baltimore, MD, October 16-21, 2016. The paper is included as an appendix to this report.

ACKNOWLEDGMENTS

We are grateful to the Sandia LDRD program, and particularly the Exploratory Express program, for funding our quick-turnaround project and enabling us to publish a paper that we have long contemplated.

1. INTRODUCTION

Consider the many tools claiming to help analysts find valuable information in very large datasets: Palantir, SAS TextMiner, IBM LanguageWare, and others. The core of any analytical tool is the ability to select smaller subsets of that data that share things in common/ How do I know if one of these analysis/visualization tools is "better" than another for analytic workflows at enabling a person to distill data? Informatics researchers have focused on "insight" as a key outcome measure for comparing the goodness of analytic software – but what counts as an "insight," and how to measure those events, is an open question and not general. Most measures are subjective and/or tied to context. This makes it very difficult to compare across different tools and problems. This isn't satisfactory. We need a robust, quantitative, context-independent metric that indicates how well an analytic tool promotes human performance. That metric should be based in a theory of human-information interaction that aligns with known principles of human perception and cognition. We have a candidate metric that meets these criteria: complexity reduction, measured using algorithms such as *normalized compression distance* (NCD).

This report briefly describes how we developed a dataset that is enabling us to examine NCD and other information theoretic measures as process metrics to assess whether a visual analytics tool is enabling the user to impose order on otherwise unwieldy text datasets.

2. PROBLEM STATEMENT

Information Visualization and Visual Analytics (InfoVis/VA) comprise the interdisciplinary science, engineering, and design of using computer hardware and software to augment the perceptual and cognitive experience of discovery, understanding, and insight. The idea that visualization of data and information promotes insight has inspired an entire generation of computer scientists to develop visualizations and interactive visual analytics aimed at helping humans discover, learn, explain, and make decisions with data. However, 'insight' is a notoriously difficult event to measure, which makes it difficult to implement in the applied context of evaluating a software tool.

In its place, we propose process metrics that enable researchers to examine the degree to which a system enables people to reduce the amount of information they are dealing with, so they can focus their attention on the items that are germane to the topic at hand. These process metrics recognize that all analysis workflows begin with information collection and categorization, or "triage," which reduces the complexity of large information sources while enabling users to build a mental model of content in relation to a question of interest. In the context of text-based analysis workflows, we propose that an effective analytic tool facilitates triage by a) enabling users to identify and set aside irrelevant content (i.e., reduce the complexity of information in a dataset) and b) develop a working mental model of which items are most relevant to the question at hand. Effective analytics tools help people impose order, reducing complexity in measurable ways.

One such metric is *normalized compression distance* (NCD), as a task, user, and context-insensitive measure of categorization outcomes. NCD is a feature- and parameter-free way of evaluating the similarity of two or more information objects [53]. The simplicity and generalizability of such compression-based measures lend them to a wide range of applications, and we are intrigued by the idea of using them as a process measure of information interaction in visual analytic systems.

2. RESEARCH ACTIVITY

In summer 2016, the Exploratory Express program in Sandia National Laboratories LDRD office provided $60K in funding for our team to develop a dataset that would enable us to study NCD and similar information-theoretic measures for assessing the impact of automated support in data and information categorization tasks.

We proposed and executed a counterbalanced, between-subjects data collection activity in which 18 research participants drawn from the Sandia National Laboratories' professional population used one of three tools to a) categorize a diverse set of documents and b) use a subset of the documents to examine a real-world intelligence problem, namely the involvement of Russian assassins in the polonium poisoning of Russian oligarch-turned-dissident Alexander Litvenenko in London (November 2006). We provided participants with an electronic set of 370 documents, ranging from dense journal articles, to government reports, to articles from popular news magazines. Articles directly related to Litvenenko murder comprised approximately 20% of the total dataset.

The Sandia Human Studies Board approved our proposed data collection activity in July 2016. We recruited participants through an advertisement in the Sandia Daily News. Participants were accepted on a first-come basis and randomly assigned to one of three different workflows, or conditions. Two of the conditions use the Sandia National Laboratories' Citrus document analysis platform, while a third "control" condition requires participants to perform the task using the Windows filing system.

Participants spent a total of four hours over two days in this activity. On the first day of the activity, we trained all participants in basic use of the Durian tool in the Sandia Citrus text analysis platform. We then randomly assigned participants to one of the two Citrus conditions (both of which used different functionalities in Durian), or to the Windows file system condition. We then told the participants to review and sort as many of the documents as possible into topical categories, using whatever categorization scheme they deemed appropriate. We gave participants 90 minutes to complete the categorization activity. Once the time was up, each participant filled out the NASA Task Load Index (NASA TLX), a widely used tool for eliciting the subjective experience of workload.

On the second day of the activity, participants were given two hours to use the same dataset to answer a set of questions about the Alexander Litvinenko murder. We asked them to create a document folder or "bin" containing the items they used to respond to the questions. We also requested that they create a separate bin containing items they would recommend for understanding the impact of the Litvinenko murder on relations between the Russian government and its European and American counterparts. At the two hour time mark we stopped work and requested the participants to fill out a second copy of the NASA TLX.

3. OUTCOMES

This LDRD investment enabled us to assemble a unique experimental dataset that we can use to evaluate our ideas about compression and other distance measures as process metrics for evaluating automated support for text analysis tools and workflows. We also wrote a position paper that included a summary of our then-ongoing study for submission to the ACM Beyond Time and Error: Novel Methods in Information Visualization Evaluation (BeLIV) conference, held as part of the IEEE VIsWeek/VAST

Conference (Baltimore, MD, October 16-21, 2016). The paper was accepted in the first round of reviews after minor revisions and will be presented at the conference.

We are continuing to analyze the data collected in FY16 and will be compiling our findings in a journal-quality publication. We expect the data to show that people can successfully sort documents into semantically coherent sets, and that our selected compression metric should indicate greater *within-category* similarity compared with the dataset treated as a single category. At the same time, we expect that *between-category* similarity to fall as people create "patches" of semantically related information; i.e., as they add structure to the unstructured mess of information that they began with. We also expect that automated support for categorization will facilitate users' efforts to impose structure on the chaos we give them. Specifically, when using Citrus, we expect research participants will organize documents into categories more quickly and efficiently, with greater reductions in NCD compared to performance of the same task using the Windows file system on a computer desktop.

We will be completing data collection and analyzing the resulting data in the summer and fall of 2016. Our follow-on publication will document our experience applying compression distance measures to assess how these computer-mediated interaction models facilitate the reduction of complexity in a realistic analytic workflow.

Please see Appendix A for the ACM BeLIV paper, which details the principles and ideas that generated this research and a complete bibliography.

4. APPENDIX A

McNamara, L.A., Bauer, T.L., Haass, M.J., Matzen, L.E. 2016 "Information Theoretic Measures for Visual Analytics: The Silver Ticket?" BELIV '16, October 24 2016, Baltimore, MD, USA DOI: http://dx.doi.org/10.1145/2993901.2993920

Information Theoretic Measures for Visual Analytics: The Silver Ticket?

Laura A. McNamara
Sandia National Laboratories
Albuquerque, NM 87185 USA
(505) 844-1366
lamcnam@sandia.gov

Travis L. Bauer
Sandia National Laboratories
Albuquerque, NM 87185 USA
(505) 284-8723
tlbauer@sandia.gov

Michael Haass
Sandia National Laboratories
Albuquerque, NM 87185 USA
(505) 844-4060
mjhaass@sandia.gov

Laura Matzen
Sandia National Laboratories
Albuquerque, NM 87185
(505) 844-1505
lematze@sandia.gov

ABSTRACT
In this paper, we argue that information theoretic measures may provide a robust, broadly applicable, repeatable metric to assess how a system enables people to reduce high-dimensional data into topically relevant subsets of information. Explosive growth in electronic data necessitates the development of systems that balance automation with human cognitive engagement to facilitate pattern discovery, analysis and characterization, variously described as "cognitive augmentation" or "insight generation." However, operationalizing the concept of insight in any measurable way remains a difficult challenge for visualization researchers. The "golden ticket" of insight evaluation would be a precise, generalizable, repeatable, and ecologically valid metric that indicates the relative utility of a system in heightening cognitive performance or facilitating insights. Unfortunately, the golden ticket does not yet exist. In its place, we are exploring information theoretic measures derived from Shannon's ideas about information and entropy as a starting point for precise, repeatable, and generalizable approaches for evaluating analytic tools. We are specifically concerned with needle-in-haystack workflows that require interactive search, classification, and reduction of very large heterogeneous datasets into manageable, task-relevant subsets of information. We assert that systems aimed at facilitating pattern discovery, characterization and analysis – i.e., "insight" - must afford an efficient means of sorting the needles from the chaff; and simple compressibility measures provide a way of tracking changes in information content as people shape meaning from data.

CCS Concepts
• Human Centered Computing→Empirical Studies in HCI • Human Centered Computing→User Models
Keywords Visual Analytics; Evaluation; Human-Information Interaction; Complexity Reduction; Metrics.

BELIV '16, October 24 2016, Baltimore, MD, USA
© 2016 ACM. ISBN 978-1-4503-4818-8/16/10…$15.00
DOI: http://dx.doi.org/10.1145/2993901.2993920

1. INTRODUCTION
"The purpose of visualization is insight, not pictures" [11]. This declaration has inspired an entire generation of computer scientists to develop visualizations and interactive visual analytics aimed at helping humans discover, learn, explain, and make decisions with data.

Information Visualization and Visual Analytics (InfoVis/VA) comprise the interdisciplinary science, engineering, and design of using computer hardware and software to augment the perceptual and cognitive experience of discovery, understanding, and insight [8, 21, 48]. In an era of unprecedented data fecundity, insight is exactly what we need, as we collectively try to figure out what all this data *stuff* can tell us about ourselves and the world around us. Stored in electronic bits, our masses of data require the mediating power of computational hardware and software if we are to glean any meaning from what we are creating. Vision is the dominant human sense and is an obvious channel for engaging data and information. Elegant, engaging visualizations and interactive visual environments that facilitate human perception and cognition are critical if we are to realize – even democratize – the opportunity to gain knowledge from data.

Like many of our counterparts in informatics, visualization, and visual analytics, we would love to deploy software environments that facilitate insight through an enjoyable interactive experience. Like them, we have struggled to operationalize the concept of insight in our work. What does it mean to design for insight? What perceptual and cognitive processes are involved, and how can we design tools that augment those? How can we evaluate whether our prototypes are meeting insight-related design goals? What *is* insight, really?

2. MOTIVATION AND OVERVIEW
The authors are an organizational anthropologist, a cognitive neuroscientist, a cognitive/ computer scientist and a physicist who are part of a larger Human Analytics research function at Sandia National Laboratories. Human Analytics is really about "analytics for humans." We are interlocutors in the world of national security informatics, moving back and forth between analyst-users and researcher-developers. Our methodological toolbox includes everything from ethnographic observation to cognitive task analysis to controlled experimental studies that use gaze tracking to document visual search strategies. Our projects have brought us into contact with a wide range of user communities, including imagery analysts, nuclear weapons engineers, network security analysts, and all-source intelligence analysts across the United States' federal government. In this paper, we refer to these communities collectively as "analysts."

Our raison *d'etre* is providing empirical data on human-information interactions in the analytic communites we engage; to inform the design, development and evaluation of new algorithms, software, and visual workflows for analysts. This entails close partnerships with remote sensing scientists, computer scientists, mathematicians and software engineers. In this paper, we collectively refer to these colleagues as "developers."

Developers want to create tools that are usable, useful, and ultimately adoptable. When they are successful, their algorithms, software and workflows can help analysts make sense of patterns, anomalies, and signatures that might otherwise be difficult to discern. This goal is important as analysts face increasingly unwieldy collections of electronic data and information (practically speaking, *unwieldy* means more electronic data than analysts can manage with conventional desktop applications).

In reflecting on our experiences for this paper, we realized that the idea of insight has not played a salient role in our team's approach to analytic workflows. When conversing with our InfoVis/VA colleagues, we often talk about systems to help identify complicated geospatial and/or temporal patterns, to detect anomalies and characterize signatures, or to identify critical sections of text for addressing a key intelligenc question. All such events arguably constitute "insights" in the domains we work with.

In reality, however, the biggest challenge for analysts is not detecting a signature or putting together information to gain knowledge – they are quite good at that. What they really need are efficient, dependable ways to clear away irrelevant data and information, so they can focus their attentional resources on the messages that matter. As any analyst will attest, efficiently discovering and characterizing signatures depends on one's ability to quickly and efficiently eliminate irrelevant information.

2.1 From Insight to Complexity Reduction

In this paper, we discuss why we believe that *complexity reduction* is critical for the visual analytics systems that InfoVis/VA researchers are pursuing. Our ideas are informed by our interactions with analysts, coupled with our ongoing interdisciplinary discussions about exploring ways to evaluate analytic systems and measure complexity reduction in dynamic data environments [4, 16, 19].

In regards to visualization and visual analytics, the idea of user-directed complexity reduction is hardly new. Consider, for example, Shneiderman's famous mantra: *overview, zoom and filter, details-on-demand* [47]. Each of these design principles can be framed as a means for accomplishing user-driven complexity reduction. It follows that process measures that focus on the intermediate outcome of complexity reduction may provide a quantitative indicator that an analyst is narrowing their input deck to items they deem relevant.

Moreover, making complexity reduction an explicit requirement for visual analytics has some intriguing implications for design and evaluation: *First*, we assert that complexity reduction is a necessary condition for people to experience the happy event of insight. It is probably not sufficient. However, we sincerely doubt that visualization tools that fail to afford complexity reduction will effectively facilitate the sensory, perceptual, and cognitive work associated with insight.

Second, in terms of design studies, emphasizing complexity reduction suggests looking for sources of complexity and examining native strategies for managing it. It also suggests that visual analytics will be more effective in supporting analytic work to the extent that technologies enable user-driven complexity reduction. In doing so, it provides both an overarching common goal toward which analytical tools can strive along with quantitative, straightforward, a comparable ways to measure their success.

Third, emphasizing complexity reduction as a design goal opens the door to computationally tractable metrics for evaluating how well visual analytic systems support this requirement. In particular, information theoretic metrics (many of which derive from Claude Shannon's ideas; see [22]), can be used to determine how well a system enables users to reduce complexity (and, by extension, enrich an information patch; see [40]). Broadly speaking, information theoretic measures (for example, normalized information compression distance; see [53]), may provide a source of unobtrusive, parsimonious, quantitative, generalizable metrics for evaluating how well a visualization workflow helps people manage information complexity, and facilitates the happy event of insight [11].

2.2 Topical Overview

This paper begins by discussing the concept of insight, both in Infovis/VA and cognitive psychology. We then discuss how InfoVis/VA researchers have addressed the fuzziness of insight by focusing on the empirically documented, concrete processes of information foraging and sensemaking. Such processes can be understood in terms of information theoretic measures. We provide some examples, both factual and fantastic, of ways in which humans experience perceptual and cognitive overload when dealing with undifferentiated information spaces. We emphasize that the class of measures we are proposing are not the "golden ticket" of insight evaluation (to borrow a metaphor from Roald Dahl's Willy Wonka). However, we do believe they may provide a "silver ticket" for tracking the nitty gritty of human-information interaction as people are working toward discovery and insight.

3. YOU KNOW IT WHEN YOU HAVE ONE

Insight is an intuitively elegant design goal, but it is difficult to operationalize practically in system design and evaluation. Key issues include defining what constitutes an insight, how to measure it, and valid evaluation models that express insight in relation to specific actions and processes supported by visualization systems.

3.1 What is Insight?

InfoVis/VA researchers have not arrived on practical definition of insight for their field, though this is neither for lack of thought, enthusiasm nor effort; for example [36, 39, 43, 45, 49]. Certain qualities pertaining to insight have been proposed. North, for example, suggests that insights are complex, deep, qualitative, unexpected, relevant. In his model, the more strongly any insight expresses such qualities, the more significant it is likely to be [36]. Similarly, Shneiderman and Plaisant relate insight to discovery, which they describe as an individualistic, unpredictable event that occurs in the context of engaging data and information [49]. Chang et al. suggest that the visualization community is actually working with two different definitions of insight: one that pertains to a moment of enlightenment, the other to an advance in knowledge-building [13].

Cognitive psychologists have been studying the problem for decades and have yet to arrive at a consensus definition of "insight." Decades' worth of research has examined what defines insight, how it differs from other cognitive events, the relationship between meaning and insight, conditions that give rise to insight, the mental processes of insight; and, more recently, areas of the brain implicated in meaning-making and insight events [5, 7, 8, 9, 38, 44, 51]. The topic remains an active, even cacophonous area of research activity, and we mean that in an entirely positive way. The literature is diverse, creative, lively, even intense – which is not surprising considering how fundamental insight is to the human experience.

The approaches to insight most germane to InfoVis/VA stem from research on expertise, tacit knowledge, and decision-making (see discussions in [17, 31, 40, 51]). Experimental and observational studies examine how individuals approach poorly constrained problems: situations that are ambiguous, unpredictable, open-ended, and/or dynamic, and for which effective intervention is not easily discerned. This literature tends to view insight as a re-framing of the problem leading to a strategy for an acceptable solution.

Although we commonly associate insight with the excitement of an “a-HA!” moment, an individual’s perspectives and understandings can also shift more subtly over time as the brain consolidates new information and evolves its frameworks [51]. Insight events may feel spontaneous, but research indicates that they tend to occur only when certain conditions have been met. First, the individual must be motivated to address a problem that challenges existing mental models. The problem cannot be overly constrained, lest the solution be obvious. Pre-existing information and experiential knowledge support initial problem framing, but poorly-defined problems often force the individual to articulate, examine, and question taken-for-granted or latent mental models. The individual is likely to seek additional information that addresses perceived gaps in the mental model. Some of this information will be integrated into established semantic frameworks, some will be discarded. As the semantic frameworks evolve, so does the subjective framing of the problem as new conceptual relations are formed. At some point, a new perspective crystalizes, opening the door to a strategy for addressing the problem [32, 51].

This process of organizing, consolidating, updating, and then re-appying our mental models is how we establish stable meaning from lived experience. The human brain has evolved to impose order on the continuous chaos of sensory input, which is inherently meaningless if we cannot parse signals from the sea of noise [5].

We are reminded of a scene in the book *Mind Wide Open*, in which science writer Steven Johnson recounts his experience as a test subject in an fMRI study examining brain activity associated with cognition, problem solving, and insight. Part of the experiment uses a stimulus set consisting of a black-and-white checkerboard, followed by an image consisting of randomly scattered black dots on a white background. Joy Hirsch, the researcher who is demonstrating fMRI to Johnson, explains how participants respond to both: “When you have noise (the field of random dots), the whole brain seems to light up trying to make sense of it,” she tells him. In contrast, when the field of random dots disappears and is replaced by the checkerboard, brain activity localizes to a few key areas, including the prefrontal cortex. “The checkerboard is reassuring,” she says. Randomness engages our brain to a degree that Johnson admires: “There was something lovely in that image: the brain, faced with apparent chaos, leaning on all of its resources looking for some hope of order in the mix” (179-182 in [28]). It seems our brains are continuously reaching for order, even when order is objectively impossible.

Yet actually measuring insight events remains a difficult challenge. Psychologists have traditionally relied on observational studies to capture behavioural manifestations of insight, such as pattern recognition and strategic problem-solving. For example, experimenter might document how research participants move chess pieces in a game, if they solve a puzzle, or if participants report learning something new or changing a problem-solving strategy.

More recently, cognitive neuroscientists have started examining the physiological and neural correlates of insight events [32, 44]. Much of this work relies on non-invasive sensing, such as MEG, EEG or fMRI, all of which provide evidence of brain activity with varying degrees of spatiotemporal accuracy and precision. New nmeasurement technologies are facilitating entirely new understandings about how our brains create meaning from the chaos of experience. This is very exciting for cognitive neuroscientists, but rather frustrating for those of us who would like a straightforward way to assess insight events among study participants.

3.2 Insight Evaluation in Visual Analytics

In visual analytics, the problem of evaluating insight has generated a quite impressive body of literature that examines how well visualizations and analytic workflows achieve their goals.

Over the past two decades, researchers in InfoVis/VA have pulled creatively from cognitive psychology and other disciplines, including anthropology, design, human-computer interaction, to develop a vis-oriented perspective on technologies of insight. The result is a rich literature of design and evaluation frameworks that address every facet of visually-oriented analytic systems, from the nitty-gritty of algorithmic scalability all the way to interactive workflows, user experience, and ecological fit (for example [35, 36, 46]). Visualization practitioners may pick from a rich array of quantitative, and mixed-method approaches for evaluating visualization tools, with designs ranging from controlled laboratory experiments to in-situ longitudinal observation [12, 34].

Many of the community’s frameworks emphasize insight as a desired outcome in their studies. Insight-focused evaluations have tended to rely solutions to problems or the subjective reporting of an insight event – not dissimilar to the way cognitive psychologists have approached the problem. Field study narratives can be quite rich in detailing how visualization supports interactive discovery, learning, and information synthesis [3]. For example, InfoVis/VA researchers have looked to qualitative field study approaches, such as Pair Analytics or the Multidimensional In-depth Long Term Case Study (MILCS) as a way of capturing insight events [3, 49]. These usually entail researchers collecting both structured and unstructured qualitative data to capture goals, strategies, and tasks in a workflow; users participate by providing details about the occurrence and qualities of insight events [37, 45]. The MILCS approach is holistic and participatory, and is perhaps most informative when users are highly motivated to record their experiences with a visualization tool. These subjective user accounts are integrated with the researcher’s observational findings and available quantitative data (e.g., log files). Insight events are captured through self-report and may include details about the insight, such as its semantic content, novelty, or depth.

3.3 Evaluating Processes and Outcomes

In her thoughtful tour of InfoVis/VA evaluation methodology, Carpendale describes the characteristics of an optimal evaluation metric: it should be generalizable, precise, and ecologically valid [12]. To Carpendale’s criteria, we would add that it should be unobtrusive and uncomplicated to deploy. In addition, it is important to differentiate between *process* and *outcome* metrics. The former pertain to the *how* of a task, while the latter pertain to the *what* of a task, such as a measurable change on one’s state-of-knowledge [18]. In relation to analytic reasoning and insight, process metrics speak to the behaviours and actions involved in generating insights, while outcome metrics tell us something about the event of insight, perhaps even how well that insight accords with the real world (ground truth).

Insight-related outcome metrics that maximize Carpendale's criteria set are particularly tricky. One outcome metric is whether or not the user arrived at a correct judgment or conclusion, which is only measurable if ground truth exists for the problem at hand. Visualization demonstrations, contests and user studies that incorporate ground truth measures usually rely on synthetic and/or training data for which outcomes are well-characterized (for example [29, 30]). This is less feasible for problems that lack ground truth, which unfortunately describes the majority of real-world analysis activity that visualization tools support.

To borrow from Roald Dahl's Willy Wonka, outcome metrics that meet Carpendale's criteria are the "golden ticket" of insight evaluation. Judging the correctness of an insight will always be difficult without ground truth. However, in principle, detecting the *event* of insight during analytic work should be possible, assuming we have an adequate model of the relevant perceptual and cognitive processes in the context of visual interaction with electronic information [21, 23, 24, 13]. Such a measurement framework might be rooted in a theory of insight that incorporates emerging findings from cognitive neuroscience [32]. It would probably also require neuroscientific data collection systems capable of recording insight-related brain activity as users perform tasks with a visualization tool. Neither exists yet, so we will have to wait for other research communities to catch up with the knowledge goals of the InfoVis/VA community. Perhaps in a few decades, we will have our golden ticket: a metric (probably a suite of metrics) for insight that is generalizable, precise, ecologically valid, unobtrusive, safe, and robust to variations across tasks, individuals, and work ecologies.

In the meantime, however, developing *process* measures for interactive visual analytic workflows may be a more tractable challenge. This brings us to the topic of human-information interaction and theories of information foraging and sensemaking, which describe the processes through which we reduce complexity and establish order in data and information.

3.4 Sensemaking, Foraging, and "Insight"

Given the difficulties associated with defining and measuring insight, it is perhaps no surprise that InfoVis/VA researchers have also examined the processes and activities that are likely to enable insight [25, 33]. Amar and Stasko, for example, identified ten analytic primitives that underlie exploratory query activities in visual analytic workflows, positing that visual analytic systems should support these primitives to enable higher-order problem solving [1, 2]. Chang *et al*. suggest that visual analytics might aspire to expanding a user's knowledge base, thereby increasing the probability of a spontaneous insight [13]. Munzner and colleagues have focused attention on the challenges to analytic system validity and proposed a design framework to articulate and manage such challenges [35, 46].

Sensemaking theory has also provided a rich source of guidance for interactive system design [21]. Sensemaking may refer to any of several distinct research traditions in organizational theory, decision science, cognitive psychology, and information science [17, 31, 52]. Despite being rooted in different academic disciplines, these literature generally agree that sensemaking is a fundamental sociocognitive process through which individuals establish orderly models of the world from what is sensed and perceived.

3.4.1 Sensemaking and Foraging

Within the InfoVis/VA literature, Pirolli and Card's representation-construction model is the dominant model of sensemaking [15, 40, 41]. Derived from empirical observations of intelligence analysts, this model provides explicit guidance for developing systems to facilitate exploratory human-information –a process that supports what Stuart Card has described as "knowledge crystallization" [10].

A distinguishing theme in Pirolli and Card's sensemaking model is the integration of their ideas about information foraging, which draws upon behavioural ecology to model information search as an adaptive strategy that trends toward optimal efficiency. In any information space, relevant sources are unevenly and probabilistically distributed. As human "informavores" learn to navigate an information space, they become increasingly efficient identifying the relevant sources; i.e., they spend less time navigating *between* information patches and more time harvesting *within* the patches that are more likely to contain valuable resources.

Foraging theories tell us that effective information systems should enable users to a) quickly learn the contours of an information space and b) associate the contours of that space with the task at hand. Moreover, a well-designed system will also enable informavores to create their own patches of semantically desirable resources that support completion of the task. Doing so minimizes the need for global search, freeing limited attentional resources for the cognitive work of assimilating and integrating new knowledge [9, 40, 42, 51]. This latter activity is what Pirolli, Card, Russell and others describe as "sensemaking."

3.4.2 Sensemaking in InfoVis/VA Design

Ideas about information foraging and sensemaking have moved outside of psychology and informatics and are widely accepted as a basis for the design of analytic systems. In the analytic communities we work with, sensemaking is understood as process through which analysts attend to selected sources, integrating percepts and concepts into existing representational schemas to establish a plausible, evidence-appropriate explanatory narrative for events [41, 42]. Information foraging and sensemaking theories also underpin a great deal of InfoVis/VA work, informing design guidance (see for example [26, 27]) and working technologies for both individual and collaborative workflows ([6, 30, 50]).

Sensemaking design guidance tends to emphasize the positive interaction elements of a visualization tool. By "positive" we mean those affordances thought to *promote* perceptual and cognitive engagement with data and information. However, another tack for framing the design question is to ask, "What conditions *undermine* effective sensemaking?" Flipping the problem on its head opens the door to evaluating analytic workflows in terms of how effectively they mitigate *against* certain specified conditions, as opposed to promoting others.

This is somewhat analogous to Munzner's ideas about threats to validity. Her nested model asserts that visualizations will fail when developers get the problem, abstraction, idiom and/or algorithms wrong. Getting these right many not be *sufficient* for realizing a useful, usable, adoptable system, all are *necessary* to this end [35]. Similarly, visual analytic workflows should not only promote conditions associated with effective problem-solving; they should explicitly minimize conditions that undermine it. To this end, we might consider the kinds of human-information interaction processes that undermine our native sensemaking abilities, so we can design against those conditions.

3.5 When Sensemaking Fails

Although sensemaking comes naturally to most of us, it is not failure-proof; there are conditions under which we are more or less effective in putting our experience into order. Pirolli and Card's sensemaking models derive from empirical descriptions of analytic workflows, and so they do not address sensemaking failure; nor is it an explicit topic in InfoVis/VA literature. However, it is worth identifying the attention to the conditions under which sensemaking becomes more difficult.

Organization theorist Karl Weick has devoted much of his career to understand the conditions under which sensemaking fails. When our experience of events does not conform to the semantic structures that have served us in the past, and/or when the flow of sensory and perceptual input overwhelms our ability to put it into order, we lose track of what makes sense. By definition, unpredictable events challenge our mental models; under conditions of high information throughput, our experiential knowledge may not enable us to navigate an unmanageable wave of unexpected sensory inputs. We can adjust and repair our models if we have time and space to do so. But when circumstances do not give us time to create and test new mental models, tragedy can ensue - as Weick's studies of the Tenerife airline disaster and the Mann Gulch fire illustrate so poignantly [52].

Weick's studies of sensemaking failure focus on human communication and decision-making in life-and-death situations. Computer scientists and interaction designers rarely have to consider life-and-death risk when designing their systems. But Weick's observations drive home a key point: *Under conditions of high cognitive and emotional load, ambiguity, and rapidly shifting patterns, the subjective experience of disorder undermines sensemaking.*

By extension, if analytic technologies are to minimize sensemaking failures, we should pay attention to the conditions that stress human perception and cognition. Among other things, visual analytic systems should enable users to manage volume and variability in data. This is particularly true for the open-ended, cognitively challenging, exploratory work of data analysis, synthesis, and discovery that so many VA researchers seek to support.

4. INFORMATION, ENTROPY, AND REDUCING COMPLEXITY

To summarize our ideas thus far, we assert that unmanageable amounts of data and information overwhelm our native sensemaking abilities; by extension, insight is impossible when meaningful, recognizable signals are lost in disorder. If we are to design systems that enable effective perceptual and cognitive engagement, we should emphasize features that empower people to reduce unmanageability, according to the contextual and subjective requirements of the work being performed. Happily, this design principle lends itself to evaluation through the application of information theoretic measures, such as entropy and compressibility, that may indicate how effectively a given analytic system enables users to impose sensible order [14]. This brings us to Claude Shannon's ideas about information and entropy.

4.1 A Short Introduction to Shannon Entropy

Shannon information, uncertainty, and entropy are intricately related concepts and notoriously tricky to grasp and explain in written English (or any language, for that matter). A full treatment of Shannon's ideas is well beyond the scope of this paper (although Chen et. Al. [14] write about information theory in visual analytics; also see James Gleick's very entertaining and accessible history of information theory [22]). However, a vernacular discussion of Shannon entropy may be helpful in understanding why we consider complexity reduction as a critical function of visual analytics technologies.

One can think of entropy as a probabilistic way of estimating what we can learn about an entire set, if we can only examine one element from that set. In other words, how well does a randomly selected element represent all items in the set?

Mathematically, Shannon entropy expresses the predictability of any variable X as:

$$H(X) = -\sum p_x \log_b p_x \qquad (1)$$

where p_x expresses probability that X is in state x, b is the base of the logarithm, and $p \log_2 p$ is zero when $p= 0$. The precise unit of measurement depends on the base of the logarithm. For example, if base two is used, the unit is "bits," while a base ten system gives us "Hartleys," named for another of information theory's early pioneers.

Shannon's entropy equation enables quantification of variability in a system. Higher variability means less predictable content; because there is more information, there is more entropy. This is the key point, because most people do not think of "entropy" as informative in the conventional sense. However, remember that Shannon quite explicitly excluded meaning from his formulation of information, as it was not conducive to the engineering problem he was trying to solve [4]. Instead, entropy reflects the *complexity of the vocabulary required to encode a stream of messages from some system*.

This is an engineering problem, not a semantic one. Systems that require more bits to represent their messages contain a greater volume of information; are therefore more difficult to compress; and by extension, entail higher entropy. Put simply, the more information we are dealing with, the more difficult it becomes to characterize the entire content of a set using any single element drawn from that set.

For example, consider a weather sensor that only indicates whether it is currently raining. This sensor displays "raining" if there is rain, and "not raining" if there is no rain. Furthermore, let us say that it is raining 50% of the time (in other words, the sensor it not in the authors' home state of New Mexico). If we look at the sensor, we have a 50% chance of capturing the system in either state. Zero indicates that it is not raining and a one tells us that it is, so the entropy of our weather system is one bit. The answer to the question of whether it is raining or not takes *one bit*.

However, let us assume we are in a place where it never rains; i.e., you can be certain that it is not raining. In this case, our rain sensor is providing no additional knowledge, because it never rains in this place. In other words, it takes *zero bits* of information to encode the answer to the question; the state of the system is entirely predictable.

A more interesting case occurs when rainfall is more variable - let's say one can expect rain 25% of the time. Responding to the question, "Is it raining?" requires something less than one bit of information. This is a bit of a strange concept, because we tend not to think of information in fractional bits. Arithmetically, however, the representation is quite straightforward and provides a mathematical foundation for tracking changes in the predictability of content based on a single message (sample) from the set.

4.2 The Tedium of Yes-No Questions

Another useful metaphor for thinking about Shannon entropy is the yes/no game of twenty questions: How many yes/no questions must be posed and answered to determine the state x for variable X? Simple systems require very few yes/no questions, as in the weather system described above: "Is it raining? No." Case closed.

As information, entropy and therefore complexity increase, however, more yes/no questions are required to determine the system state. Consider a weather station that measures barometric pressure, temperature, precipitation, etc., in a geographic area with very unstable weather conditions. One (very bad) way of interrogating the system is asking a series of yes/no questions to determine the state of all parameters of interest: "Is barometric pressure below 50 millibars? Yes? Is it above 30 millibars? No?" Etcetera.

Although unrealistic, this example helps with thinking through the implications of Shannon entropy for the perceptual and cognitive work of sensemaking. Serial questions are an inefficient, tiring way of figuring out the state of a system; and the more noise we encounter, the more questions we have to ask, and the more tiring the task.

4.3 The Worst Analysis Workflow in the World

Here is another hypothetical example: consider a group of all-source analysts tasked with producing a detailed report on the history, current production capabilities, and international expansion plans for Willy Wonka's candy factory. Our analysts have access to the International Confectionary Data Portal (ICDP), a repository containing over eight million documents pertaining to confectionary industries around the world. The analysts know that only a small number of available documents will be useful in compiling their report, so they get to work winnowing down the dataset to the items they need. In Pirolli and Card's terminology, they must forage for resources before they can use them in assessing Willy Wonka's industrial capacity.

Now, imagine for a moment that the ICDP lacks any support for user queries. Instead, the interface forces analysts into an item-by-item evaluation: it retrieves an abstract at random and asks the analyst, "Is this item related to Willy Wonka?" If the answer is no, the item goes into the "Junk" category. If the answer is yes, the item is put in the "Wonka" category. Over time, the analysts evolve the Wonka category into enriched information patch comprising a selection items pertaining to Willy Wonka. Assuming the analysts are judging document content correctly, the entropy of that patch is gradually falling relative to the entire ICDP, even as the number of items in the patch increases, because the items are in some way related to Willy Wonka.

However, the analysis group experiences high employee turnover because the work is so tedious; while the government customer is frustrated by the delays in receiving the analysis she has requested. Finally, the lead intelligence analyst decides she's had enough and writes a script to perform Boolean searches against the database. This is nothing short of revolutionary: a sorting task that seemed endless can now be performed with one simple operation. Morale improves as the analysts use the new tool to winnow the ICDP database to fewer than one thousand relevant items. In less than a week, the group has assembled convincing evidence that Willy Wonka is actually verging on bankruptcy and is unlikely to declare a monopolistic chocolate empire anytime soon.

We included this counterfactual fantasy to illustrate the importance of automation in enabling a user to cultivate and enrich a patch of information using whatever semantics she deems appropriate for the problem. When automated tools are not available, people will use whatever is at hand to sort, winnow, and texturize information, structuring it according to experiential knowledge and subjective understanding of the task.

4.4 Enriching Patches, Reducing Complexity

In 2009, we were involved in a Sandia project to develop an document categorization and visualization tool for all-source intelligence analysts. To inform prototype development and evaluation, we designed a task analysis activity with all-source intelligence analysts to capture variations in their strategies for managing text document collections. We used paper copies of intelligence reports in this study so that we could watch analytic selection strategies in action.

As one might expect, which reports a particular analyst found "useful" depended how that analyst interpreted the problem, her experience working in the intelligence community, her familiarity with the type of reporting that we provided, and her self-reported background knowledge for the mock analysis problem we presented. That said, all but one analyst began the task by skimming each document and identifying the less-relevant items. Once this reporting chaff was set aside, the analyst settled into the task of putting the remaining wheat into semantically meaningful categories. The number of sorting iterations, the final number of categories, which documents were grouped together, and the semantic labels assigned to categories differed among individuals. However, every analyst created some form of a "junk" pile to corral items they deemed irrelevant for their analysis (with the exception of one analyst, who told us he would rely on his domain knowledge rather than use the reports we provided).

This kind of categorization is probably familiar to all of us. Categorization is the antidote to entropy: as we concentrate semantically related information into patches, we enhance our ability to predict the content we will encounter as we engage a patch. Categorization also mitigates perceptual and cognitive load, particularly in relation to working memory. When we put items into subjectively meaningful, thematically consistent categories, we are expressing a cognitive model of semantic connections both within and across categories. This reduces the load associated with remembering random items. Adding patchiness to an information space helps people avoid the tiresome task of repeatedly interrogating elements to determine knowledge relevance. It is akin to transforming the random collections of dots into a tidy checkerboard.

Interestingly, in one of their early essays on information foraging and sensemaking, Pirolli and Card get tantalizingly close to information theoretic models of human interaction with data. They describe experts' heuristics for setting noise thresholds, asserting that experience makes experts more efficient at filtering noise to focus on relevant signals. Their foraging loop describes *enriching* an information space in terms of *narrowing* the set of items selected for review [41]. As counterintuitive as it may seem, enriching an information patch necessarily entails the *reduction* of complexity, because patches concentrate like items into meaningful categories. Another way to think about this is to say we enrich information patches by increasing predictability: One glance at the document on top of that pile on your desk tells you what sits underneath, so you can decide whether that particular topic merits your attention at that time. Making content predictable facilitates sensemaking by freeing our attentional resources from the burden of identifying and classifying items, so we can engage the complex work of interpretation.

5. FROM PRINCIPLE TO APPLICATION: COMPLEXITY REDUCTION IN SYSTEM EALUATION

We are not the first to suggest that visual analytics and visualizations can be framed in terms of information theoretic principles. As Chen points out, information foraging theory is highly consonant with Shannon's theories of information, entropy, and uncertainty; measures of Shannon entropy and Komolgorov complexity can be used to indicate heterogeneity, uncertainty, thematic saliency, and shifts in the semantic content of an information stream [14].

It is rather surprising, however, that information theoretic frameworks are not more explicit in the InfoVis/VA literature. For one thing, complexity reduction is a major design requirement in visual analytics: of *course* information visualizations and visual analytics systems should enable users to boil down information to the most relevant attributes, parameters, and content. In fact, most InfoVis/VA design guidance can be interpreted to support user-driven complexity reduction in one way or another.

Going forward, we suggest that our colleagues consider how emphasizing complexity reduction as an explicit design goal can facilitate evaluation methodologies that emphasize process measures associated with effective sensemaking:

Design. When working with user communities to understand their workflows, it makes sense to pay attention to the strategies, heuristics, and representations that people use to get rid of irrelevant information. Paying attention to these native strategies and heuristics might prove very useful in developing automated techniques that enable people to efficiently ringfence irrelevant information. After all, getting rid of junk can be a quite desirable attribute of a visual analytics tool; we are thinking here of the ForceSPIRE users who were surprised and gratified to find that Endert et. Al.'s semantic interaction models resulted in the right items being relegated to the junk pile [20].

Evaluation. Information theoretic measures are probably not the golden ticket for measuring insights. However, if one accepts the principle that complexity reduction is a necessary element in the sensemaking processes that support insight, then it makes sense for us to evaluate visualizations and analytic workflows in terms of how well they enable people to reduce complexity. The subjectivity and idiosyncrasy of analytic workflows is one of the major barriers to developing generalizable, precise, and ecologically valid indicators to evaluate the goodness of a visual analytics workflow. However, everyone needs to get rid of the items that are irrelevant to a problem. Meanwhile, information theoretic measures are generalizable across users, problems, and visual analytic workflows, precisely because such are explicitly *not* about semantics or meaning (thank you, Claude Shannon).

5.1 Is There a 'There,' There?

Recognizing the importance of empirically testing one's ideas (or, as we say in the United States, eating one's own dog food), we are currently developing a dataset that will help us characterize the performance of information theoretic measures in analytic processes. In summer 2016, we are running an experimental study in which we present participants with a hypothetical intelligence analysis question and a rather obnoxious pile of references.

The information theoretic measure we have selected for testing is a normalized information compression distance (abbreviated as NCD), which is a feature- and parameter-free way of evaluating the similarity of two or more information objects [53]. The simplicity and generalizability of such compression-based measures lend them to a wide range of applications, and we are intrigued by the idea of using them as a process measure of information interaction in visual analytic systems. Specifically, we suggest that an effective visual analytic interaction model should enable people to sort data and information into categories of semantically similar items. The emergence of within-category semantic coherence should be expressible as a change in compression distance, both within categories and across the document set as a whole.

To evaluate this idea, we have designed a counterbalanced, between-subjects experiment involving 15-18 research participants drawn from the Sandia National Laboratories' professional population. Volunteers will be asked to assess evidence supporting the British government's assertion that Russian assassins used polonium to kill Alexander Litvenenko in London, November 2006. They will work with a set of 370 documents, ranging from dense journal articles, to government reports, to articles from popular news magazines. A small number of the documents are directly relevant to the challenge questions; many are tangentially related but do not contain the information required to assess the challenge questions; and a plurality have nothing to do with the topic (distractors). To perform the task, particpants will be randomly assigned to one of three different workflows, or conditions. Two of the conditions use the Sandia National Laboratories' Citrus document analysis platform, while a third "control" condition requires participants to perform the task using the Windows filing system.

We hypothesize that research participants across all experimental conditions will arrange the documents in a manner that increases within-category similarity; at the same time, we expect between-category similarity to fall as people create "patches" of semantically related information. We also expect that automated support for categorization will facilitate users' efforts to impose structure on the chaos we give them. Specifically, when using Citrus, we expect research participants will organize documents into categories more quickly and efficiently, with greater reductions in NCD compared to performance of the same task using the Windows file system on a computer desktop.

For each experimental condition, we will collect both process and outcome measures.

- *Process measures:* Total number of sorting iterations, time per iteration, subjective workload (measured using the NASA Task Load Index, or TLX; see nasatlx.com) at the end of the work session; and a normalized compression distance measure applied to the categories at the end of each sorting iteration.
- *Outcome measures:* End-of-task NASA-TLX, time-on-task, NCD per category created; difference between starting and ending categories, and the number of correct answers to challenge questions about the Litvinenko event.

We will be completing data collection and analyzing the resulting data in the summer and fall of 2016. Our follow-on publication will document our experience applying compression distance measures to assess how these computer-mediated interaction models facilitate the reduction of complexity in a realistic analytic workflow.

ACKNOWLEDGMENTS

This research was funded by Sandia National Laboratories' Laboratory Directed Research and Development Program. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94-AL85000.

REFERENCES

[1]
Amar, R.A., Eagan, J., and Stasko, J., 2005. Low-level components of analytic activity in information visualization. In *Proceedings of IEEE Symposium on Information Visualization* (23-25 October 2005), 111-117.

[2] Amar, R.A. and Stasko, J., 2004. A knowledge task-based framework for the design and evaluation of information visualizations. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)* (2004), 143-149.

[3] Arias-Hernandez, R., Kaastra, L.T., Green, T.M., Fisher, B., 2011. Pair analytics: capturing reasoning processes in collaborative visual analytics. In Proceedings of System *Sciences (HICSS), 2011 44th Hawaii International Conference on* (2011) 1-10.

[4] Bauer, T., 2009. *Information and meaning: Revisiting Shannon's theory of communication and extending it to address today's technical problems.* Albuquerque, NM: Sandia National Laboratories, *prod.sandia.gov/techlib/access-control.cgi/2009/098168.pdf*

[5] Bergen, B., 2012. *Louder Than Words: The New Science of How The Mind Makes Meaning*. Basic Books, New York.

[6] Bradel, L., Endert, A., Koch, K., Andrews, C., and North, C., 2013. Large high resolution displays for co-located collaborative sensemaking: display usage and territoriality. *International Journal of Human-Computer Studies 71* (Nov) 1078-1088.

[7] Bruner, J.S., 1978. Acquiring the uses of language. *Canadian Journal of Psychology-Revue Canadienne De Psychologie 32*, 4 (1978), 204-218.

[8] Bruner, J.S. and Goodman, C.C., 1947. Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology 42*, 1 (1947), 33-44.

[9] Bruner, J.S. and Tajfel, H., 1965. Width of category and concept differentiation: A note on some comments by Gardner and Schoen. *Journal of Personality and Social Psychology 2* (1965-Aug), 261-264.

[10] Card, S.K., 2008. Information visualization. In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*, A. Sears and J.A. Jacko, eds. Lawrence Earlbaum and Associates, Mahwah, NJ, 510-542.

[11] Card, S.K., Mackinlay, J., and Shneiderman, B., 1999. *Readings In Information Visualization: Using Vision To Think* Academic Press, San Diego, CA.

[12] Carpendale, S., 2008. *Evaluating Information Visualizations*. Springer-Verlag, Berlin-Heidelberg.

[13] Chang, R., Ziemkiewicz, C., Green, T.M., and Ribarsky, W., 2009. Defining insight for visual analytics. *IEEE Computer Graphics and Applications 29*, 2, 14-17.

[14] Chen, M. and Jaenicke, H., 2010. An information theoretic framework for visualization. IEEE *Transactions on Visualization and Computer Graphics 16*, 6, 1206-1215.

[15] Chi, E.H., Pirolli, P., Chen, K., and Pitkow, J., 2001 Using information scent to model user information needs and actions on the web. In *Proceedings of SIGCHI* (Seattle, WA, March 31-April 4, 2001), ACM.

[16] Colbaugh, R., Glass, K., and Bauer, T., 2013. Dynamic information-theoretic measures for security informatics. In *Proceedings of the ISI 2013* (Seattle, WA 2013), IEEE.

[17] Dervin, B., 1999. Chaos, Order and sensemaking: a proposed theory for information design. *Information Design*, 33-57.

[18] Dillon, A., 2002. Beyond usability: process, outcome and affect in human-computer interactions. *Canadian Journal of Library and Information Science*, August 2002.

[19] Dornburg, C.C., Matzen, L.E., Bauer, T.L., and McNamara, L.A., 2009. Working memory load as a novel tool for evaluating visual analytics. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST),* Atlantic City, NJ, 2009, 214-215

[20] Endert, A., Fiaux, P., and North, C., 2011. Unifying the sensemaking loop with semantic interaction. In *Proceedings of The IEEE VisWeek Workshop on Interactive Visual Text Anlaytics for Decision Making* (Providence, RI 2011), IEEE.

[21] Fisher, B., Green, T.M., and Arias-Hernandez, R., 2011. Visual analytics as a translational cognitive science. *Topics in Cognitive Science 3*, 3 (Jul), 609-625.

[22] Gleick, J., 2011. *The Information: A History, A Theory, A Flood*. Random House, New York.

[23] Green, T.M., Arias-Hernandez, R., and Fisher, B., 2014. Individual differences and translational science in the design of human-centered visualizations. Huang, W., ed. *Handbook of Human Centric Visualization*, 93-113.

[24] Green, T.M. and Fisher, B., 2012. Using translational science in visual analytics. *2012 IEEE Conference On Visual Analytics Science and Technology (VAST)*, Seattle, WA, 221-222.

[25] Green, T.M., Wakkary, R., and Arias-Hernandez, R., 2011.. Expanding the scope: interaction design perspectives for visual analytics. In *Proc System Sciences (HICSS), 2011 44th Hawaii International Conference on*, (Kauai, HI, 2011), 1-10.

[26] Heer, J. and Agrawala, M., 2006. Software design patterns for information visualization. *Visualization and Computer Graphics, IEEE Transactions on 12*, 5, 853-860.

[27] Heer, J. and Agrawala, M., 2007. *Design Considerations For Collaborative Visual Analytics* Information Visualization 7, 1 (Mar) 49-62.

[28] Johnson, S., 2004. *Mind Wide Open: Your Brain and*

The Neuroscience of Everyday Life. Scribner, New York.

[29] Kang, Y.-A., Goerg, C., and Stasko, J., 2011. How can visual analytics assist investigative analysis? Design implications from an evaluation. *IEEE Transactions on Visualization and Computer Graphics 17*, 5 (May), 570-583.

[30] Kang, Y.-A. and Stasko, J., 2012. Examining the use of a visual analytics system for sensemaking tasks: case studies with domain experts. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (Dec), 2869-2878.

[31] Klein, G., Moon, B., and Hoffman, R.R., 2006. Making sense of sensemaking 2: a macrocognitive model. *IEEE Intelligent Systems 21*, 5 (Sep-Oct), 88-92.

[32] Kounios, J. and Beeman, M., 2009. The Aha! moment: the cognitive neuroscience of insight. *Current Directions In Psychological Science 18*, 4, 210-216.

[33] Lam, H., 2008. A framework of interaction costs in information visualization. *Visualization and Computer Graphics, IEEE Transactions on 14*, 6, 1149-1156.

[34] Lam, H., Bertini, E., Isenberg, P., Plaisant, C., and Carpendale, S., 2012. Empirical studies in information visualization: Seven scenarios. *Visualization and Computer Graphics, IEEE Transactions on 18*, 9, 1520-1536.

[35] Munzner, T., 2009. A nested model for visualization design and validation. *Visualization and Computer Graphics, IEEE Transactions on 15*, 6, 921-928.

[36] North, C., 2006. Toward measuring visualization insight. *IEEE Computer Graphics and Applications 26*, 3, 6-9.

[37] North, C., Saraiya, P., and Duca, K., 2011. A comparison of benchmark task and insight evaluation methods for information visualization. *Information Visualization 10*, 3 (Jul), 162-181.

[38] Panizza, D., 2012. Formal neurosemantics: Logic, meaning and composition in the brain. *Journal of Neurolinguistics 25*, 5, 460-488.

[39] Pike, W.A., Stasko, J., Chang, R., and O'Connell, T.A., 2009. The science of interaction. *Information Visualization 8*, 4, 263-274.

[40] Pirolli, P., 2007. *Information Foraging*. Oxford University Press, Oxford, UK.

[41] Pirolli, P. and Card, S.K., 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. *Proceedings of International Conference on Intelligence Analysis*, pp 2-4.

[42] Pirolli, P. and Russell, D.M., 2011. Introduction to this special issue on sensemaking. *Human-Computer Interaction*, 26, 1-8.

[43] Plaisant, C., Grinstein, G., Scholtz, J., Whiting, M., O'Connell, T., Laskowski, S., Chien, L., Tat, A., Wright, W., Gorg, C., Liu, Z., Parekh, N., Singhal, K., and Stasko, J., 2008. Evaluating visual analytics at the 2007 VAST symposium contest. *IEEE Computer Graphics and Applications 28*, 2 (Mar-Apr), 12-21.

[44] Pulvermüller, F., 2012. Meaning and the brain: The neurosemantics of referential, interactive, and combinatorial knowledge. *Journal of Neurolinguistics 25*, 5 (Sept), 423-459.

[45] Saraiya, P., North, C., Lam, V., and Duca, K.A., 2006. An insight-based longitudinal study of visual analytics. *Visualization and Computer Graphics, IEEE Transactions on 12*, 6, 1511-1522.

[46] Sedlmair, M., Meyer, M., and Munzner, T., 2012. Design study methodology: Reflections from the trenches and the stacks. *Visualization and Computer Graphics, IEEE Transactions on 18*, 12, 2431-2440.

[47] Shneiderman, B., 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of The Visual Languages, 1996. Proceedings., IEEE Symposium On* (Boulder, Co1996), IEEE.

[48] Shneiderman, B., 2002. Inventing discovery tools: Combining information visualization with data mining. *Information Visualization 1*, 1 (Mar), 5-12.

[49] Shneiderman, B. and Plaisant, C., 2006. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI Workshop Beyond Time and Errors: Novel Evaluation Methods for Information Visualization*, ACM, 1-7.

[50] Stasko, J., Gorg, C., and Liu, Z., 2008. Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization 7*, 118-132.

[51] Sternberg, R.J. and Davidson, J.E., 1996. *The Nature of Insight*. MIT Press, Cambridge, MA.

[52] Weick, K., 1995. *Sensemaking in Organizations*. Sage, Thousand Oaks, CA.

[53] Vitanyi, P.M.B., Balbach, F.J., Cilibrasi, R.L. and Li, M. Normalized Information Distance. In Emmert-Streib, F. and Dehmer, M., eds., *Information Theory and Statistical Learning*. Springer US, 45-82.

DISTRIBUTION

1	MS0899	Technical Library	9536 (electronic copy)
1	MS0359	D. Chavez, LDRD Office	1911

Sandia National Laboratories