Pathways for Theoretical Advances in Visualization

Min Chen
University of Oxford

Jessie Kennedy
Edinburgh Napier University

Georges Grinstein
University of Massachusetts

Melanie Tory
Tableau Software

Chris R. Johnson
University of Utah

More than a decade ago, Chris Johnson proposed the "Theory of Visualization" as one of the top research problems in visualization.[1] Since then, there have been several theory-focused events, including three workshops and three panels at IEEE Visualization (VIS) Conferences. Together, these events have produced a set of convincing arguments:

- As in all scientific and scholarly subjects, theoretical development in visualization is a necessary and integral part of the progression of the subject itself.
- Theoretical developments in visualization can draw on theoretical advances in many disciplines, including, for example, mathematics, computer science, engineering science, psychology, neuroscience, and the social sciences.
- Visualization holds a distinctive position, connecting human-centric processes (such as human perception, cognition, interaction, and communication) with machine-centric processes (such as statistics, algorithms, and machine learning). It therefore provides a unique platform to conduct theoretical studies that may impact on other disciplines.
- Compared with many mature disciplines (such as mathematics, physics, biology, psychology, and philosophy), theoretical research activities in visualization are sparse. The subject can therefore benefit from a significantly increased effort to make new theoretical advances.

Visualization theory is commonly viewed as a research focus for only a few individual researchers. Its outcomes, perhaps in the form of theorems or laws, are perceived as too distant from practice to be useful. Perhaps inspired by well-known theoretical breakthroughs in the history of science, visualization researchers may unconsciously have high expectations for the originality, rigor, and significance of the theoretical advancements to be made in a research project or presented in a research paper.

On the contrary, although textbooks tend to attribute major breakthroughs to single pioneers at specific times and places, in most cases, such breakthroughs generally take years or even decades and are usually the product of numerous incremental developments, including a substantial number of erroneous solutions suggested by the pioneers themselves as well as many lesser-known individuals. Many complex discoveries did not initially appear to have elegant proofs, and it has taken some challenging and often questionable steps to obtain the well-formulated solutions we find in modern textbooks. For example, most readers associate the theory of general relativity with Albert Einstein's November 1915 discovery in Berlin. According to Petro Ferreira,[2] Einstein first speculated about the generalization in 1907. He then published two papers with Marcel Grossmann (Zurich) in 1913 that sketched out the theory and worked with David Hilbert (Göttingen) on the problem in June 1915. Some of the most important discoveries related to the theory of general relativity are

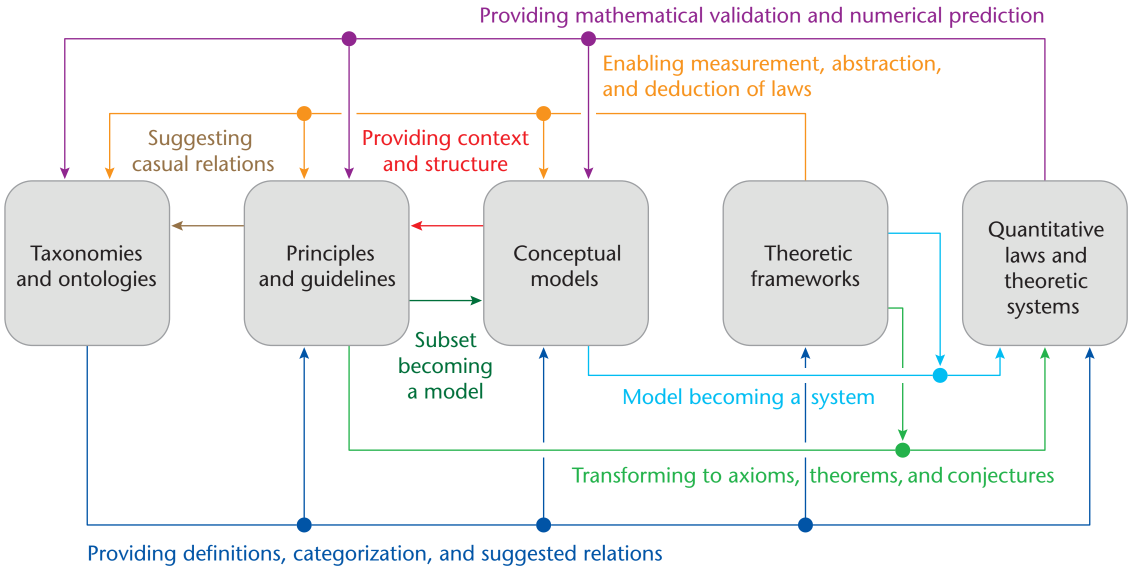

Figure 1. A theoretical foundation typically evolves through iterative developments. The development of each aspect both influences and benefits from that of others. A successful transformation between different aspects indicates a theoretical enhancement of understanding. Note that the third aspect, "Conceptual Models and Theoretic Frameworks," is represented by two boxes in the figure.

Mercury's perihelion shift (Le Verrier, 1859), the 1919 eclipse expedition (Edington, Cottingham, Crommelin, and Davidson), the evolving universe (Fredmann, 1922; Lemaître, 1927), the expanding universe (Slipper, 1915; Lundmark, 1924; Hubble and Humason, 1929), the big bang (Lemaître, 1931), and the black hole (Schwarzschild 1916, Chandrasekhar, 1935; Landau, 1938; Oppenheimer and his students, 1939). Some ill-fated solutions also followed Einstein's 1915 discovery, the most notable of which were perhaps the static universe (Einstein and de Sitter) and the suspended universe (Eddington).

During an Alan Turing Institute event in London in April 2016 on the theoretical foundation of visual analytics, discussions on the need to build such a theoretical foundation varied greatly, with opinions ranging from "Visualization should not be physics-envy" to "It is irresponsible for academics not to try." After two days of presentations, discussions, and debates, the attendees gradually converged on a common understanding that a theoretical foundation consists of several aspects (see the "Major Aspects of a Theoretical Foundation" sidebar for more details) and that every visualization researcher should be able to make direct contributions to some aspect of the theoretical foundation of visualization.

During the IEEE VIS Conference in Baltimore, Maryland, in October 2016, a discussion panel took this viewpoint further by outlining avenues for pursuing theoretical research in each aspect. This article is a structured reflection by the panelists about the discussions during that IEEE VIS 2016 panel. In this article, we first review four major aspects of a theoretical foundation and then discuss the interactions and transformations between them. Figure 1 provides an overview of the discourse in this article.

Taxonomies and Ontologies

For millennia, humans have been classifying things in the world around them by describing and naming concepts to facilitate communication (see the "History of Taxonomy and Ontology" sidebar). The most significant and enduring effort is the classification of life on Earth, which commenced in Aristotle's time, became mainstream through the work of Linnaeus, and continues to this day with new species being identified and alternative classifications of existing species being proposed.[3] Alternative classifications (*taxonomies*) arise over time as a result of differing opinions about the importance of differentiating characteristics used in creating the concepts (*taxa*). These differing opinions are usually the result of new information becoming available, often through technological advances, which can result in the same organism being classified according to different taxonomic opinions and subsequently having several alternative names, which may in turn lead to miscommunication. Newer classifications are usually improvements on previous ones, but sometimes the existence of alternative classifications reflects a disagreement as to how to interpret the data on which the classification is based.

Ontologies are representations of different relationships among various concepts. Naturally, they are built on the taxonomic classification of the concepts of both entities and relationships. Taxonomies and ontologies are means of conceptualizing, understanding, organizing, and reasoning about these entities and relationships. They are central to communicating about the world around us. They play an increasing role in understanding in the visualization field, allowing us to organize and formalize our knowledge.

A brief review of the literature over the past three decades reveals at least 70 publications containing some form of visualization taxonomy.[3] Three questions are relevant when considering visualization taxonomies: What is being classified (domain)? Why is the taxonomy being developed (purpose)? How is the taxonomy constructed (process)? Taxonomies have been proposed to classify many aspects of visualization, including systems, tools, techniques, interaction approaches, data types, user tasks, visual encodings, input methods, and evaluation strategies. These aspects can be classified according to different criteria. For example, visualization techniques can be classified by the analytical tasks they support, the visual encoding or algorithm used, the data type, or the domain in which they are employed.

The visualization community has found taxonomies useful in their research. Taxonomies offer a shared vocabulary with which we can communicate effectively and reduce misunderstanding.[4] They orientate us among the vast number of techniques and tools that have already been developed, often across disparate domains. Taxonomies are therefore frequently adopted in literature surveys to categorize existing work. Furthermore, using taxonomies as design spaces can reveal novel research opportunities, for example, by conducting gap analysis.

In comparison, the term "ontology" appears much less frequently in the visualization literature. This is partly because some studies on ontological relationships are presented as qualitative models. Because ontologies are typically described in ontology languages, such as OWL (Web Ontology Language) and RDF (Resource Description Framework), they can be used by algorithms in visualization systems. For example, ontologies can be used to generate annotations and filter or highlight visual objects automatically in visualization, to enable the automated creation of visualization, and to integrate keyword search and visual exploration in a user interface.[5]

For designers, taxonomies and ontologies play a role in systemizing the design process and can be employed at multiple stages, such as domain characterization and abstraction, selection of appropriate visual encodings and interaction techniques, and formulation of data and information flows. In addition, taxonomies and ontologies provide the basis for studying causal relationships, thereby facilitating the development of guidelines and qualitative models.

Building taxonomies and ontologies is an investigative science because they often feature partial and evolving hypotheses. A number of considerations therefore arise during the process, including

Major Aspects of a Theoretical Foundation

Attendees at an Alan Turing Institute event in London in April 2016 on the theoretical foundation of visual analytics reached a consensus that a theoretical foundation consists of the following aspects.

Taxonomies and Ontologies

In scientific and scholarly disciplines, a collection of concepts are commonly organized into a taxonomy or ontology. In the former, concepts are known as *taxa* and are typically arranged hierarchically using a tree structure. In the latter, *concepts*, often in conjunction with their instances, attributes, and other entities, are organized into a schematic network, where edges represent various relations and rules.

Principles and Guidelines

A *principle* is a law or rule that must be followed and is usually expressed in a qualitative description. A *guideline* describes a process or a set of actions that may lead to a desired outcome or, alternatively, actions to be avoided to prevent an undesired outcome. The former usually implies a confidence in the high degree of generality and certainty of the causality concerned, whereas the latter suggests that a causal relation may be subject to specific conditions.

Conceptual Models and Theoretic Frameworks

The terms *models* and *frameworks* have broad interpretations. Here we consider that a *conceptual model* is an abstract representation of a real-world phenomenon, process, or system, featuring different functional components and their interactions. A *theoretic framework* provides a collection of measurements and basic operators and functions for working with these measurements. The former provides a tentative description of complex causal relations in the real world, and the latter provides a basis for evaluating different models quantitatively.

Quantitative Laws and Theoretic Systems

A *quantitative law* describes a causal relation of concepts using a set of measurements and a computable function confirmed under a theoretic framework. Under a *theoretic framework*, a conceptual model can be transformed into a theoretic system through axioms (postulated quantitative principles) and theorems (confirmed quantitative laws). Unconfirmed guidelines are thus conjectures, and contradictory guidelines are paradoxes.

History of Taxonomy and Ontology

The term *taxonomy* comes from the Greek word *taxis* (meaning "order" or "arrangement") and the suffix *-nomos* (meaning "law" or "science").[1] Plato was among the first to formulate methods for grouping objects based on their similar properties. Aristotle wrote *Categories*, which provides an in-depth study of classes and objects.

Naming and classifying plants and animals dates back to the origin of human languages. The development of modern botanical and zoological taxonomy is often attributed to Carl Linnaeus (1707–1778), a Swedish botanist, who defined many of the rules that taxonomists use today. The development of taxonomy in biology facilitated the paradigm shift in the 19th century when the theory of evolution was proposed.

The automatic construction of a hierarchical categorization scheme began in the 1960s with applications such as decision-tree based classification, computational phylogenetics, and topic analysis in text mining.

The term *ontology* comes from the Greek prefix *onto-* (meaning "being" or "that which is") and suffix *-logia* (meaning "logical discourse," "study," or "theory").[2]

The term *ontologia* first appeared in the works by German philosophers Jacob Lorhard (1606) and Rudolf Göckel (1613). It refers to the philosophical study of the concept of "being" and its variants—for example, "becoming," "existence," and "reality" as well as the categorization of the concept and the relationships between different categories. Taxonomy is often viewed as a subset of ontology, which primarily considers the grouping relationships. Ontology can be seen as a generalization of taxonomy by allowing for different types of relationships among different entities.

An ontology is a form of knowledge representation,[3] where entities are defined with names, types, properties, and different relationships with other entities. Its applications in computer science include artificial intelligence, the Semantic Web, biomedical informatics, library science, systems engineering, software engineering, and many more. The methodology has also been used in visualization.[4]

References

1. P.F. Stevens, *The Development of Biological Systematics*, Columbia Univ. Press, 1994.
2. J.F. Mora, "On the Early History of 'Ontology,'" *Philosophy and Phenomenological Research*, vol. 24, no. 1 1963, pp. 36–47.
3. J.F. Sowa, *Conceptual Structures: Information Processing in Mind and Machine*, Addison Wesley, 1984.
4. O. Gilson, et al., "From Web Data to Visualization via Ontology Mapping," *Computer Graphics Forum*, vol. 27, no. 3, 2008, pp. 959–966.

determining the subpopulation to study; identifying the characteristics used to define a class, a relation, or the level of specificity; comparing the importance of different characteristics; differentiating among various terms used for specifying characteristics; selecting the effective visualization techniques for visualizing large taxonomies and ontologies; automatically generating a taxonomy or ontology from text analysis of visualization literature; and automatically evolving a taxonomy or ontology using machine learning.

Taxonomies and ontologies are fundamental tools that help with understanding, communication, and development in the visualization field. Still, a number of challenges and open questions remain: Can we define a methodology for creating, comparing, and integrating taxonomies and ontologies? At what levels and granularity should taxonomies or ontologies be specified? How do we select one or more taxonomies (or ontologies) for our work? The visualization field continues to change, so taxonomies and ontologies must evolve as well. We must continue to improve their construction and use.

Principles and Guidelines

A *guideline* embodies a wisdom advising a sound practice. This may be a course of action to take or to avoid in achieving a goal. Guidelines are commonly outlined based on accumulated experience and knowledge about some causal relations in a process. It takes courage and conviction to propose a new guideline. And it takes a lot more courage and fair-mindedness to accept critiques about the proposed guideline and then retract or refine it.

Some guidelines stand the test of time and become principles. Many others are effective in only specific circumstances. Because of the qualitative nature of framing guidelines and the typically self-directed mechanism for creating and evolving guidelines, now and then some may be defined without rigorous care, generalized beyond their intended application, become out of date, or conflict with other guidelines. Many documents about guidelines often contain a disclaimer: "By definition, following a guideline is never mandatory. Guidelines are not binding and are not enforced."[6]

In many disciplines, such as biology and medicine, guidelines have played an indispensable role and are rigorously evaluated, critiqued, and maintained. In other disciplines, such as physics, chemistry, and engineering, old wisdoms have gradually been transformed into quantitative laws and quantitative process management. In the visualization field, guidelines have no doubt played a positive role in designing and developing visualization systems as well as in education. (See the "Examples of Visualization Guidelines" sidebar for more details.) For example, Miriah Meyer and her colleagues considered guidelines to be an integral part of an agile process for developing visual designs and visualization systems, helping designers make choices.[7] Torre Zuk and his colleagues argued that guidelines can be used as heuristics for evaluating visual designs and visualization systems.[8]

These recommendations inevitably place a huge burden on the correctness and effectiveness of guidelines. If visualization guidelines are going to play a pivotal role, as these researchers suggested,[7,8] we will need to take several steps:

- develop mechanisms for curating, evaluating, critiquing, and refining guidelines in an open and transparent manner;
- establish a culture of open, democratic, evidence-based discourse on the guidelines and enable broader participation in the discourse beyond the current scale of a few papers and blogs; and
- inspire researchers to study guidelines, including their evolution and applicability in different conditions using scientific methods, and when appropriate opportunities arise, transform guidelines into quantitative laws and process management.

Social scientists have established research methods for collecting and analyzing qualitative data in order to infer concrete theoretical insights, which include taxonomies, ontologies, guidelines, and conceptual models. One such method is *grounded theory*.[9] It involves observing practical phenomena in the wild (to ground the theory in real-world data), identifying categories of the instances (events, processes, occurrences, participants, and so on), making links between categories, and establishing relationships between them. The method utilizes descriptive labeling (referred to as coding) to conceptualize discrete instances of phenomena systematically. It advocates continuous comparative analysis and negative case analysis to ensure the coding is comprehensive, meticulous, and up to date. It encourages researchers to interact with data by asking questions, broadening the sampling space by exploring related phenomena, and writing memos.

By enabling categorization and relationship discovery, the grounded theory method can facilitate the development of visualization taxonomies and ontologies by supporting the analysis of causal relationships. By pursuing both positive and negative case studies and undertaking continuous comparative analysis, we facilitate the evaluation, critique, revision, and improvement of guidelines. By enabling the curation of a relatively complete and coherent set of causal relationships functioning in a system, we help establish a conceptual model.

Examples of Visualization Guidelines

In the visualization field, various books, research papers, and online media recommend several hundred different guidelines. Here we offer just a subset of examples:

- **Maximize the data-ink ratio:** E.R. Tufte, *The Visual Display of Quantitative Information*, Graphics Press, 1983, p. 93.
- **Overview first, zoom and filter, then details on demand:** B. Shneiderman, "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations," *Proc. IEEE Symp. Visual Languages*, 1996, pp. 336–343.
- **Rainbow color map guidelines:** B.E. Rogowitz and L.A. Treinish, "Data Visualization: The End of the Rainbow," *IEEE Spectrum*, vol. 35, no. 12, 1998, pp. 52–59, and D. Borland and R.M. Taylor II, "Rainbow Color Map (Still) Considered Harmful," *IEEE Computer Graphics & Applications*, vol. 27, no. 2, 2007, pp. 14–17.
- **10 guidelines for data visualization:** C. Kelleher and T. Wagener, "Ten Guidelines for Effective Data Visualization in Scientific Publications," *Environmental Modelling & Software*, vol. 26, no. 6, 2011, pp. 822–827.
- **14 guidelines for data visualization:** schoolofdata.org/2013/04/26/data-visualization-guidelines-by-gregor-aisch-international-journalism-festival/.
- **Six guidelines for creative visualization:** www.tut.com/article/details/12-6-guidelines-for-creative-visualization/?articleId=12.

Conceptual Models and Theoretic Frameworks

A conceptual model can be a representation of an idea, process, or system. It is typically used to describe and explain the causal relationships exhibited in phenomena in a physical, biological, economic, social, or any other type of system that may be intuitively observable, cannot be experienced directly, or may be totally hypothesized.

The descriptions of many models are accompanied by visual representations that help link conceptualization with observation. The physicist Richard Feynman created new visual abstractions of the physics and mathematics of quantum electrodynamics so that he could more easily reason about the complex mathematics.[10] Feynman famously had his van painted with his illustration of the interactions of subatomic particles (see Figure 2).

Figure 2. Richard Feynman's 1975 Dodge van. Feynman had the behavior model of subatomic particles painted on the sides. (Courtesy of ArtCenter College of Design.)

In most disciplines, model development has been a driving force for progression. It fuels and guides the advancement of a subject by enabling abstraction, proposition, prediction, and validation (using experimentation, mathematics, and computation). Models are central to what researchers do, both in their research and when communicating their explanations. The development of the standard model in particle physics was a collective effort of scientists around the world throughout the second half of the 20th century. In the same way, the discovery of the double helix model of DNA was an iterative research endeavor in the early 1950s. Many intermediate steps, ranging from the partial model alpha helix and the incorrect triple helix model by Linus Pauling to x-ray diffraction experiments by Rosalind Franklin and others, paved the way for James Watson and Francis Crick to formulate the landmark model in biology.

In the visualization field, researchers have proposed more than a dozen conceptual models for describing the relationships among data, visualization systems, analytical techniques, interaction methods, human perception and cognition, user tasks, and application contexts.[3] The goal of such models is to help us describe, understand, reason about, and predict what people can do in a visualization process and environment, which actions might lead to which results in given circumstances, and which workflow is more efficient or effective than others.

For example, a personal visualization model for fitness tracker data[11] helped explain why the on-calendar visualization approach was more effective than a traditional fitness feedback tool, and more importantly, it provided a theoretical basis from which general design guidelines for behavior feedback tools can be derived. Another example is a human cognition model for visualization.[12] Based on human ergonomics and cognitive psychology, the model defines human leverage points, where cognitive experiments can be conducted for quantitative and qualitative evaluation of visualizations. Similarly, sense-making models have played an important role in supporting the design of interactive analysis tools.[13]

Hence, building correct and effective conceptual models for visualization must be an endeavor on the part of the visualization community. Learning from other disciplines, we must significantly increase our efforts in experimentation, theorization, and computational simulation and validation.

Experimentation and Qualitative Theorization

The visualization literature includes more than 40 empirical studies for studying human perception and cognition in visualization as well as more than 40 others for comparing different visualization techniques. In addition, through numerous application case studies, visualization researchers have had firsthand experience observing a variety of data, visualization systems, analytical techniques, interaction methods, human perception and cognition, user tasks, and application contexts in the wild.

These empirical studies and application case studies provide opportunities to formulate new models, perform continuous comparative analysis, probe negative experience, critique and improve existing models, broaden theoretical sampling, and explore model unification and theoretical saturation, all of which are advocated by the grounded theory methodology we mentioned earlier. Rigorously building and analyzing qualitative models will inevitably motivate further theorization through the development of quantitative models.

Quantitative Theorization

In many applications, especially in the physical sciences, models are often formulated using a particular mathematical framework. For example, in physics, Newton invented calculus (also credited to Leibniz) to underpin classic mechanics. Einstein used Riemannian geometry to underpin his general theory of relativity. Today, we commonly see publications entitled mathematical framework *X* for model *Y*. In some situations, a model *Y* may itself have evolved into an elegant mathematical framework that can be used to underpin other models. For example, information theory, which is underpinned by probability theory, has become a fundamental framework for telecommunication, data communication, data compression, and data encryption.

Several mathematical frameworks have been proposed for underpinning quantitative theorization in visualization, including information theory[14] and algebra.[15] Naturally, we hope that some qualitative models in the visualization literature can be described using such a framework with quantitative measurements, which may not be quite accurate initially. Lack of accuracy does not always mean wrong, however. We must remember that Newton's first law of motion could not be fully validated until the technology for creating the conditions for a vacuum became available. Having errors is not always unhelpful. We must remember that the discrepancy between the prediction of Newtonian gravity and the observed orbit of Mercury inspired the discovery of the theory of general relativity.

Computational Simulation and Validation

In most disciplines where visualization techniques are routinely deployed, together with experimentation and theorization, computational science now constitutes the "third pillar" of scientific inquiry, enabling researchers to build and test models of complex phenomena. Advances in computing and connectivity make it possible to capture, analyze, and develop computational models for unprecedented amounts of experimental, observational, and simulation data to address problems previously deemed intractable or beyond imagination.[16]

Once we have quantitative models of visualization phenomena and processes, we can simulate such models computationally, validating them against experimental results and making predictions about causal relations in a visualization process. For example, we can model the relationships among volume datasets, volume-rendering algorithms, and resultant imagery data. The model can be used to predict the discretization errors, order of accuracy, and convergence performance as well as verify if they meet the requirements of the application concerned.[17] The cognition literature shows that human observers' perception errors may not linearly correlate with discretization errors, so it would be exciting to extend such a model to include more elements of human perception and cognition.

Quantitative Laws and Theoretic Systems

In all branches of science, many quantitative laws are regarded as disruptive discoveries because they represent great leaps in our understanding about causal relationships from numerical uncertainty to numerical certainty.[18] As we discussed earlier, any visualization guideline that has stood the test of time should be regarded as a principle. Furthermore, any principle in visualization can be formulated and proved under a theoretical framework. (See the "An Example Theoretic System: Probability Theory" sidebar for an example.) For example, part of Ben Shneiderman's guideline "overview first, zoom, then details on demand" was proved using information theory (including an anomaly investigation).[14] The filtering part of the guideline likely requires a more complex proof because defining the filtering that would result in desired details might require additional variables.

In many disciplines, some laws have parameters that may be constants. The discovery of such fundamental constants (such as the speed of light and absolute zero temperature) transforms postulated laws into truly quantitative laws. Discovering values that would fit such parameters often requires extensive experimentation. For example, in psychology, Fitts's law has two parameters that vary according to the choice of input device, and Stevens' law also has two parameters that vary according to the choice of physical stimulus. These parameters suggest that a more general quantitative law may be hidden underneath. For example, if Newton's second law of motion had used volume instead of mass, it would have required an object-dependent parameter that we now know as density. Worse, if it were surface area instead of mass, one would need more object-dependent parameters.

Computational science now constitutes the "third pillar" of scientific inquiry, enabling researchers to build and test models of complex phenomena.

The visualization discipline provides great opportunities for postulating parameterized laws and for discovering values for such parameters in different scenarios. From such discoveries, we could potentially make more fundamental leaps in our understanding as long as we continue to investigate the causes of the unattractive parameterization.

When a number of quantitative laws share a common measure space that includes all variables to be measured and all measurement functions, they indicate the existence of a theoretical system, where new quantitative laws can be inferred from

An Example Theoretic System: Probability Theory

$(\Omega, \mathbf{E}, P)$ is a measure space, where Ω is the sample space, $\mathbf{E}$ is the event space, and $P(e)$ is the probability measure of an event $e \in \mathbf{E}$.

Axioms

1. The probability of an event is a nonnegative real number: $P(e) \in \mathbf{R}$, $P(e) \geq 0$, $\forall e \in \mathbf{E}$.
2. The probability that at least one of the elementary events in the entire sample space will occur is 1: $P(\Omega) = 1$.
3. Any countable sequence of mutually exclusive events $(e_1, e_2, \ldots)$ satisfies the following: $P\left(\bigcup_{i=1}^{\infty} e_i\right) = \sum_{i=1}^{\infty} P(e_i)$.

Example Law: Monotonicity

If $\mathbf{E}_A$ is a subset of or equal to $\mathbf{E}_B$, then the probability of $\mathbf{E}_A$ is less than or equal to the probability of $\mathbf{E}_B$. That is, if $\mathbf{E}_A \subseteq \mathbf{E}_B$, then $P(\mathbf{E}_A) \leq P(\mathbf{E}_B)$.

A Skeleton of a Theoretic System for Visualization

(Ω, Θ, Ξ) is a measure space, where Ω is the sample space, Θ is a state space defined by a subset of all possible alphabets in visualization (such as data (**D**), task (**T**), medium (**M**), visual representation (**V**), human capability (**H**), and interaction (**I**)), and Ξ is a subset of all possible measures in visualization (such as probability, mutual information, accuracy, time, cognitive load, error, and uncertainty).

Axioms

1. It may be defined based on a principle (that has stood the test of time), and it cannot be deduced from other axioms.
2. …

Example Law: Optimal Visual Representation

Let $v \in \mathbf{V}$ be a visual representation, where v is optimal under a particular goodness measure $M \in \Xi$. Let $\mathbf{S}$ be the state space based on all variables $\Theta - \{\mathbf{V}\}$—that is, the subset of Θ without the visual representation $\mathbf{V}$. With appropriate definitions of M and $\mathbf{S}$, we have $M(v, s) \geq M(w, s)$, $\forall w \in \mathbf{V}$, $\forall s \in \mathbf{S}$.

existing ones. (See the "A Skeleton of a Theoretic System for Visualization" sidebar for an example.) In mathematics, axiomatization has been one of the driving forces in discovering rich axiomatic systems, each of which is underpinned by a set of primitive axioms. Historically, the early efforts that aimed to derive a self-complete axiomatic system motivated many innovations (such as in geometry), but they often failed to achieve the aim itself. Such failures led to Gödel's incompleteness theorems, which confirmed that such a self-complete axiomatic system is unattainable for any slightly complex theoretical system. Nevertheless, discovering axioms in the theoretical system is a noteworthy achievement in itself as long as we are aware of the axioms' limitations. Such a discovery is analogous to the pursuit of curating, evaluating, critiquing, and revising guidelines to discover principles.

One challenge in formulating a theoretical system for visualization is that there appear to be many variables in a visualization process, such as the source datasets, visualization tasks, display media, interaction devices, human viewers' knowledge and experience, interaction actions, application contexts, and so on. Some measurements are more attainable, such as data size, accuracy, and time. Other measurements may be problematic in terms of their theoretical conceptualization or practical implementation, such as information, knowledge, cognitive load, and task performance. Nevertheless, a theoretical system can be built bit by bit. We might start with a subset of these variables, while fixing other variables to a set of constants related to a scenario. We could also identify principles applicable to such a scenario and use them to formulate axioms and laws. Then, we could derive new laws based on existing axioms and laws in the system and test these new laws using experimentation and simulation. Any negative testing results will motivate further investigations into the theoretical system itself as well as the experimentation and simulation methods, yielding new improvements and advancements. New laws derived and confirmed in this way can be disseminated as new guidelines in practice.

The development of small theoretical systems will naturally lead to new advancements through integration and unification. For example, one theoretical system may focus on cognitive load in its measure space, and another may focus on training costs. Their unification would result in a more elegant and applicable theoretical system. We can expand our horizons in the endeavor to build theoretical systems for visualization, for example, addressing the relationships between visualization and emotions, aesthetics, language, social objects, or ethics.

Building a Theoretical Foundation

The visualization field has already seen more than 100 research papers on different aspects of a theoretical foundation for visualization. A recent keyword search using the term "visualization theory," for example, returned a wide variety of topics. Intriguingly, all returned items contained the word "measure" or variants of it. All included some ordered or numerical measurements, such as reliability, accuracy, correctness, limits, or optimality. Some papers discussed these measurements in the context of a framework, a model, or some form of a theory, and most included the word "quantify" or its variants. In addition to traditional quantities such as accuracy, precision, and time, the search results revealed some ambitious attempts to measure particular forms of human insight, understanding, performance, creativity, knowledge, cognitive load, learning, confidence, and many other attributes. A similar search of the visualization community returned more than 200 individual authors within the community.

Building a theoretical foundation should not be equated with creating a theory. Theoretical research is about creating new fundamental knowledge in each aspect and about making transformations, as shown in Figure 1. Taxonomies are essential for identifying all concepts (variables) and their states (values) in visualization. Ontologies are essential for identifying the interactions among these concepts (functions and relational variables). Under the contextual framework of taxonomies and ontologies, guidelines and principles postulate causal relationships. By organizing a collection of causal relationships coherently in an ontology that may also define other relationships, we can establish a qualitative model. In return, the development of a model informs us of any need for a new concept in a taxonomy or a new relationship in an ontology, while motivating us to discover new guidelines or study the conflicts of guidelines. The grounded theory method and other research methods in social sciences can help us achieve such transformations methodologically and systematically.

Using a quantitative theoretic framework, we can transform a qualitative model into a quantitative model, providing opportunities for model validation using experiments and computational simulation. Similarly, guidelines and principles can be quantitatively defined, leading to a more formal approach to defining causal relationships in visualization. When a quantitative model is structured as a theoretical system, we can infer new laws and prove or disprove a postulated law (for example, formulated based on a guideline) using existing axioms and laws in the system. A quantitative model, law, or theoretical system is predictive and therefore falsifiable. In turn, developing theoretical systems and investigating their extension and unification will stimulate new taxonomies, ontologies, guidelines, and models, thereby enriching visualization's theoretical foundation.

Building a theoretical foundation for visualization is the collective responsibility of the visualization community. In the literature, hundreds of authors have already contributed to different aspects of the foundation. The visualization community has demonstrated its ability in formulating taxonomies, proposing guidelines, and creating models. It possesses the unparalleled experience of working with a spectrum of visualization users and has accumulated much insight about the cost-benefits of many visualization and visual analytics workflows in different applications. Through collaboration, the community has acquired knowledge for empirical studies, mathematical modeling, and computational simulation and is continuing to learn new skills.

Building a theoretical foundation for visualization is the collective responsibility of the visualization community.

The community needs to build its confidence in directing a new generation of research students and postdoctoral researchers to tackle fundamental problems. Perhaps reviewers need to adjust their expectations of novelty to reflect the actual theoretical research activities of other scientific disciplines. For example, arvix.org lists 6,202 articles in 2016 alone in the category of "High Energy Physics – Theory." The collective effort to build a theoretical foundation in physics is enormous, making any significant breakthroughs much less romantic than portrayed by the media.

Making significant theoretical advances will lead to significant advances in practical visualization applications. For example, we all talk about "design" as an action in practice. A design space is commonly defined by a taxonomy or ontology. Most guidelines are proposed for improving designs. Most models suggest that designs or design processes can be optimized. When we have mathematically proven the correctness of a design guideline, this implies that the guideline must be obeyed in practice under the conditions defined by the corresponding quantitative law.

We hope every visualization researcher can find at least one pathway in this article through which to explore unanswered questions, known problems, and identified deficiencies in the theoretical foundation of visualization. No doubt, there are other pathways featuring unasked questions, unknown problems, and unidentified deficiencies. Like any research, building a theoretical foundation for visualization presents many challenges. It may not be all smooth sailing. We must always respect such challenges "in theory," but we should never be afraid of them "in practice."

References

1. C. Johnson, "Top Scientific Visualization Research Problems," *IEEE Computer Graphics & Applications*, vol. 24, no. 4, 2004, pp. 13–17.
2. P.G. Ferreira, *The Perfect Theory*, Little, Brown, 2014.
3. M. Chen et al., "References for Theoretic Researches in Visualization," 2017; sites.google.com/site/drminchen/themes/theory-refs.
4. H.-J. Schulz et al., "A Design Space of Visualization Tasks," *IEEE Trans. Visualization and Computer Graphics*, vol. 19, no. 12, 2013, pp. 2366–2375.
5. S. Carpendale et al., "Ontologies in Biological Data Visualization," *IEEE Computer Graphics & Applications*, vol. 34, no. 2, 2014, pp. 8–15.
6. US Dept. of Veterans Affairs, "TRM Glossary," www.va.gov/trm/TRMGlossaryPage.asp.
7. M. Meyer et al., "The Nested Blocks and Guidelines Model," *Information Visualization*, vol. 14, no. 3, 2015, pp. 234–249.
8. T. Zuk et al., "Heuristics for Information Visualization Evaluation," *Proc. AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization* (BELIV), 2006, pp. 1–6.
9. B. Glaser and A.L. Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Aldine Transactions, 1967.
10. D. Kaiser, *Drawing Theories Apart*, Univ. of Chicago Press, 2005.
11. D. Huang, M. Tory, and L. Bertram, "A Field Study of On-Calendar Visualizations," *Proc. 42nd Graphics Interface Conf.* (GI), 2016, p. 3.
12. R.E. Patterson et al., "A Human Cognition Framework for Information Visualization," *Computer & Graphics*, vol. 42, 2014, pp. 42–58.
13 P. Pirolli and S. Card, "The Sensemaking Process and Leverage Points for Analyst Technology as Identified through Cognitive Task Analysis," *Proc. Int'l Conf. on Intelligence Analysis*, vol. 5, 2005, pp. 2–4.
14. M. Chen et al., *Information Theory Tools for Visualization*, AK Peters/CRC Press, 2016.
15. G. Kindlmann and C. Scheidegger, "An Algebraic Process for Visualization Design," *IEEE Trans. Visualization and Computer Graphics*, vol. 20, no. 12, 2014, pp. 2181–2190.
16. D. Reed et al., *Computational Science: Ensuring America's Competitiveness*, President's Information Technology Advisory Committee (PITAC), Jun. 2005; www.nitrd.gov/pitac/reports/20050609_computational/computational.pdf.
17. T. Etiene et al., "Verifying Volume Rendering Using Discretization Error Analysis," *IEEE Trans. Visualization and Computer Graphics*, vol. 20, no. 1, 2014, pp. 140–154.
18. M. Chen et al., "From Data Analysis and Visualization to Causality Discovery," *Computer*, vol. 44, no. 10, 2011, pp. 84–87.

Min Chen *is a professor at the University of Oxford. Contact him at min.chen@oerc.ox.ac.uk.*

Georges Grinstein *is a research professor at University of Massachusetts. Contact him at ggrinstein@cs.umass.edu.*

Chris R. Johnson *is a distinguished professor of computer science and directs the Scientific Computing and Imaging (SCI) Institute at the University of Utah. Contact him at crj@sci.utah.edu.*

Jessie Kennedy *is a professor at Edinburgh Napier University. Contact her at j.kennedy@napier.ac.uk.*

Melanie Tory *is a senior research scientist at Tableau Software. Contact her at mtory@tableau.com.*

Contact department editor Theresa-Marie Rhyne at theresamarierhyne@gmail.com.