Arguments for a Theory of Visualization

Georges Grinstein

Institute for Visualization and Perception Research, University of Massachusetts Lowell

ABSTRACT

At the IEEE Visualization Conference in 2008 on a panel I organized on Grand Challenges in Visualization I proposed that one fundamental key grand challenge, was to develop a theory of visualization. I suggest that if such a theory of visualization existed we would be able to use that theory to solve some thorny and wicked problems and that these solutions could be arrived in many different ways. I give some examples of this here and will hint at solutions for some thorny and wicked problems at the workshop.

KEYWORDS: Visualization Theory, Information Theory, Measuring Information, Grand Challenge.

INDEX TERMS:

1 INTRODUCTION

In the 2008 Grand Challenge Panel [1] I described the key fundamental grand challenge as the development of a Visualization Theory and as would have it we now are seeing more papers on such theories. However most are still rephrasings of taxonomies or pipelines, a valid start, but not really reaching the goal I had in mind. That goal is to provide measurable outcomes and be able to make predictions. A visualization theory must be falsifiable and must make predictions. We discuss in more details why we need a theory and how to go about building such a theory, and we show how such a theory could be used with simple examples. This is work in progress.

2 WHY A THEORY

Jules-Henri Poincaré said that " Science is built up of facts, as a house is built of stones; but an accumulation of facts is no more science than a heap of stones is a house." [2] Visualization is currently a collection of stones – we have lots of facts, wonderful facts; just look at all the papers at this and similar conferences. A good theory is effective and practical as it guides us toward both identifying and resolving crucial and foundational questions. This has the side effects of enhancing knowledge and educating decision makers.

Figure 1 highlights the need for such a theory and clearly the value of measurement. It describes the preliminary results from the comparison of two hypothetical clinical trials both treating the same condition: a generic conventional and an investigational treatment. Elting [3] asked which of four visualization, three classic and one more iconic, would lead to clinicians deciding to stop the trial because one of the treatments was obviously better than the other.

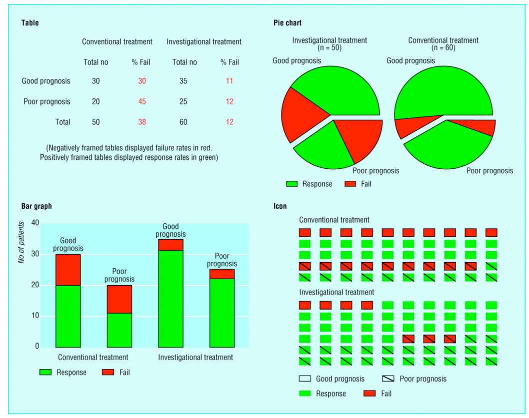

Figure 1 Visualization of a Clinical Trials experiment for a decision making evaluation [3]

The decision to stop varied significantly depending on the presentation of the data. Correct decisions were 56% with both the bar and pie charts, 68% with the table, and 82% with the icon display ($p = 0.03$). In actual practice this would mean that up to 25% of the patients treated based on the bar or pie charts would have received inappropriate treatment. Although this is a low power experiment (34 physicians) it does highlight the value of a theory if it could predict such outcomes, especially if such predictions were independent of the data.

A theory is not a theorem nor is it right or wrong. It can be well tested and even well validated but there is not guarantee that it will not be refuted in the future. Much activity is not necessarily focused on validating a theory but on invalidating it, on refuting it, on proving it wrong. That's an easy step as all one has to do is contradict any of its predictions. Hence a theory has to predict. A theory of visualization has to make strong statements about the field, in all its aspects. It must be falsifiable via experiments or mathematics or some other mechanism. If it does not make any predictions it is useless as it is unfalsifiable. At that point it is just an opinion, a belief. [4]

3 HOW DO WE GO ABOUT BUILDING A THEORY

We must explain how, when and why the theory works. We must explain when and where it fails. We must provide a range of examples of its failings and successes. And for each we must provide many different perspectives including analysis, perception, cognition, collaboration, stability, consistency, utility, and usability, for example, each having many facets.

The goal is to discover or identify or highlight some pattern predicted by the theory, explain them, measure them, highlight their structure, define their probabilities, explore special cases, events, and more.

A theory requires that we define its fundamental abstractions, and objects, its basic operations, its structure, what is being measured and which section of the model. It should also help identify if not do so, what is hard, what is doable, why, what is a correct representation and what context it applies.

This problem is hard but we have much in favor of the timeliness of building and evaluating such a theory (or such theories). We now have a few large data sets that are publicly sharable. The problem has gathered interest to both researchers, the public and policy makers, and a solution will have significant benefit to the field [5].

4 EXAMPLE OF A THEORY'S USE

We assume the following.

Conjecture 1. Given a data set D, given a task T, for a given display, there exists a visualization V such that the perceived information I is such that task T is optimally perceptually/cognitively "resolved". This means that no other visualization will solve task T as well or that the perceived information I is the best for resolving task T.

There are lots of unknowns, undefined terms, measures to be defined or clarified. There are as well a number of dependencies. For example the perceived information I depends on the user.

However with such a Conjecture we could attempt to solve a large number of difficult problems.

For example, we could build an average perceived information measure across all users or classes of users. Here's an example of another implication.

Corollary 1. Given a data set D, given a task T, given a visualization V such that the perceptive information I is such that task T is optimally perceptually/cognitively "resolved", then no other visualization W will solve task T as well.

In other words the perceived information I is the best for solving task T with this given V. Another way to phrase this is that if m is a measure of perceptual information then for V of corollary 1 we have

Corollary 2. (1) $m(V, T) \geq m(W, T)$ for any other visualization W.

An example of an application of (1) is to generate a number of data sets with ground truth bound to task T (so that the task clearly is solved by the discovery of the ground truth) and design experiments that would identify which visualizations are capable of discovering the ground truth.

5 CONCLUSION

The biggest problem is not necessarily the development of a theory but more in its evaluation. We need to focus on metrics for information transfer at all levels of activity. The most important and likely the reason for the lateness of our developing a theory is metrics involving the user's perception, cognition, and interpretation. Another way to phrase this is "how well did the user get it?"

6 NOTE

The need for a theory was discussed at the IEEE Grand Challenge Panel [1]. The mathematical arguments for a theory are part of my visualization class lectures. This approach is now being continued with several researchers including Robert Kosara from the University of North Carolina at Charlotte and from the Air Force Research Laboratory Kristen Liggett, Paul Havig, Jason Moore, David Kaveney, Tim Lebo and Robert Patterson.

REFERENCES

[1] G. Grinstein, D. Keim, and T. Munzner, Grand Challenges for Information Visualization, IEEE VisWeek 2008.

[2] H. Poincaré, H. "Les relations entre la physique expérimentale et la physique mathématique", Revue générale des sciences pures et appliquées 11: 1163–1175 . Reprinted in "Science and Hypothesis", Ch. 9–10.

[3] L.S. Elting, C.G. Martin, S.B. Cantor, E.B. Rubenstein. Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures. British Medical Journal. 1999; 318:1527-31

[4] K. Popper, The Logic of Scientific Discovery, 1934 as Logik der Forschung, English translation 1959

[5] G. Piatetsky-Shapiro, R. Grossman, C. Djeraba, R. Feldman, L. Getoor, and M. Zaki, Is There a Grand Challenge or X-Prize for Data Mining?, ACM KDD 2006 Panel, Philadelphia