Theory of Visualization: From an Information-Theoretic Framework to an Information-Theoretic System

Part 1: Track Record

Oxford Engineering Science (http://www.eng.ox.ac.uk/)

From 1 August 2017, Oxford e-Research Centre (where Professor Min Chen has been based since 2011) became part of Department of Engineering Science. Oxford Engineering Science has an international reputation for its research in all the major branches of engineering and its research groups under the theme of “Information, Control & Vision” are a leading force in developing data science and knowledge engineering. The broad collection of existing research activities (such as machine learning, driverless cars, economic modelling, video and image understanding, and so on) and the upcoming creation of the Oxford Information Science Institute will provide this proposal with an exceptional backdrop as well as the inspiration for making fundamental breakthroughs.

Professor Min Chen (http://www.oerc.ox.ac.uk/people/min-chen)

Professor Min Chen is an internationally-established scientist in visualization. His main research interests have been visualization, computer graphics, interactive systems and aspects of computer vision. In particular, he has made significant contributions in volume graphics (e.g., [mc1]), video visualization (e.g., [mc2]), information visualization (e.g., [mc3]), visual analytics (e.g., [mc4]), theory of visualization (e.g., [mc5]), quality metrics for visualization (e.g., [mc6]), and empirical studies for visualization (e.g., [mc7]). He teaches the subject of visual analytics to 4th year and MSc students, and gives short courses to two DTC programmes. His broad knowledge and experience enables him to tackle the theoretical challenges discussed in this proposal.

Chen has co-authored over 200 publications, including some 90 journal papers. He has been awarded over £12M research grants from EPSRC, JISC (AHRC), TSB (NERC), Royal Academy, Welsh Assembly Government, HEFCW, industry, and several UK and US government agencies. Since 1992, under his supervision or co-supervision, 24 PhD and 3 MPhil students have successfully completed their research programs. He also supervised some 37 research officers in both academic and industrial environments. He is currently leading visualization activities at Oxford, working on a broad spectrum of interdisciplinary research topics, ranging from the sciences (e.g., [mc8]) to sports (e.g., [mc9]), from digital humanities (e.g., [mc10]) to cyber security (e.g., [mc11], and from colour perception (e.g., [mc12]) to philosophical discourse (e.g., [mc13]).


Professor Min Chen’s Work on Theoretical Foundation of Visualization

Professor Min Chen has made significant contributions to the theoretical research in visualization, in particular in applying information theory to the field of visualization [mc14]. Together with his co-authors, he proposed an information-theoretic framework for underpinning the subject of visualization [mc5]; used information theory to explain complex phenomena in visualization, such as visual embellishment [mc7] and visual multiplexing [mc15]; and deployed information-theoretic measures in visual quality metrics [mc5], anomaly detection [mc11], and glyph design [mc16]. Recently, together with an economist and information-theorist (Golan), he proposed an information-theoretic metric for measuring the cost-benefit of machine-centric processes (statistics and algorithms) and human-centric processes (visualization and interaction) in data intelligence [mc17]. The metric provides a theoretic utility function for optimizing data intelligence workflows, enabling a deep understanding of the complementary relationship between machines and humans in data intelligence. This new theoretical insight was confirmed in a field-based observational study that compares human-centric and machine-centric processes for building classification models.
and a laboratory-based manipulation study for measuring humans' soft knowledge in time series visualization.

A Selection of References

A full list of publications can be found at https://sites.google.com/site/drminchen/pub-chron


Part 2: Proposed Research

1. Ambition and Potential Transformation
In 2010 when the UK led the proposal for using information theory to underpin the field of visualization [mc5], many were sceptical. The work [mc5] stimulated more research effort into a few proposals of alternative theories of visualization. In 2016 and 2017 when the UK led a series of theoretic breakthroughs [mc17,mc18,mc19], together with a new book [mc14], information theory became the focus of the theoretic development in visualization. There are ongoing research activities in the USA, Spain, and Austria as well as new initiatives in Canada, Germany, and China.

Our ambition is to transform the current information-theoretic framework with a few measures and laws to a theoretic system for visualization. The latter will consist of a set of coherently defined axioms that reflect the most fundamental concepts and principles in visualization and a set of information-theoretic operations for generating new laws and validating any hypothesised laws. Using a theoretic system, various wisdoms about the best practices in visualization can each be transcribed as a law. Since a piece of wisdom is not guaranteed to work everywhere, the theoretic system can help scrutinise the operational conditions of such a transcribed law rigorously. Furthermore, when new laws are derived from some axioms and existing laws, they can be enriched and augmented with visualization contexts, resulting in practical guidelines [1].

The cost-benefit metric proposed by Chen and Golan [mc17] is generic to both machine- and human-centric processes. The notion of cost-benefit has been used to carry out comparative analysis of model-development using machine learning and visual analytics [mc18]. The metric has been measured in a recent empirical study that was designed to detect the effect of human knowledge in data visualization [mc19]. These early advancements indicate the potential that information theory can bring several major aspects of data science together, including visualization, statistical inference, and cognitive science [2,3]. This would lead to a transformative advancement in data science. Although this travel grant focuses mainly on a theoretic system for visualization, such a transformative advancement will always be in our mind during the project.

2. The State of the Art
Visualization is a form of communication. Figure 1 shows an abstract representation of a general communication system considered by Shannon and Weaver [4]. The source and destination of the message can be a person or a machine. The encoder (also referred to as transmitter) and decoder (also referred to as receiver) transform messages into signals and vice versa. Conceptually, the term “signal” is a generalization encapsulating both digital and analogue signals. In modern communication systems, we can simply consider both messages and signals as “data”. Traditionally, the term “channel” refers to a transmission medium. In abstract, it is a function or process that operates on an input signal and sometimes adds noise, resulting in an output signal.

It is not difficult to observe that a visualization process can have more or less the same abstraction. For instance, a person or a machine may encode some data using visual objects,
resulting in a visualization to be transmitted via a display device. A viewer decodes the visualization and makes sense of the data being conveyed. Figure 2 depicts a general visualization pipeline. To make the comparison easier, we can group the eight processes, from filtering to cognition, into three subsystems as shown in Figure 2, which we colloquially refer to as vis-encoder, vis-channel, and vis-decoder. This results in a model analogous to Figure 1. As discussed in [mc5, mc13], feedbacks involving human-computer, human-human, or inter-process interaction can be modelled using the pipeline in Figure 2 through process sequentialisation.

The application of information theory in visualization began in the late 1990s, when entropy-based metrics were used for measuring scene complexity and optimising illumination, view selection, and feature highlighting. A suggestion that information theory could be used as a theoretic framework for visualization was first featured in a report in 2008. Due to the constraints of space, we refer readers to [mc5] for the references to these early developments. In 2010, Chen and Jänicke [mc5] presented a detailed description of an information-theoretic framework for visualization, covering a range of aspects including metrics, laws, and the relevance of major concepts (e.g., source encoding, channel encoding, and data processing inequality). They showed that information theory could explain well-known phenomena in visualization (e.g., logarithmic plots, overview-zoom, and motion parallax). They first pointed out that interactive visualization broke the conditions of data processing inequality that is a bottleneck in automated data analysis.

Following the establishment of the theoretic framework, there has been increasing effort in using information theory to underpin various technical developments in visualization. There were about 16 papers that described the uses of information theory in visualization during 1999-2009. In comparison, during 2011-2016, there were over 70 papers (estimated based on a list of papers that cited [mc5]). The new applications of information theory in visualization include (i) a technique for automated transfer function design [5], (ii) a discourse on visual uncertainty in visualization [6], (iii) the discovery and explanation of a perceptual phenomenon, visual multiplexing [mc15], and (iv) a new metric, called quasi-Hamming distance, for optimizing glyph designs [mc16].

In 2015, Chen and Golan [mc17] made a theoretic breakthrough by proposing a new metric for optimising data intelligence workflows, theorising the measurement and comparison of the cost-benefit ratio of human- and machine-centric processes in data intelligence. The metric represents a further generalization of Shannon’s grouping property [7] by encompassing the notion of soft knowledge through the use of Kullback-Leibler divergence [7]. In 2016, Tam et al. [mc18] investigated two case studies where human-centric processes performed better than machine-centric processes, and provided quantitative estimation (in the information unit bits) of the soft knowledge featured in the two case studies. In 2017, Kijmongkolchai et al. [mc19] reported an empirical study confirming quantitatively that humans’ soft knowledge have a positive benefit in reducing potential distortion in data intelligence by steering analytical inference towards the underlying probability distribution of the data.

2016, when we commemorated Claude Shannon’s 100th birthday, saw a new surge of interests in theory of visualization. In April, Chen et al. organised the Alan Turing Institute Symposium on Theoretic Foundation of Visual analytics [8]. In May, Chen et al. gave a tutorial on Information Theory in Visualization in Eurographics Conference [9]. In October, the tutorial was repeated in IEEE VIS, where Chen also convened a plenary panel on Pathways for Theoretical Advances in Visualization [1,10]. Also during IEEE VIS, CRC Press (Taylor & Francis) launched a new book, Information Theory Tools for Visualization, by Chen, Sbert and other four co-authors [mc14].

2. Research Questions

Chen et al. outlined four components of a theoretic foundation [1,10]:

- **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, which 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 high degree of generality and certainty of the causality
Conceptual Models and Theoretic Frameworks: The terms of 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, which is 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.

In the field of visualization, it is estimated that there are currently nearly a hundred papers on taxonomies and ontologies [11]; several hundreds of principles and guidelines recommended by various books, research papers, and online media; more than a dozen conceptual models [11]; a few theoretic frameworks [11]; and a few quantitative laws that are mathematically confirmed guidelines. There is not yet any theoretic system proposed in the literature. Much more effort will be required to develop the theoretical foundation of visualization.

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 [12]. 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 theoretic 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 theoretic system. Nevertheless, discovering axioms in the theoretic 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 [1].

One challenge in formulating a theoretic 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 theoretic 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 theoretic 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. Figure 3 juxtaposes an example theoretic system, Probability Theory and a skeleton of a theoretic system for visualization [1]. This suggests a pathway for achieving a theoretic system for visualization.

The development of small theoretic systems will naturally lead to new advancements through integration and unification. For example, one theoretic 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 theoretic system. We can further expand our horizons in the endeavour to build theoretic systems for visualization, for example, addressing the relationships between visualization and emotions, aesthetics, language, social objects, or ethics [1].

3. Aim and Objectives

The overall aim of this project is transform the current information-theoretic framework [mc5] to an information-theoretic system. Based on the definition of in Section 2, we would need to identify a conceptual model, a set of axioms (postulated quantitative principles) and theorems (confirmed quantitative laws. This thereby leads to the following objectives:

O1. Axiomatisation. To outline an information-theoretic system for visualization based on a conceptual model, and identify a set of axioms through a careful examination of their mathematical properties as well as practical implications in visualization as postulated quantitative principles. This strand of research will span the entire project period because axioms can sometimes be misinterpreted as theorems and their identification will be influenced by the search for theorems or quantitative laws for visualization.

O2. Theorems. To examine the existing theorems (including equalities and inequalities) in information theory, and relate them to practical guidelines in the field of visualization. As some guidelines, such as “overview first, zoom, details on demand” have already been confirmed as quantitative laws [mc5][mc14], more discoveries are expected to be made. We plan to examine over 200 theorems systematically during a 24-month period. With a pessimistic estimation that 20% of the theorems would be found relevant to visualization, the discovery of some 40 new quantitative laws for visualization would represent a major leap in comparison with fewer than 10 such laws in the current visualization literature.

O3. Theoretic Systems and their Unification. We are confident that we will be able to establish a few of narrowly-defined theoretic systems towards the end of the second year. We will focus our third year research on unification. Historically theoretical unification has always been a driving force in the advancement of the sciences. Alan Turing’s Universal Computer exemplifies such a great endeavour. Using the terms in Figure 3 [1], for example, we may initially have a theoretic system involving a data alphabet (D) and a visualization alphabet (V), and another theoretic system involving a task alphabet (T) and an interaction alphabet (I). We will certainly attempt to extend the first system to include a display-medium alphabet (M) and make a further attempt to unify the two systems.

4. Methodologies

For Objective O1, we will first identify a conceptual model that is suitable for the transformation based on the following criteria:

- The model can be represented adequately under the information-theoretic framework [mc5],
- The model can encompass a large portion of phenomena, activities, and processes in visualization,
The model should reach a sufficient level of abstraction, with a small number of axiomatic variables.

As briefly surveyed in [mc17], there are over 10 pipelines and over 10 conceptual models in the visualization literature. Although these pipelines and conceptual models have been studied during the work of [mc17], we will revisit some of them in detail. Our current conjecture is that the conceptual model used in [mc17] is the most promising candidate, because (i) it is already underpinned by the information-theoretic framework, (ii) its connections to the four levels of visualization and six common workflows indicate its wide coverage, (iii) its coverage of parallelism, interaction and iteration through sequentialisation offers a balance between general applicability and theoretic abstraction, and (iv) it has only three elementary measures and four axiomatic variables at each processing step. Nevertheless, we will be open-minded about other conceptual models.

For **Objective O2**, we will conduct the research through a combination of collaborative brainstorming discussions (e.g., face-to-face meeting, whiteboard, coffee, and walk) and individual intensive research (e.g., pen, paper, book, search engine, and latex). We will begin with revisiting the five quantitative laws proposed in [mc5,mc17] by rewriting them out using the axiomatic variables and elementary measures established in the process for achieving **Objective O1**. We anticipate that this initial effort will be mutually beneficial to both **Objectives 1 and 2**, enabling the formulation of a theoretic system featuring appropriate abstraction for **O1** as well as supporting a large number of theorems for **O2**. This initial effort will be followed by systematic examinations of the theorems in [7], and later in other books and papers (including the two books by Golan [13,14]). We plan to cover roughly 20 theorems (including equalities and inequalities) in each collaborative visit (see also the WORK PLAN document).

The theorems in information theory will be coarsely placed into five groups such that group 1 consists of basic theorems, groups 2 and 3 are intermediate, and groups 4 and 5 are relatively more advanced. Meanwhile, in relation to the two collaborative partners, groups 2 and 4 are closer to Professor Golan’s research interests and expertise, and groups 3 and 5 are closer to Professor Sbert’s research interests and expertise. The theorems in [7] are tentatively grouped as:

- **Group 1 (Basic, Chen, Golan, and Sbert)**: Chapter 2 in [7],
- **Group 2 (Intermediate, Chen and Golan)**: Chapters 3, 4, 6, 8 in [7]
- **Group 3 (Intermediate, Chen and Sbert)**: Chapters 5, 7, 9, 10 in [7]
- **Group 4 (Advanced, Chen and Golan)**: Chapters 11, 12, 15, 16 in [7]
- **Group 5 (Advanced, Chen and Sbert)**: Chapters 13, 14, 17 in [7]

Other theorems in the literature will be selectively added into these groups.

For **Objective O3**, we plan to adopt the methodology that we used to develop the information-theoretic metric for cost-benefit analysis [mc17]. For that piece of work, we started with a very specific application, i.e., time series aggregation. We then extended the scope to time series visualization. We further extended the scope to cover correlation analysis. At that point, we had an initial mathematical formula for the metric. After we applied the formula to several data intelligence processes, we improved the formula. Similarly, we will start with an abstraction of a specific application and develop a theoretic system may suit this application. Through continuing extension, unification, evaluation, and validation, we hope to arrive a theoretic system that provides a good coverage of various alphabets in visualization (e.g., Data (D), Task (T), Medium (M), Visual Representation (V), Human Capability (H), Interaction (I), and so on).

The work plan for delivering these three objectives consists of 12 tasks, which are detailed in the WORK PLAN document.

5. Potential Impact and Outputs

This proposed project timely reflects the latest research agenda for building a theoretical foundation of visualization [1], which was first outlined in the Alan Turing Institute Symposium on Theoretical Foundation of Visual Analytics (April 2016) [8], and then elaborated in IEEE VIS Panel on Pathways for Theoretical Advances in Visualization (October 2016, Best Panel Award) [1,10]. The short-term impact (3-5 years) that this project may generate will be in the forms of (i) the provision of a consistent theoretic system for validating any guideline proposed for visualization, (ii) the consolidation of the information-theoretic framework for facilitating qualitative modelling of
visualization and visual analytics processes, and (iii) the advancement from a loosely-defined theoretic framework to a tightly-defined theoretic system as a landmark breakthrough in theory of visualization. We also hope that this work will yield some long-term impact (6-15 years) to a variety of visualization applications and the theoretical foundations of data science and cognitive science. Further details of the potential impact and the pathways to reach such impact can be found in the PATHWAYS TO IMPACT document.

We plan to write and submit at least four high impact journal papers during the 36-month project, and to present our work in IEEE VIS, EuroVis, and Info-Metrics Conferences.

6. Future Development and Other Potential Benefits

Visualization is an integral part of data science. The theory that we are building covers interfaces among statistics, algorithms, visualization, and interaction [1]. It is very likely that information theory, which is underpinned by probability theory, can bring statistics, visualization and interaction under a common theoretic framework. When appropriate connections with algorithms can be made (e.g., through algorithmic information theory, probabilistic analysis of algorithms, and probabilistic programming), a theoretic foundation for data science will not be beyond our reach. This is a medium-to-long-term endeavour where the UK can lead and flourish. The theoretical insight to be gained in this project and beyond will be essential in transforming various techniques for data intelligence to a causality discovery technology [12].

References

PATHWAYS TO IMPACT (maximum 2 sides of A4)

Potential Impact

Visualization is a ubiquitous technology and plays an important role in everyday life as well as in complex data intelligence processes. Its benefits have already been mathematically confirmed by the recent theoretical advances [mc5][mc17][mc18][mc19]. However, the theory of visualization is still at its infancy. Although there are numerous visualization tools and systems around the world, we are yet to establish a coherent mathematical system that can be used to model and reason algorithmic, visual, interactive, and cognitive phenomena or to underpin design principles and guidelines in visualization. In an analogy, one has pyramids, walls, buildings, and bridges, but does not have the Euclidean geometry system.

If this project successfully develops a theoretic system for visualization through axiomatisation and theorem discovery, such a theoretic system will have a transformative impact on the discipline of visualization including its main subfields (information visualization, scientific visualization, and visual analytics) and many domain-specific areas (e.g., biomedical data visualization, software visualization, and visualization literacy). Such impact may include:

- It has been estimated that there could be several hundreds of guidelines in visualization recorded in different books, research papers, blogs, and lecture slides. A vast majority of have not been defined or validated mathematically. Inevitably some may not be applicable under many conditions and some are often in conflict. **This work will provide a consistent theoretic system for validating any guideline proposed for visualization.**

- There are at least a dozen conceptual models proposed to describe visualization processes, but hardly any of them feature sufficient number of qualitative measures that would enable these models to be rigorously tested, evaluated, and validated in practical applications. The information-theoretic framework proposed in [mc5] offers the potential to enable the transformation from conceptual models to quantitative models [1], which can be measured in laboratories and field trials and be computationally simulated. **This work will strengthen the role of information-theoretic framework in qualitative modelling of visualization and visual analytics processes.**

- The information-theoretic framework proposed in [mc5] has already provided a basic set of quantitative measures. Three further measures were introduced in [mc17]. Under such a framework, a quantitative law (theorem) describes a causal relation of concepts using a set of measurements and a computable function. Given a set of interrelated quantitative laws under the same framework, there potentially exists a theoretic system. **This work will make such a landmark transformation from an information-theoretic framework to an information-theoretic system for visualization.** With a theoretic system, postulated quantitative principles are axioms, confirmed quantitative laws are theorems, unconfirmed guidelines are conjectures, and contradictory guidelines are paradoxes.

In addition to the above direct impact to the field of visualization, this work is expected to have some indirect and/or long-term impact:

- **Applications.** Although visualization has been shown to be useful in many applications and its cost-benefit in data intelligence was recently confirmed mathematically, it will take some time for many potential users, especially decision makers and experts in other computing technologies (e.g., machine learning) to be convinced by the merits of the visualization technology. By having information theory as the framework [mc5], a theoretic system as the foundation, and a collection of techniques as the building blocks, potential users can be much more confident in using visualization in their data intelligence workflows.

- **Data Science.** Data science is a new interdisciplinary subject. Although many different mathematical methods are being applied to various aspects of data science, a more coherent mathematical framework will likely emerge in the future, which can bring together many or most strands of data science (e.g., machine learning, data mining, visual analytics, network analysis, cryptography, decision theories, dynamic system, algorithmic complexity, etc.). This work might only be a drop [2] in the ocean of such an enormous challenge, “the secret of getting ahead is getting started.” [Mark Twain]
Cognitive Science. Visualization offers an exciting connection between cognitive science and information theory. Any attempt of using information theory to explain or model visualization will inevitably reach out to aspects of cognitive science. For example, the cost-benefit metric proposed in [mc17] has been linked to the phenomena such as parallel and sequential visual search, selective attention and inattentional blindness, gestalt grouping and illusion, heuristic and biases, and memory representation and reconstruction [3].

Pathways to Impact
The pathways to the abovementioned impact will be through high impact research publications, and presentations in conferences. The PI’s track record in publications indicated the feasibility of this pathway. In addition, the PI on average give 2-4 invited speeches per year, which provide opportunities for wider dissemination. For example, the PI’s recent speech schedule includes:

- **Four Levels of Visualization.** Invited Talk, Swansea CODAH Workshop, January 2016.
- **Four Levels of Visualization.** Seminar, Oxford Brookes University. UK, March 2016.
- **Discovering Treasure Troves in Digital Humanities.** Keynote Speech. 2016 Workshop on Visualization for Digital Humanities, Baltimore, USA, October 2016.
- **Can Soft Knowledge in Data Intelligence be Measured?** Seminar, American University, DC, USA, November 2016.
- **Shannon’s Dilemma.** Invited Talk, Info-metrics Institute Workshop for the Commemoration of Claude Shannon’s 100th Birthday, American University, DC, USA, November 2016.
- **Information Theory and Visualization.** Seminar, Leipzig University, November 2016.
- **Information Theory and Visualization and Glyph-based Visualization.** Two Invited Lectures, CSHL Course on Immersive Approaches to Biological Data, Cold Spring Harbor Laboratory, New York, December 2016.
- **Cost-benefit of Visual Analytics and Four Levels of Visualization.** Two Invited Talks, DBVIS 2017, Bühlerhof, Germany, April 2017.
- **Cost-benefit Analysis of Data Intelligence.** Seminar, Joint Computational/Mathematical Biology Seminar Series, Mathematical Institute, Oxford, UK, May 2017.
- **Four-Levels of Visualization in Digital Humanities.** Invited Talk, Lorentz Center Workshop on Visualizing Digital Humanities, Leiden, The Netherlands, June 2017.
- **Data Visualization.** Invited Talk, Symposium and Workshop on Crops In Silico, Oxford, UK, June 2017.
- **What is More Uncertain: Data, Visualization, or Insight?** SFB/Transregio 161 Workshop on Uncertainty Visualization, Konstanz, Germany, July 2017.
- **What is More Uncertain: Data, Visualization, or Insight?** Seminar, Hong Kong, August 2017.
- **Myths and Promises of Data Intelligence,** Keynote Speech, Cyberworlds 2017, Chester, UK, September 2017.
- **The Value of Interaction in Data Intelligence.** Keynote Speech, BCS HCI conference in Belfast, UK, July 2018.
The research activities for this project are planned based on the three objectives defined in Section 3 and the methodologies outlined in Section 4. The PI will lead and coordinate all these activities. Broadly speaking, the PI and Professor Golan will work closely on aspects featuring probability theory, statistical inference, information-theoretical optimization (e.g., maximum entropy), and applications in economics and decision-making. The PI and Professor Sbert will work closely on aspects featuring data communication, data compression, data encryption, and other computing-related topics. In addition to O2, Professor Golan’s mathematical expertise will contribute to the overall aim for developing theoretic systems, and in particularly objectives O1 and O3. Professor Sbert’s expertise in visual computing, including aspects of computer graphics, image processing, and visualization will help the PI in relating the theoretical discoveries to practical applications in visualization and beyond. In general, with the broad knowledge and experience in visualization and visual analytics as well as the internationally-leading expertise on theory of visualization, the PI will take a significant amount of workload in delivering tasks related to objectives O1 and O3.

The delivery of this project will also indirectly benefit from the PI’s ongoing research activities, including several collaboration projects with University of Konstanz (on ontology, cognition in visual analytics, etc.), University of Leipzig (on colormaps in visualization), and University of Texas Austin and University of Chester (on information-theoretic analysis of VR/VE environments).

We plan to submit at least four papers to the two major publication venues in the field of Visualization, i.e., EuroVis 20xx (Wiley Computer Graphics Forum) and VIS 20xx (IEEE Transactions on Visualization and Computer Graphics), starting from the last quarter of Year 1. As part of the plan for these research and publication activities, a detailed travelling plan is given in the document JUSTIFICATION OF RESOURCES REQUESTED.

<table>
<thead>
<tr>
<th>Objectives and Tasks</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
</tr>
<tr>
<td></td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
</tr>
<tr>
<td></td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
</tr>
<tr>
<td><strong>O1: Axiomatisation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1: In-depth analysis of VIS pipelines</td>
<td>□</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2: Identify candidature VIS principles</td>
<td>□</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T3: Define a model &amp; a set of axioms</td>
<td>□</td>
<td>□</td>
<td></td>
</tr>
<tr>
<td>T4: Review T2 and T3</td>
<td></td>
<td></td>
<td>□</td>
</tr>
<tr>
<td><strong>O2: Theorems</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T5: Mapping Groups 1 and 2 theorems</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>T6: Mapping Groups 1 and 3 theorems</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>T7: Mapping Group 4 theorems</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>T8: Mapping Group 5 theorems</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>T9: Integrated study of all five groups</td>
<td>□</td>
<td>□</td>
<td></td>
</tr>
<tr>
<td><strong>O3: Theoretic Systems &amp; Unification</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T10: Develop 2-3 simple systems</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>T11: Study unification</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td><strong>General Research Activities</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T12: Publication effort (submission)</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

□: Activities involving the PI and both Golan and Sbert. □: Activities involving the PI and mainly Golan. □: Involving the PI and mainly Sbert.
JUSTIFICATION OF RESOURCES REQUESTED (maximum 2 sides of A4)

Investigator

This is a theoretical research project to be conducted through close collaboration with experienced scholars who are knowledgeable about information theory and have strong research interests in visualization in particular and data science in general. Chen, who led the previous UK initiatives in applying information theory to visualization [mc5, mc11, mc13~mc19], will lead this new initiative.

The two main collaboration partners, whose research activities are being funded by their own funding sources, are both well-established scholars in their own fields. Chen has already had close collaboration with both Golan [mc17] and Sbert [mc14]. Three of them are currently working together on a theoretical problem arising from a recent empirical study on soft knowledge visualization [mc18]. During Sbert's recent visit to Oxford (July 2017), they managed to find a solution. They are preparing a paper to be submitted to a journal.

Professor Amos Golan (American University, DC, USA) is an internationally-leading expert in information-theoretic inference in economics [13,14]. He is the director of the Info-Metrics Institute, which brings together leading scientists and scholars in different disciplines to a think-tank focusing on the fundamental questions related to “information”. His knowledge in both information theory and economical prediction and decision-making will enable any proposal for a theoretical formulation (e.g., a measure, an axiom, a law) to be cross-examined from the perspective of probabilistic inference and decision making based on information theory. Professor Golan will focus on the validation of various theoretical notions and mathematical formulae developed in this project. He will contribute to the project with his own research funding in the USA, and will help disseminate the project outputs through the Info-Metrics Institute.

Professor Mateu Sbert (University of Girona, Spain) is an internationally-leading expert in applying information theory to computer graphics, image processing, and visualization. He has co-authored three books on these topics [15,16,mc14]. He was recently appointed as a part-time 1000 Plan professor at Tianjin University, China, leading the development of a programme for applying information theory to visual computing and cognitive science. Professor Sbert will focus on the various inequalities in information theory, exploring their relevance to visual computing.

Note. It has been suggested that the work might be carried out by one or two PDRAs. In fact, there was an EPSRC proposal “Underpinning Data Intelligence Using Information Theory” from the Chen and three co-PI in 2015 for a project involving two PDRAs. Although that proposal was not funded, Chen pursued the theoretical aspects of the proposal through his collaboration with Professor Amos Golan. During 2015 and 2016, Chen and Golan exchanged 5 visits, totalling about 60 days (mostly funded by Info-Metrics Institute). Through intensive brainstorming discussions, they delivered the cost-benefit metric [mc17], which was one of the four objectives defined in the unsuccessful EPSRC proposal. This experience suggests that it is highly feasible to conduct serious theoretical research through an EPSRC Overseas Travel Grant. On the other hand, training a PRDA without the knowledge of both information theory and visualization will likely take 1~2 years. For the PI to obtain funding for providing such training opportunities has so far been difficult during the past 36 months.

Travel and Subsistence

Being able to visit Professor Golan in DC, USA and Professor Sbert in Girona, Spain will be critical to the collaborative research activities. In addition, in order to keep abreast of new developments in visualization and visual analytics and to present the outcomes of this research, it is essential for the PI to attend major conferences. IEEE VIS (including VAST, Infovis and SciVis), EuroVis (including EuroVA) are the prime venues, where top conference papers are also accepted as journal papers, and papers from reputable journals (IEEE Transactions on Visualization and Computer Graphics, and Wiley Computer Graphics Forum) are offered presentation slots.

Eighteen trips were planned over a three year period for collaborative visits and attending conferences. The detailed research tasks related to these trips are given in the document WORK PLAN. The timing of these trips assumes that the PI will have teaching duties during January-March, 2018, 2019, and 2020. For the 2017/2018 session, the PI does not have any teaching duty. The timing also assumes that Professor Golan may visit Oxford in February 2019, 2020, and 2021,
and Professor Sbert may visit Oxford in September 2019 and 2020. The budgetary summary of all PI’s trips is given below:

<table>
<thead>
<tr>
<th>Proposed Time</th>
<th>Duration</th>
<th>Destination</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>April 2018</td>
<td>two weeks</td>
<td>American U., DC, USA</td>
<td>£2900</td>
</tr>
<tr>
<td>May 2018</td>
<td>two weeks</td>
<td>U. Girona, Spain</td>
<td>£2180</td>
</tr>
<tr>
<td>June 2018</td>
<td>two weeks</td>
<td>EuroVis (Brno, Czech) + U. Girona, Spain</td>
<td>£2980</td>
</tr>
<tr>
<td>August 2018</td>
<td>two weeks</td>
<td>American U., DC, USA</td>
<td>£2900</td>
</tr>
<tr>
<td>October 2018</td>
<td>two weeks</td>
<td>IEEE VIS (Berlin) + U. Girona, Spain</td>
<td>£2980</td>
</tr>
<tr>
<td>November 2018</td>
<td>two weeks</td>
<td>American U., DC, USA</td>
<td>£2900</td>
</tr>
<tr>
<td>April 2019</td>
<td>two weeks</td>
<td>American U., DC, USA</td>
<td>£2900</td>
</tr>
<tr>
<td>May 2019</td>
<td>two weeks</td>
<td>U. Girona, Spain</td>
<td>£2180</td>
</tr>
<tr>
<td>June 2019</td>
<td>two weeks</td>
<td>EuroVis (Portugal) + U. Girona, Spain</td>
<td>£2980</td>
</tr>
<tr>
<td>August 2019</td>
<td>two weeks</td>
<td>American U., DC, USA</td>
<td>£2900</td>
</tr>
<tr>
<td>October 2019</td>
<td>two weeks</td>
<td>IEEE VIS (USA) + American U., DC, USA</td>
<td>£3900</td>
</tr>
<tr>
<td>December 2019</td>
<td>two weeks</td>
<td>U. Girona, Spain</td>
<td>£2180</td>
</tr>
<tr>
<td>April 2020</td>
<td>two weeks</td>
<td>American U., DC, USA</td>
<td>£2900</td>
</tr>
<tr>
<td>May 2020</td>
<td>two weeks</td>
<td>U. Girona, Spain</td>
<td>£2180</td>
</tr>
<tr>
<td>June 2020</td>
<td>two weeks</td>
<td>EuroVis (Sweden) + U. Girona, Spain</td>
<td>£3900</td>
</tr>
<tr>
<td>August 2020</td>
<td>two weeks</td>
<td>American U., DC, USA</td>
<td>£2900</td>
</tr>
<tr>
<td>October 2020</td>
<td>two weeks</td>
<td>IEEE VIS (USA) + American U., DC, USA</td>
<td>£2980</td>
</tr>
<tr>
<td>December 2020</td>
<td>two weeks</td>
<td>U. Girona, Spain</td>
<td>£2180</td>
</tr>
</tbody>
</table>

**Total:** **£50920**
Timeline of the Submission, Review and Decision of this Proposal

Submission on 19/09/2017

01/10/2017

Rejection notification on 16/01/2018

ICT Prioritisation Panel 11/01/2018

Pl's response is submitted 13/12/2017

08/12/2017 Reviews sent to PI

01/01/2018
Applicant Details

Applicant | Professor Min Chen
---|---
Organisation | University of Oxford

Title of Research Project

Theory of Visualization: From an Information-Theoretic Framework to an Information-Theoretic System

Review Information

Response Due Date | 06/11/2017
Reviewer Reference | 127832747

Quality

Primary criterion. Please comment on the degree of research excellence of the proposal, making reference to:
(1) The novelty, relationship to the context, and timeliness;
(2) The ambition, adventure, and transformative aspects identified;
(3) The appropriateness of the proposed methodology.
(For multi-disciplinary proposals please state which aspects of the proposal you feel qualified to assess)

Since the late 1980s, visualisation of data has become a field that has led to remarkable advances in computer science -- in both hardware and software -- and to remarkable achievements in its many scientific and social applications. The subject is essential for the recently created world of Big Data applications, from which its technical development and its impact will certainly prosper. Initiatives in eScience and eResearch, together with high performance computing, are testimony to the recognition of the maturity, potential and timeliness of data visualisation. However, the theoretical understanding of the field of data visualisation is neglected and immature in comparison. Yet visualisation poses intriguing and pressing questions about its fundamental nature and has a need for well founded explanations of practical techniques. The researchers have identified a range of important foundational questions and have proposed a novel, timely and completely viable project to explore them.

The aims of the project are ambitious, adventurous and transformative. They propose a deep and wide analysis of practical techniques and heuristics in visualisation, which they will make precise and, in particular, analyse using the concepts and methods of Information Theory. The scale of the work is ambitious and the depth of the theoretical analysis is without precedent in computer graphics. That visualisation is very broadly conceived will add to the adventure.

The researchers have developed initial methods and results that demonstrate the viability; they have been published and expounded in state of the art tutorials at top conferences. The methodology is appropriate.
Importance

Secondary major criterion. Comment on the national importance of the research. How it:
(1) Contributes to/helps maintain the health of other disciplines, contributes to addressing key UK societal challenges and/or contributes to future UK economic success and development of emerging industry(s);
(2) Meets national needs by establishing/maintaining a unique world leading activity;
(3) Complements other UK research funded in the area, including any relationship to the EPSRC portfolio.

Data science is a major target for UK government and commercial investment. One thinks of the establishment of the Alan Turing Institute which is devoted to data science, for example.

There are many detailed expositions of the importance of data to the UK that I expect are well known: manifestos and mission statements abound -- open data initiatives of the Government Digital Services, Health Data UK, security sector, Google and the start-ups. Visualisation is integral to progress in all of these areas.

The work expands theoretical computer science, in which the UK excels. It does this in two ways: (i) by pioneering theoretical studies of visualisation and data, and (ii) by the challenge of applying and adapting Information Theory to a new area. In expanding the scope of theoretical computer science it certainly qualifies as a new development in which the UK leads.

I think the work is relevant for the EPSRC Information Economy strategy, and to themes Information and Communication Technologies (ICT), Digital Economy, Manufacturing the Future, Healthcare Technologies, Cyber Security.

Impact

Secondary criterion. Please comment on the pathway to impact identified for this work, particularly:
(1) How complete and realistic are the impacts identified for this work;
(2) The effectiveness of the activities identified to help realise these impacts, including the resources requested for this purpose;
(3) The relevance and appropriateness of any beneficiaries or collaborators.

Data visualisation has a close relation with the world's work and cares. Progress in visualisation quickly impacts on its users.

The project is truly foundational and the natural pathway to impact is via its impact on data visualisation. The account presented explains this clearly and realistically.

Applicant

Secondary criterion. Please comment on the applicant's ability to deliver the proposed project, making reference to:
(1) Appropriateness of the track record of the applicant(s);
(2) Balance of skills of the project team, including academic partners.

The applicant is a first rate computer scientist with a vast range of experience and many technical achievements. Chen is one of the top visual computing scientists in Europe.

Most importantly, the proposal is based on truly novel and exciting work by the applicant, making surprising foundational discoveries in data visualisation.

The partners involved are experienced co-workers of the applicant who have worked with him on the initial first steps of his information-theoretic derivation of visualisation theory.
Resources and Management

Secondary criterion. Please comment on the effectiveness of the proposed planning and management and on whether the requested resources are appropriate and have been fully justified. Please comment explicitly on the viability of the arrangements described to access equipment needed for this project, and particularly on any university or third party contribution.

The proposal request reasonable funds for experienced scientists to meet regularly to work together.

It is sound and easily implemented with little risk.

Overall Assessment

Please summarise your view of this proposal

This is an excellent proposal in all respects.

The science is very original, it has great potential to be transformative of data visualisation, and to have impact on the world.

I welcome the fact that, in this pioneering project, reasonable funds are requested for experienced people to work together intensively for some years on a seriously challenging and exciting problem.

I expect it to be highly productive of excellent research -- and excellent value for money.

My judgement is that:
1) This proposal is scientifically or technically flawed
2) This proposal does not meet one or more of the assessment criteria
3) This proposal meets all assessment criteria but with clear weaknesses
4) This is a good proposal that meets all assessment criteria but with minor weaknesses
5) This is a strong proposal that broadly meets all assessment criteria
6) This is a very strong proposal that fully meets all assessment criteria

My confidence level in assessing this is:

1 2 3 4 5 6 Low Medium High

Reviewer Expertise

Please indicate your areas of expertise that are relevant to your assessment. Take care not to reveal your identity to the applicant.

I am a theoretical computer scientist.

I have research publications in many areas of computer science (e.g., data, programming language semantics, visual computing, hardware design, parallel computing, analogue-digital systems) and their applications.
Applicant Details
Applicant | Professor Min Chen  
Organisation | University of Oxford

Title of Research Project
Theory of Visualization: From an Information-Theoretic Framework to an Information-Theoretic System

Review Information
Response Due Date | 10/11/2017  
Reviewer Reference: | 064463485

Quality
Primary criterion. Please comment on the degree of research excellence of the proposal, making reference to:
(1) The novelty, relationship to the context, and timeliness;
(2) The ambition, adventure, and transformative aspects identified;
(3) The appropriateness of the proposed methodology.
(For multi-disciplinary proposals please state which aspects of the proposal you feel qualified to assess)

The need for underpinning frameworks based on sound, well-grounded theoretical systems is essential for the development of any scientific discipline and visualisation is no exception. As a growing and maturing field, such well-grounded frameworks are vital if the discipline is to thrive and have wider impact on other academic disciplines and the wider society. The proposed project promises to take some fundamental steps in developing such an underpinning, unifying theory for the field of visualisation. This is an highly important and ambitious goal, and the successful delivery of the project objectives will have serious impact not only for core visualisation research but will also strengthen the position of visualisation research in the wider domain of Data Science.

The proposed research agenda builds on the earlier research outcomes of the PI on developing an information theoretical perspective on the visualisation pipeline. The outlined project plan and the stated objectives are in very good alignment with the stated aim of developing a theoretical system for visualisation grounded on information theoretical concepts. The stages of the proposed plan are viable and well informed, and a detailed plan to deliver each objective is presented.

One aspect to ensure the potential impact of the project's outcomes would be to develop guidelines and clear pointers on how the developed system can be adopted and utilised by visualisation researchers and practitioners. This can be done through example case studies, follow-up PhD projects, or review studies that position the current literature within the context of the developed framework.
Another question is related to developing the human-related parts of the framework. The suitability of information theory to more tangible aspects of visualisation, such as data or visual representations, is clear and evidenced by the earlier work of the PI. However, when it comes to the human side of the "communication through visualisation", namely, interaction, sense-making, perception, or cognition, looking through an information theoretical lens will be less straightforward. One idea could be to turn to theories of cognition, sense-making or perception and build bridges with information theory. It would be really interesting to see how this project could develop such bridges, however the current agenda do not aim to go too much in depth on those aspects. This might be a topic of a next, follow-up project or some other side projects, but my recommendation would be to consider these aspects as well.

**Importance**

**Secondary major criterion. Comment on the national importance of the research. How it:**

1. **Contributes to/helps maintain the health of other disciplines, contributes to addressing key UK societal challenges and/or contributes to future UK economic success and development of emerging industry(s);**
2. **Meets national needs by establishing/maintaining a unique world leading activity;**
3. **Complements other UK research funded in the area, including any relationship to the EPSRC portfolio.**

Visualisation is a maturing field with growing impact on the current data-intensive society. Data science is only one major and important field where visualisation has an important role, in particular in addressing the communicability, interpretability and the accountability of algorithmic approaches. The UK has a stable and growing community in visualisation and developing a theoretical system, as proposed in this project, will strengthen the position of UK in the research discipline. The industrial market benefiting from visualisation is also constantly growing in the UK and the US, and the potential applications of the framework could be economically significant in the longer run.

**Impact**

**Secondary criterion. Please comment on the pathway to impact identified for this work, particularly:**

1. **How complete and realistic are the impacts identified for this work;**
2. **The effectiveness of the activities identified to help realise these impacts, including the resources requested for this purpose;**
3. **The relevance and appropriateness of any beneficiaries or collaborators.**

The project aims to deliver impact through several academic papers and the proposed outcomes are all at the most impactful venues and journals. The PI has already organised workshops and panels at IEEE VIS and the Alan Turing Institute and follow-up activities along similar lines could strengthen the potential impact. Being an overseas travel grant, such activities are rightly not costed but the project team has already demonstrated that good academic impact can be generated with such workshops that are not necessarily funded.

Being a theoretical project, the more practical impact could take a while to develop but as long as the project team provides clear guidance for further development, the likelihood of this happening is high.

**Applicant**

**Secondary criterion. Please comment on the applicant's ability to deliver the proposed project, making reference to:**

1. **Appropriateness of the track record of the applicant(s);**
2. **Balance of skills of the project team, including academic partners.**

The PI has an excellent track record and one of the pioneers in the field of visualisation research. Through a series of papers, workshops and panels, the PI has already paved the way for an information theory based research agenda and has generated interest and a community of researchers working on developing theoretical frameworks for visualisation. With his experience and vision in this area, the PI is in the best position to deliver this important research goal.
The collaborators on the project are all well-known experts in their area and come with a great depth of knowledge, in particular, in the area of information theory - which is essential for the project. While Prof. Golan will contribute significantly on the theoretical foundations, Prof. Sbert will contribute not only on theoretical aspects but also can bring in his expertise in applications of information theory in computer graphics and visualisation.

The team has a very good record of working together in the recent years which further ensures the potential success of the collaboration. The recent papers and books also demonstrate that the team could deliver high quality research without the need for researchers working on the project.

Resources and Management

Secondary criterion. Please comment on the effectiveness of the proposed planning and management and on whether the requested resources are appropriate and have been fully justified. Please comment explicitly on the viability of the arrangements described to access equipment needed for this project, and particularly on any university or third party contribution.

This is a modestly costed project and all the costs are justified. The partners are also contributing to the hosting of the PI which makes this an even better value project.

Given that no funding is requested for research personnel, some additional support from the host institution of the PI could be significantly instrumental for this promising endeavour. Given the duration of the proposal, I wonder whether the host institution could be "recommended" to fund a PhD student who can specifically work in relation to this project.

Overall Assessment

Please summarise your view of this proposal

Overall, this is an ambitious and scientifically highly important project. Developing sound theoretical foundations for visualisation is a crucial aim with significant impact on the development of the field as a whole. Even though information theory might not be able to address all the aspects of the visualisation pipeline and a "unified theory of all" could not be achieved, the project will surely deliver important contributions that will take the field closer to this aim. The PI is one of the key researchers in the field and together with the team, it is highly likely that the project is will succeed in many ways, hence the overall proposal is a very good value for the funding requested.

My judgement is that:

1) This proposal is scientifically or technically flawed
2) This proposal does not meet one or more of the assessment criteria
3) This proposal meets all assessment criteria but with clear weaknesses
4) This is a good proposal that meets all assessment criteria but with minor weaknesses
5) This is a strong proposal that broadly meets all assessment criteria
6) This is a very strong proposal that fully meets all assessment criteria

My confidence level in assessing this is:

[ ] Low [ ] Medium [ ] High

Reviewer Expertise

Please indicate your areas of expertise that are relevant to your assessment. Take care not to reveal your identity to the applicant.

My expertise is mostly on visualisation and visual analytics with an emphasis on applications in data science and machine learning.
learning, in addition I have done some work on Information Thoery which is of importance for the project.
Applicant Details

Applicant | Professor Min Chen | Organisation | University of Oxford

Title of Research Project

Theory of Visualization: From an Information-Theoretic Framework to an Information-Theoretic System

Review Information

Response Due Date: 24/11/2017 | Reviewer Reference: 115469128

Quality

Primary criterion. Please comment on the degree of research excellence of the proposal, making reference to:

1. The novelty, relationship to the context, and timeliness;
2. The ambition, adventure, and transformative aspects identified;
3. The appropriateness of the proposed methodology.

(For multi-disciplinary proposals please state which aspects of the proposal you feel qualified to assess)

Based on the current publications in this field, and acknowledging the pace of research in this area in recent years it is clear that this research proposal is timely, and fills a niche in current thinking. The proposal itself contains a considerable number of appropriate supporting references, which is reassuring.

The project proposed is ambitious, in that what is being attempted through the visits planned goes beyond the scope one would normally expect of such a project. The visits planned would support the completion of novel and valuable research in this area. The proposed methodology is on the whole sensible, based on the PI's existing published research approaches. The justification for the methodology relating to objective 3 is however quite weak. Therefore the novelty, timeliness and ambition of the proposal are considerable strengths, with one minor issue within the methodology.

Importance

Secondary major criterion. Comment on the national importance of the research. How it:

1. Contributes to helps maintain the health of other disciplines, contributes to addressing key UK societal challenges and/or contributes to future UK economic success and development of emerging industry(s);
2. Meets national needs by establishing/maintaining a unique world leading activity;
3. Complements other UK research funded in the area, including any relationship to the EPSRC portfolio.

The successful conclusion of the project proposed would in my opinion be of considerable importance to the country, and

The request was sent on 14/11/2017. Not sure why the two dates are inconsistent.
would potentially allow the UK to gain a lead in this area; something the applicant himself alludes to in a number of places within their proposal. While reference is not made explicitly to the EPSRC portfolio I believe this proposal sits firmly in support of the Digital Economy Theme; more specifically research into ICT & HCI. I believe the research has the potential to impact on a number of current projects, both, EPSRC and beyond, however more effort could usefully have been made within the proposal to identify current EPSRC projects which could benefit from the outcomes of the proposed interactions. I do not view this as a significant issue for funding though.

Impact

Secondary criterion. Please comment on the pathway to impact identified for this work, particularly:
(1) How complete and realistic are the impacts identified for this work;
(2) The effectiveness of the activities identified to help realise these impacts, including the resources requested for this purpose;
(3) The relevance and appropriateness of any beneficiaries or collaborators.

The project as proposed would lead to sufficient impact to be worthwhile, and those impacts have been suitably outlined. The outlined pathway to impact identified for this project however is weaker. The main limitation to the current pathway to impact is its generic nature, not taking into account how the realisation of the impacts outlined may require different pathways, or pathways beyond those provided by research papers and speeches at research events in general. The applicant makes reference to the difficulties associated with convincing outsiders of the merits of the approach when discussing impact, but does not propose a long term pathway to appropriately address it.

While it is clear the applicant takes part in many events I believe some of the space used to outline the applicant's current upcoming speeches could more usefully have been used to address the area noted above.

Applicant

Secondary criterion. Please comment on the applicant's ability to deliver the proposed project, making reference to:
(1) Appropriateness of the track record of the applicant(s);
(2) Balance of skills of the project team, including academic partners.

The track record of the applicant is appropriate and the academic partners associated with the proposal appear valuable to the project, supportive and enthusiastic. I see the applicant's track record as a key strength to the proposal as it stands.

Resources and Management

Secondary criterion. Please comment on the effectiveness of the proposed planning and management and on whether the requested resources are appropriate and have been fully justified. Please comment explicitly on the viability of the arrangements described to access equipment needed for this project, and particularly on any university or third party contribution.

Considering the contents of the work-plan and the justification of resource documents together, each of the travel items has sufficient justification, with the resources requested for each justifiable. The relatively light approach to documenting management activities within the proposal is not a significant concern given the scope of the proposed research activities, and the track record of the PI. The supporting letters from the two institutions involved in hosting visits indicate an appropriate level of support, in terms of meetings spaces and time to ensure successful undertaking of the project. There is no specialist equipment requirement identified, which helps minimise risk.

Overall Assessment

Please summarise your view of this proposal

In assessing this proposal I have placed greater weighting on quality, importance of the research, and the track record of
the applicant than I have on the pathway to impact as currently defined. In summary I believe this proposal would be a worthwhile investment for the EPSRC and as a reviewer I am supportive of it. The weaknesses that do exist within the proposal are in relation to the pathway for impact and to a much smaller extent within the justification for the methodology and linking to current projects.

My judgement is that:

1) This proposal is scientifically or technically flawed
2) This proposal does not meet one or more of the assessment criteria
3) This proposal meets all assessment criteria but with clear weaknesses
4) This is a good proposal that meets all assessment criteria but with minor weaknesses
5) This is a strong proposal that broadly meets all assessment criteria
6) This is a very strong proposal that fully meets all assessment criteria

My confidence level in assessing this is:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td></td>
<td>Medium</td>
<td></td>
<td>High</td>
<td></td>
</tr>
</tbody>
</table>

Reviewer Expertise

Please indicate your areas of expertise that are relevant to your assessment. Take care not to reveal your identity to the applicant.

I have worked for over 10 years in the area of information modelling, with particular emphasis on graphical notations, and the use of ontologies in support of formalising the relationships between concepts.
Title of Research Project
Theory of Visualization: From an Information-Theoretic Framework to an Information-Theoretic System

Review Information
Response Due Date: 24/11/2017  
Reviewer Reference: 063376429

Quality

Primary criterion. Please comment on the degree of research excellence of the proposal, making reference to:

1. The novelty, relationship to the context, and timeliness;
2. The ambition, adventure, and transformative aspects identified;
3. The appropriateness of the proposed methodology.

(For multi-disciplinary proposals please state which aspects of the proposal you feel qualified to assess)

This proposal sets out to provide a potentially transformative information-theoretic system in the context of information visualization.

The PI does an excellent job of conveying the timeliness of the proposed research. There is now enormous practical usage of information visualization techniques which has really exploded over the last 15 or 20 years. Currently there is substantial research interest in providing both new visualization methods and understanding of what makes them effective, or 'optimal' to use the PI's phrasing from figure 3 in the proposal. Despite this practical and research interest, there is no unified system that allows their rigorous investigation. Providing insight into how to develop and understand qualities of information visualization techniques more effectively, through the proposed information-theoretic system, will be a clear advantage to the entire visualization community.

As for the research itself, it is without doubt both ambitious and challenging. I agree with the PI's comments that it would be costly and time-consuming to train PDRAs to undertake the proposed research project. It is clearly difficult and requires a broad, deep and advanced understanding of the research landscape across all areas related to visualization research. This spans mathematics and cognitive sciences, whilst keeping a keen eye on a huge range of application areas to ensure sufficient generality of the results.

The methodology seems appropriate for the planned research, involving partners with both theoretical and applied
Importance

Secondary major criterion. Comment on the national importance of the research. How it:
(1) Contributes to/helps maintain the health of other disciplines, contributes to addressing key UK societal challenges and/or contributes to future UK economic success and development of emerging industry(s);
(2) Meets national needs by establishing/maintaining a unique world leading activity;
(3) Complements other UK research funded in the area, including any relationship to the EPSRC portfolio.

There is ample scope for other disciplines to benefit from the proposed work, albeit indirectly. Data visualization arises across numerous disciplines - the importance of having access to effective visualizations is clearly evident. Being able to exploit data has economic advantages, as well as human ones. Effective visualization can reveal insights that would otherwise be opaque. If we have a better understanding of what makes a visualization effective, we will be able to design new visualization methods that are fit-for-purpose more quickly, analyse them more systematically, and ultimately allow their end-users to better understand their underlying data. These comments should indicate that there is clear value in devising a unified information-theoretic system that allows 'optimal' visual representations to be produced.

The proposal describes the UK's leading position in this area, placing it in the broad global context. Indeed, much of the leading work can be attributed to the PI, which will be discussed in the Applicant section. Funding this proposal will help to maintain the UK's leading position in this area.

Unfortunately, the proposal does not discuss the national picture in any great depth, particularly in relation to the EPSRC portfolio, beyond the PIs own contributions.

Impact

Secondary criterion. Please comment on the pathway to impact identified for this work, particularly:
(1) How complete and realistic are the impacts identified for this work;
(2) The effectiveness of the activities identified to help realise these impacts, including the resources requested for this purpose;
(3) The relevance and appropriateness of any beneficiaries or collaborators.

The pathways to impact described in the proposal are entirely sensible, although the plan to deliver impact could be criticised for being a bit 'traditional': publications and talks. However, one must be cognisent of the fact that this is a proposal for an overseas travel grant and not a standard three-year responsive mode project. Inevitably, this limits what it is possible to achieve in terms of impact activities. I think the big challenge for the PI will be ensuring that the research community, in the long term, makes use of the results. In this regard, I think the PI's standing in the field is particularly important - being such a respected figure provides a platform for publicising and promoting the work. This gives confidence that it is realistic for the results to have the envisaged impact.

The resources requested -- conference expenses -- are appropriate for the impact plan, although I was a little surprised by
the high number of conference trips planned given the (roughly) ten months of PI time on the project. Perhaps this could have been a little better justified.

Regarding the appropriateness of the Partners, this is covered under the Applicant section.

**Applicant**

*Secondary criterion. Please comment on the applicant's ability to deliver the proposed project, making reference to:*

1. Appropriateateness of the track record of the applicant(s);
2. Balance of skills of the project team, including academic partners.

**Resources and Management**

*Secondary criterion. Please comment on the effectiveness of the proposed planning and management and on whether the requested resources are appropriate and have been fully justified. Please comment explicitly on the viability of the arrangements described to access equipment needed for this project, and particularly on any university or third party contribution.*

**Overall Assessment**

*Please summarise your view of this proposal*

This is a strong proposal from a PI with an outstanding track record. The research is ambitious and has the potential to be transformative. I believe a proposal of this kind - for the PI to undertake the research rather than a PDRA - is ideally suited due to the highly skilled nature of the work. I’d like to have seen the proposal cover more background and context beyond
the PIs own contributions, but this is a minor criticism and should not detract from the overall quality of this proposal. The project team as a whole is well thought out and the Partners have provided highly supportive letters conveying real excitement for the proposed research and collaboration.

In summary, I support the award of funding and look forward to seeing this research evolve.

My judgement is that:
1) This proposal is scientifically or technically flawed
2) This proposal does not meet one or more of the assessment criteria
3) This proposal meets all assessment criteria but with clear weaknesses
4) This is a good proposal that meets all assessment criteria but with minor weaknesses
5) This is a strong proposal that broadly meets all assessment criteria
6) This is a very strong proposal that fully meets all assessment criteria

My confidence level in assessing this is: Low Medium High

Reviewer Expertise

Please indicate your areas of expertise that are relevant to your assessment. Take care not to reveal your identity to the applicant.

Expertise in information visualization and inference problems. Knowledgable of cognitive and perceptual theories relevant in the context of visualization.
Applicant Professor Min Chen
Organisation University of Oxford

Title of Research Project
Theory of Visualization: From an Information-Theoretic Framework to an Information-Theoretic System

Review Information
Response Due Date 04/12/2017
Reviewer Reference: 169803099

Quality

Primary criterion. Please comment on the degree of research excellence of the proposal, making reference to:

(1) The novelty, relationship to the context, and timeliness;
(2) The ambition, adventure, and transformative aspects identified;
(3) The appropriateness of the proposed methodology.

(For multi-disciplinary proposals please state which aspects of the proposal you feel qualified to assess)

The proposed strategy is to seek to axiomatise visualisation, using results from information theory.

I do not believe this is a good strategy. The attempts at formal axiomatisation have not so far, in my view, brought useful concrete insights.

I see no promise in the proposed strategy of examining many theorems of information theory to see which ones might be applicable to visualisation.

I had never encountered the applicant's work before, so I read two of the applicant's previous publications that are prominently cited in the proposal: [mc1] and [mc5]. I do not believe that either of these publications contain useful or insightful results about visualisation.

The general idea of analysing a visualisation using information theory is not necessarily wrong: it is desirable, for example, that relevant distinctions in the data should be mapped to visible and interpretable distinctions in the visualisation. It is also desirable to quantify the accuracy with which features of the data may be read from a visualisation - but this idea is hardly new. It was introduced by William Cleveland at Bell Labs in the 1980s, and probably also before that.
However, the formalisms proposed in [mc1] and [mc5] seem too vague and abstract to be usefully applicable in practice. There is surely general agreement that a visualisation should map data to a representation that humans can interpret, in a sense that the user can read the representation to determine facts about the data.

However, information theory seems too abstract a tool to be useful here: the engineering requirement of a visualisation is to map particular characteristics of the data to visible features in a way that human users can interpret correctly. I feel that the applicant has not yet demonstrated that the abstractions proposed would add any insight over and above a clear definition of visualisation requirements.

Importance

Secondary major criterion. Comment on the national importance of the research. How it:
(1) Contributes to/helps maintain the health of other disciplines, contributes to addressing key UK societal challenges and/or contributes to future UK economic success and development of emerging industry(s);
(2) Meets national needs by establishing/maintaining a unique world leading activity;
(3) Complements other UK research funded in the area, including any relationship to the EPSRC portfolio.

As I have explained in the section on quality, I do not believe that this research will lead to important effects.

Impact

Secondary criterion. Please comment on the pathway to impact identified for this work, particularly:
(1) How complete and realistic are the impacts identified for this work;
(2) The effectiveness of the activities identified to help realise these impacts, including the resources requested for this purpose;
(3) The relevance and appropriateness of any beneficiaries or collaborators.

As explained in the section on quality, I do not believe that this research will lead to impact.

The aim of the research is to develop formalisms that, in my opinion, will be too abstract and vaguely defined to be of practical use.

Applicant

Secondary criterion. Please comment on the applicant’s ability to deliver the proposed project, making reference to:
(1) Appropriateness of the track record of the applicant(s);
(2) Balance of skills of the project team, including academic partners.

The applicant has a substantial track record of publications in the area of visualisation.

He has also done practical work in devising metrics for assessing visualisations.
His research track record is therefore appropriate for the research proposed.

Resources and Management

Secondary criterion. Please comment on the effectiveness of the proposed planning and management and on whether the requested resources are appropriate and have been fully justified. Please comment explicitly on the viability of the arrangements described to access equipment needed for this project, and particularly on any university or third party contribution.

The resources requested for this theoretical work are reasonable for the project as it is specified.

Overall Assessment

Please summarise your view of this proposal

The proposal is to develop a general abstract formalism to axiomatise data visualisation. I do not believe that the formalism proposed is either well specified or useful. I could of course be wrong.

My judgement is that:

1) This proposal is scientifically or technically flawed
2) This proposal does not meet one or more of the assessment criteria
3) This proposal meets all assessment criteria but with clear weaknesses
4) This is a good proposal that meets all assessment criteria but with minor weaknesses
5) This is a strong proposal that broadly meets all assessment criteria
6) This is a very strong proposal that fully meets all assessment criteria

My confidence level in assessing this is:

1 2 3 4 5 6

Low Medium High

Reviewer Expertise

Please indicate your areas of expertise that are relevant to your assessment. Take care not to reveal your identity to the applicant.


The high confidence rating and the statement about the reviewer's expertise (e.g., information theory plus visualization) were clearly intended to mislead the panel. This seriously undermines the integrity of the review process.
The PI’s Responses to the Reviews on the Application EP/R029989/1 for an Overseas Travel Grant

**Theory of Visualization:**
From an Information-Theoretic Framework to an Information-Theoretic System
PI: Professor Min Chen, Department of Engineering Science, University of Oxford

Reviews 063376429, 064463485, 115469128, and 127832747

The PI is humbled by the positive comments about the research quality, importance, potential impact, and the applicant’s ability, and appreciates the unmistakeable support from these four reviewers. The PI will focus below on their suggestions for improving the proposed research.

063376429: I’d like to have seen ... more direct references to the work of other researchers

This is partly due to the page limit of the proposal and partly due to the PI’s unconscious bias in his plan to build a theoretic system based on information theory. Most previous research efforts on information theory in visualization, noticeable by Mateu Sbert’s team in Spain and Han-wei Shen’s team in the USA, have focused on using the metrics such as Shannon entropy and mutual information to measure variances within a volume dataset, a vector field, or a visualization image. The use of information theory in these works was to provide practical metrics rather than to model visualization processes theoretically. In addition to the theoretical works by the PI and his colleagues, Kindlmann and Scheidegger (TVCG 2014) used algebra to model visualization processes. The weakness of this approach was discussed in [mc17]. The PI always cites these previous works extensively in his research papers.

063376429: Unfortunately, the proposal does not discuss the national picture in any great depth, particularly in relation to the EPSRC portfolio, beyond the PI’s own contributions.

The PI agrees that the impact document could have discussed the national picture. The visualization research activities in the UK are sparsely distributed in many universities. In addition to Oxford, the universities that have published in the top visualization venues in recent years include Bangor, Edinburgh Napier, Leeds, London (City, King’s, Middlesex), and Swansea. Considering the very limited amount of funding received by these visualization groups, the UK is doing well. For example, among the 3 best papers and 6 honourable mentions in VIS2016, the UK had 2 best papers and 1 honourable mention. However, in terms of the total number of papers per country, the UK is usually placed behind the USA, Germany, China, Austria, and the Netherlands, joining a group that also includes France, Norway, Canada, and Hong Kong.

064463485: One aspect to ensure the potential impact of the project’s outcomes would be to develop guidelines and clear pointers ...

The PI agrees with this suggestion. In addition to building a theoretic system, the PI is following two other directions suggested in the position paper [1]: setting up a guideline discourse web site, and building an ontology for visual analytics. As the funding for these two directions is more readily available in Konstanz, the PI is acting as an external adviser to a PDRA (on guidelines) and a PhD student (on ontology) there.

064463485: Another question is related to developing the human-related parts of the framework. The PI agrees with this suggestion. Any theory that can explain how visualization works has the potential to explain some phenomena in cognitive science. The PI has tentatively suggested such a possibility in a short video at https://vimeo.com/227948927, hoping to stimulate more research efforts in this direction. The PI will also give a keynote at the HCI 2018 (Belfast), discussing the connection between information theory and human-computer interaction in data intelligence.

115469128: ... more effort could usefully have been made within the proposal to identify current EPSRC projects which could benefit from the outcomes

The PI appreciates the reviewer’s suggestion to increase the impact of this theoretical research by applying the cost-benefit metric to the data intelligence workflows developed in some projects. In [mc17, mc18], the information-theoretic metric was used to analyse the cost-benefit of four past projects (from Swansea, Konstanz, and Oxford respectively). The PI and three other colleagues (in Chester and Texas, Austin) have also written a paper (under review) for analysing the cost-benefit of visualization in virtual environments including virtual reality and mixed reality. The PI will continue to explore opportunities to apply the acquired theoretical understanding to practical scenarios.
Lastly, let us address the issues raised in this staggering review. In general, the PI believes that the panel should treat this review as an anomaly or an outlier in terms of its crude review style and hasty opinions. In particular, the reviewer raised the following three issues (in the reverse order):

1. The idea of quantifying accuracy by William Cleveland in the 1980s or probably earlier.

   William Cleveland is no doubt a great pioneer of data visualization. His idea of quantifying accuracy followed the methodology of empirical studies in cognitive science. The most representative presentation of his idea is the paper by Cleveland and McGrill, “Graphical Perception: Theory, Experimentation, and Application to the Development of Graphics Model,” JASA, 1984. I am not aware of any proposal from Cleveland for quantifying the accuracy of visualization using information theory. Cleveland did apply log2 to the accuracy measure in one of his empirical studies. This might be his closest point to information theory, though he mentioned the reason as “a log scale seemed appropriate to measure relative error; ... We used log base 2 because average relative errors tended to change by factors less than 10.”

   [mc5] by Chen and Jänicke was the first paper that outlined the theoretical framework for measuring several forms of accuracy in visualization, including (i) Shannon entropy in three different aspects: data, visualization, and graphical display; (ii) information loss with a case study showing how logarithmic plots minimize information loss; and (iii) mutual information, with a case study confirming mathematically that the visual information-seeking mantra, “overview first, zoom, and details on demand” is correct.

2. [mc1] and [mc5] are neither useful nor insightful.

   The reviewer incorrectly assumed that [mc1] was “prominently cited”. In fact, [mc1] was only mentioned once in the track record section but never mentioned in the main research proposal. The other prominently cited paper is [mc17] by Chen and Golan.

   Until now, one conventional belief in visualization has been to assume that visualization must be accurate, since it would be much more difficult to suggest that visualization could be inaccurate. Most visualization scientists and practitioners, including pioneers such as Cleveland and Tufte, made passionate arguments for accuracy in their teaching. In a small way, this is a bit like the early 19th century or before, when it was much easier to assume that the earth was flat. However, many visualization researchers have privately doubted the completeness or correctness of this doctrine because it cannot explain some obvious phenomena in visualization. For example, why would a series of data points in the range of [0, 10,000] be accurate enough when it is plotted as a time series pot, from which humans can manage to distinguish perhaps no more than 500 data values per data point? How could displaying such a time series on a powerwall with 20,000x40,000 pixels hardly bring about any more benefit than a desktop display with 1,000x2,000? Why should the highly-distorted London underground map be considered as one of the best visualizations in the history?

   [mc17] by Chen and Golan provided the first mathematical formulation of the cost-benefit of data analysis and visualization processes, which can explain these phenomena.

3. Attempts at formal axiomatization have not so far brought useful concrete insights.

   There has not been any report on any attempt of axiomatization in visualization, but there have been numerous attempts in mathematics (e.g., https://en.wikipedia.org/wiki/List_of_axioms) and theoretical computer science (e.g., https://en.wikipedia.org/wiki/Turing%27s_proof). It seems to be a colossal unjust to those great forerunners by dismissing these attempts including many significant accomplishments.

   As mentioned in the proposal and a recent position paper [1], the PI is fully aware of the Gödel theorem. The goal of following a path to discover axioms in the theoretic system is to gain a theoretical understanding about what is fundamental and what can be inferred. Such a discovery is analogous to the pursuit in qualitative research for curating, evaluating, critiquing, and revising guidelines in order to discover principles as well as the pursuit in physical sciences for identifying all elementary components in a physical, chemical, or biological system. In [mc17], it is shown that the cost-benefit of visualization processes relates to only three fundamental measures. This suggests some paths for building an axiomatic system. As an esteemed colleague at STFC (Dr. Bryan Edwards) said in the 2016 ATI Symposium on Theoretical Foundation of Visual Analytics [8], “It is irresponsible for visualization scientists not to try.”

While it is hurtful to read such a review, I appreciate the acknowledgement of this reviewer: “His research track record is therefore appropriate for the research proposed” and “I could of course be wrong.”