

Min Chen
University of
Oxford

Georges
Grinstein
University of
Massachusetts

Chris R.
Johnson
University of
Utah

Jessie Kennedy
Edinburgh
Napier
University

Melanie Tory
Tableau
Software

Pathways for Theoretical Advances in Visualization

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 panels at IEEE VIS Conferences and three workshops. 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, social sciences, and so on.
- The subject of visualization holds a distinctive position connecting human-centric processes (e.g., human perception, cognition, interaction, communication, etc.) with machine-centric processes (e.g., statistics, algorithms, machine learning, etc.). It therefore provides a unique platform to conduct theoretical studies that may impact on other disciplines.
- In comparison 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.

It is not uncommon to perceive that the “Theory of Visualization” is a topic only for a few individual researchers and its outcomes, perhaps in the forms of some theorems or laws, and may be too distant from practice to be useful. Perhaps inspired by well-known theoretical breakthroughs in the history of science, researchers in the field may unconsciously have high expectations for the originality, rigor, and significance of the theoretical advancements that may be made in a research project or presented in a research paper.

On the contrary, although textbooks tend to attribute a

major breakthrough to a pioneer at a particular time and a specific place, in most cases, such breakthroughs took years or decades, and were usually supported by numerous incremental developments, including a substantial number of erroneous solutions carried out by the pioneers themselves as well as many less well-known individuals. Many complex discoveries in the past did not appear to have elegant proofs at that time, and it has taken some very challenging and often questionable steps to obtain the well-formulated solutions in modern textbooks. For example, almost every reader of this article would associate the theory of general relativity with Albert Einstein’s discovery in November 1915 in Berlin. According to Petro Ferreira [2], Einstein first speculated about the generalization in 1907, 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 Mercury’s perihelion shift (Le Verrier, 1859), the 1919 eclipse expedition (Eddington, 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 suspended universe (Eddington).

During an Alan Turing Institute event in London in April 2016 on the Theoretical Foundation of Visual Analytics, the discussions on the need for building such a theoretical foundation attracted a wide range of 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 to a common understanding that a theoretical foundation consisted of several aspects (as described in the first boxed panel of this article), and every visualization researcher should be able to make direct contributions to some aspect of

Major Aspects of a Theoretical Foundation

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 has to 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 actions to be avoided in order to prevent an undesired outcome. The former usually implies a confidence in the high degree of generality and certainty of the causality concerned, while the latter suggests that a causal relation may be subject to specific conditions.

Conceptual Models and Theoretic Frameworks

The terms frameworks and models 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 description of complex causal relations in real world in a tentative manner, while 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 to a *theoretic system* through *axioms* (postulated quantitative principles) and *theorems* (confirmed quantitative laws). Unconfirmed guidelines are thus *conjectures* and contradictory guidelines are *paradoxes*.

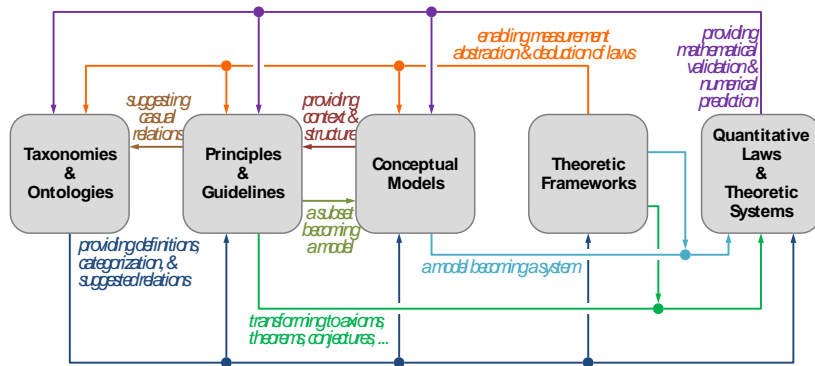


Figure 1: A theoretical foundation typically evolves through iterative developments. The development of each aspect influences as well as benefits from that of others. A successful transformation between different aspects indicates a theoretical enhancement of understanding. Note that the third aspect,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

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, which adopts the title of the panel, is a structured reflection by the panelists about the discussions during that IEEE VIS 2016 panel. In the remainder of this article, we will first discuss four major aspects of a theoretical foundation and then discuss the interactions and transformations between these aspects. We will follow with a summary of our viewpoints and provide our recommendations. 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 into concepts that are described and named to facilitate communications. 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. Alternative classifications (taxonomies) arise over time due to 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 concepts of entities and concepts of 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 field of visualization allowing us to organize and formalize our knowledge.

A brief review of the literature over the last three decades reveals at least 70 publications containing some form of visualization taxonomy [3]. Three questions are relevant when considering visualization taxonomies: (i) What is being classified (domain)? (ii) Why is the taxonomy being developed (purpose)? (iii) How is the taxonomy constructed (process)?

Methodological Development for Taxonomy [1]

The term taxonomy comes from the Greek word *taxis* (meaning “order” or “arrangement”) and suffix *-nomos* (meaning “law” or “science”).

Categorization (since 300s BC)

Plato was among the first to formulate methods for grouping objects based on their similar properties. Aristotle wrote the work of *Categories*, providing an in-depth study of classes and objects.

Taxonomy in Biology (since 3000s BC)

Naming and classifying plants and animals dates back to the origin of human languages. The development of modern botanical and zoological taxonomy 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.

Computational Taxonomy (since 1960s)

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

Methodological Development for Ontology [2]

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

Ontology in Philosophy

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 (e.g., “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 considers primarily the grouping relationships. Ontology can be seen as a generalization of taxonomy by allowing for different types of relationships among different entities.

Ontology in Computer Science

An ontology is 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 (e.g., [4]).

References

1. P. F. Stevens, *The Development of Biological Systematics*, Columbia University Press, New York, 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, Reading, MA, 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.

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 may 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 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. Further, using taxonomies as design spaces may reveal novel research opportunities, for example, by conducting gap analysis.

In comparison, the term “ontology” has appeared much less frequently in the visualization literature. This is partly because some studies on ontological relationships are presented as qualitative models. As 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 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 as they often feature partial and evolving hypotheses. A number of considerations therefore arise during the process, including 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 automatically using machine learning.

Taxonomies and ontologies are fundamental tools assisting in understanding, communication and development in the field of visualization. Yet 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 (one or more ontologies) for our work? The field of visualization continues to change. Taxonomies and ontologies must continue to evolve. 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 an aim. Guidelines are commonly outlined based on accumulated experience and knowledge about some causal relations in a process. On the one hand, it takes some courage and conviction to propose a new guideline. On the other hand, it takes a lot more courage and fair-mindedness to accept critiques about one’s guideline, and then retract or refine it. Some guidelines

Examples of Guidelines for Visualization

It is estimated that there are currently several hundreds of different guidelines recommended by various books, research papers, and online media. For example, these include:

Maximize the Data-Ink Ratio.

1. E. R. Tufte, *The Visual Display of Quantitative Information*, Graphics Press, Cheshire CT, 1983, p.93.

Overview first, zoom and filter, then details-on-demand.

2. B. Shneiderman, “The eyes have it: a task by data type taxonomy for information visualizations,” *Proc. IEEE Symposium on Visual Languages*, Washington, 1996, pp. 336-343.

Rainbow Color Map is Harmful.

3. B. E. Rogowitz and L.A. Treinish, “Data visualization: The end of the rainbow,” *IEEE Spectrum*, vol. 35, no. 12, 1998, pp. 52-59.
4. 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

5. 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.

Some 14 Guidelines for Data Visualization

6. <https://schoolofdata.org/2013/04/26/data-visualization-guidelines-by-gregor-aisch-international-journalism-festival/>, Accessed in Feb. 2017.

6 Guidelines for Creative Visualization

7. <http://www.tut.com/article/details/12-6-guidelines-for-creative-visualization/?articleId=12>, Accessed in Feb. 2017.

stand the test of time and become *principles*. Many others may be 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, out of date, or in conflict with other guidelines. Many documents about guidelines often contain a disclaimer [6]: “By definition, following a guideline is never mandatory. Guidelines are not binding and are not enforced.”

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 qualitative laws and qualitative process management. In the field of visualization, guidelines have no doubt played a positive role in designing and developing visualization systems as well as in education. Meyer et al. considered that guidelines are an integral part of an agile process for developing visual designs and visualization systems; they help designers make choices in such a process [7]. Zuk et al. proposed that guidelines can be used as heuristics for evaluating visual designs and visualization systems [8]. These recommendations inevitably place a huge burden upon the correctness and effectiveness of guidelines. If visualization guidelines are going to play a pivotal role as suggested by [7, 8], we will need to:

- develop *mechanisms* for curating, evaluating, critiquing, and refining guidelines in an open and transparent manner;

Examples of Models in Other Disciplines



Figure 2: Richard Feynman's 1975 Dodge van with the behavior model of subatomic particles painted on the sides. Image source: courtesy of ArtCenter College of Design, Pasadena, CA, USA.



Figure 3: James Watson (left) and Francis Crick (center) demonstrating the 3D physical model of DNA in 1953. The experimental results by Rosalind Franklin (right) were crucial to the discovery. Image source: <http://www.bbc.co.uk/education/guides/zsnssbk/revision/2> and <http://mentalfloss.com/article/53199/rosalind-franklin-and-search-dna>

- establish a culture of open, democratic, evidence-based discourse on the guidelines, and enable much broader participation in the discourse beyond the current scale of a few papers and blogs;
- inspire researchers to study guidelines, including their evolution and applicability in different conditions using scientific methods, and when appropriate opportunities arise, to transform guidelines into qualitative 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, etc.), 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 enabling the analysis of causal relationships, it facilitates the formulation of guidelines. 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 facilitate the establishment of a conceptual model.

Conceptual Models and Theoretic Frameworks

A conceptual model can be a representation of an idea, a process, or a 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, providing a visual way of linking

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 sub-atomic particles.

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 experiment, 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 latter half of the 20th century. The discovery of the double helix model of DNA was a research endeavor in the early 1950s. Many intermediate steps, ranging from the partial model alpha helix and the incorrect triple helix model by Pauling to X-ray diffraction experiments by Franklin and others, paved the way for Watson and Crick to formulate the landmark model in biology.

In the fields of visualization, more than 10 conceptual models [3] have been proposed for describing the relationships among data, visualization systems, analytical techniques, interaction methods, human perception and cognition, user tasks, and application contexts. The goal of such models is to help us describe, understand, reason about, and predict what people may do in a visualization process and environment, what actions may lead to what results in given circumstances, and which workflow is more efficient or effective than others.

For example, a model of personal visualization of 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 (e.g., [13]) have played an important role in supporting the design of interactive analysis tools. 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

An Example Theoretic System: Probability Theory

Measure Space

(Ω, \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 non-negative real number:
 $P(e) \in \mathbb{R}, P(e) \geq 0 \quad \forall e \in \mathbf{E}$.
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, \dots satisfies:
 $P(\bigcup_{i=1}^{\infty} e_i) = \sum_{i=1}^{\infty} P(e_i)$.

An Example of 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 :

$$\text{If } \mathbf{E}_A \subseteq \mathbf{E}_B \quad \text{then} \quad P(\mathbf{E}_A) \leq P(\mathbf{E}_B).$$

A Skeleton of a Theoretic System for Visualization

Measure Space

(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 (e.g., *Data* (\mathbf{D}), *Task* (\mathbf{T}), *Medium* (\mathbf{M}), *Visual Representation* (\mathbf{V}), *Human Capability* (\mathbf{H}), *Interaction* (\mathbf{I}), ...), and Ξ is a subset of all possible measures in visualization (e.g., probability, mutual information, accuracy, time, cognitive load, error, uncertainty, ...).

Axioms

1. *It may be defined based on a principle (that must have stood the test of time), and it cannot be deduced from other axioms.*
- ...

An Example of Law: Optimal Visual Representation

Let $v \in \mathbf{V}$ be a particular visual representation, and v is optimal under a particular goodness measure $M \in \Xi$. Let \mathbf{S} be the state space based on all variables $\Theta - \{\mathbf{V}\}$, i.e., the subset of Θ without visual representation \mathbf{V} . With an appropriate definitions of M and \mathbf{S} , we have:

$$M(v, s) \geq M(w, s) \quad \forall w \in \mathbf{V}, \forall s \in \mathbf{S}.$$

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, and more than 40 others for comparing different visualization techniques. In addition, through numerous application case studies, visualization researchers have had first-hand experience of 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 for formulating new models, performing continuous comparative analysis, probing negative experience, critiquing and improving existing models, broadening theoretical sampling, and exploring the model unification and theoretical saturation, all of which are advocated by the grounded theory methodology mentioned earlier. Building and analyzing qualitative models rigorously 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 tele- and 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. We must remember that Newton's first law of motion could not be fully validated for a long time because the technology for creating the conditions for a vacuum was not 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, and verify if they meet the requirements of the application concerned [17]. Since the cognition literature shows that the human observers' perception errors may not linearly correlate with discretization errors, 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 the sciences, many quantitative laws are regarded as disruptive discoveries, as they represent great leaps in our understanding about causal relationships from numerical uncertainty to numerical certainty [18]. As discussed earlier, any guideline in visualization 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. For example, part of Shneiderman's guideline "overview first, zoom, then details-on-demand" was proved using information theory (including an investigation of an anomaly) [14]. The filtering part of the guideline likely requires a more complex proof because defining what filtering would result in desired details may require additional variables.

In many disciplines, some laws have parameters that may be constants. The discovery of such fundamental constants (e.g., speed of light, absolute zero temperature, and so on) transforms

postulated laws to truly quantitative laws. Often discovering values that would fit such parameters requires extensive experimentation. For example, in psychology, Fitts's law has two parameters that vary according to the choice of input device, and Stevens's 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. One can imagine that if Newton's second law of motion had used *volume* instead of *mass*, it would have required an object-dependent parameter that we know now as *density*. Worse, if it were *surface area* instead of *mass*, one would need more object-dependent parameters.

The discipline of visualization 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 existing ones. 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 (e.g., in geometry) but 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 one is aware of the limitations of the axioms. Such a discovery is analogous to the pursuit for curating, evaluating, critiquing, and revising guidelines in order 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 data sets, visualization tasks, display media, interaction devices, human viewers' knowledge and experience, interaction actions, application contexts, and many more. 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. One may start with a subset of these variables, while fixing other variables to a set of constants related to a scenario. One may identify principles applicable to such a scenario, and use them to formulate axioms and laws. One can then 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 the cost of training. 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 field of visualization has already seen more than 100 research papers on different aspects of a theoretical foundation for visualization. Our recent search using the keyword "visualization theory," for example, returned a very wide variety of topics. Intriguingly, all returned items contained the term "measure" or variants of the word. All included some ordered or numerical measurements, such as reliability, accuracy, correctness, limits, optimality, and so on. Some papers discussed these measurements in the context of a framework, a model, or some form of a theory, and most included the term "quantify" or its variants. In addition to traditional quantities such as accuracy, precision, and time, the search 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 well over 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 as discussed above, and about making transformations as shown in Figure 1. *Taxonomies* are essential for identifying all concepts (i.e., variables) and their states (i.e., values) in visualization. *Ontologies* are essential for identifying the interactions among these concepts (i.e., 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, one 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 (e.g., 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 the theoretical foundation of visualization.

Conclusions

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 capability in formulating taxonomies, proposing guidelines, and creating models. It possesses the unparalleled experience of working with a wide spectrum of visualization users, and has accumulated much insight about the cost-benefit 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.

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 expectation of novelty to reflect the actual theoretical research activities of other scientific disciplinarians. For example, arxiv.org lists 6202 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 applications of visualization. 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 everyone can start 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 plain 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 London, 2014.
3. *References for Theoretic Researches in Visualization*, <https://sites.google.com/site/drminchen/themes/theory-refs>.
4. H.-J. Schulz et al., “A design space of visualization tasks,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, 2013, pp. 2366–75.
5. S. Carpendale, et al., “Ontologies in biological data visualization,” *IEEE Computer Graphics and Applications*, vol. 34, no. 2, 2014, pp. 8–15.
6. United States Department of Veterans Affairs, *TRM Glossary*, <http://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. BELIV*, 2006.
9. B. Glaser and A. L. Strauss. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Transactions, Chicago, 1967.
10. D. Kaiser, *Drawing Theories Apart*, University of Chicago Press, 2005.
11. D. Huang, M. Tory, and L. Bertram. “A field study of on-calendar visualizations.” In *Proc. Graphics Interface 2016*, May 2016.
12. R. E. Patterson et al., “A human cognition framework for information visualization.” *Computer & Graphics*, vol. 42 pp. 42–58, 2014.
13. P. Pirolli and S. Card. “The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis.” In *Proc. Int. Conf. on Intelligence Analysis*, vol. 5, 2005, pp. 2–4.
14. M. Chen, et al., *Information Theory Tools for Visualization*. A K Peters/CRC Press, 2016.

Thinking Theoretically

In the field of visualization, it is rare to find a researcher who has never had any theoretical thought. Given any topic **X** is visualization, one can pose an array of theoretical questions about **X**. The real challenge to everyone is to make an effort to answer any of these questions. For example, those who are specialized in **volume visualization** may have thought about the following questions:

- Q1. Should there be a taxonomy (or an ontology) about modalities, rendering techniques, transfer functions, colormaps, users, tasks, interactions, and so on?
- Q2. What are the guidelines for using surface rendering, amorphous effects, non-photorealistic effects, and global illumination in volume visualization?
- Q3. Does a non-physically-based integral (e.g., maximum intensity projection) incur extra cognitive load in comparison with those based on optical phenomena?
- Q4. What is an appropriate cognitive model for perceiving volumetric phenomena? How does the viewers’ soft knowledge (about the objects being visualized, the task of visualization, and the past experience) affect the perception? How do different rendering integrals affect the perception?
- Q5. Since we know that there is potentially a huge amount of information loss in volume visualization (e.g., loss through viewing only isosurfaces, or through the information integration along each ray), why is volume visualization beneficial?
- Q6. If we had a cognitive model for volume visualization, could it be unified with a cognitive model for video visualization since their basic forms of input data are very similar (i.e., a stack of images)?

15. G. Kindlmann and C. Scheidegger, “An algebraic process for visualization design,” *IEEE Transactions on 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. URL: www.nitrd.gov/pitac/reports.
17. T. Etienne, et al., “Verifying volume rendering using discretization error analysis,” *IEEE Transactions on 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,” *IEEE 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 at 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 Thereas-Marie Rhyne at
theresamarierhyne@gmail.com.