Why?

Miriah Mike (G) and Michael (S)
the doodles are MG’s fault
Why ask Why?
because they asked us to!
The Four-Level Nested Model Revisited: Blocks and Guidelines

Miriah Meyer
University of Utah
miriah@cs.utah.edu

Michael Sedlmair
University of British Columbia
msedl@cs.ubc.ca

Tamara Munzner
University of British Columbia
tmm@cs.ubc.ca

Stop The Evaluation Arms Race! A Call to Evaluate Visualization Evaluation

Position Paper
Michael Gleicher
Department of Computer Sciences
University of Wisconsin - Madison
gleicher@cs.wisc.edu

Personal Statement Summary:
While I consider myself a relative newcomer to the field of visualization, I have a longer history in other areas in which similar issues arise. My work in visualization has taken up the challenges in translating perception research into actionable design guidelines for visualization systems, which has included performing our own “low-level” evaluations to build models that inform development of visualization papers, including problem-driven design studies [7, 11, 24, 27, 29, 32], technique-driven work [18], evaluation [1, 31], models [8, 10, 17, 19, 28, 30, 35], and systems [4, 12].

We use the nested model extensively as a way to guide and reflect upon our own work, and propose an extension of the model motivated by our desire to clarify the meaning of the term guideline. This term is loosely defined within the visualization literature to describe knowledge that guides how we make design decisions. One of our goals with this work is to clarify the meaning of this term for visualization research in order to assess the impact of both design studies and technique work on guidelines.

The extension proposes blocks as a generic term for the outcomes of the design process at the three lower levels: abstractions, techniques, and algorithms. Concrete examples of blocks at each of these levels are that a network is a data abstraction block, a node-link diagram is a visual encoding block, and a specific force-directed layout approach such as GEM [13] is an algorithm block. We can then define guidelines as statements about the relationships between blocks, such as a node-link diagram is a good visual encoding of small graphs, or a specific force-directed layout algorithm is faster than another. We consider guideline and characterization to be synonyms.

Guidelines may pertain to blocks within a single level, and we call these within-level guidelines comparisons. Guidelines may also cross between levels; we call these between-level guidelines mappings to emphasize their role in moving from one level to the next. Both types of guidelines often arise from reflection after evaluation and validation efforts. Comparison guidelines are the result of putting one block against others at the same level, and often stem from validation efforts in papers that present a new technique or algorithm. Mapping guidelines provide guidance on how a block at one level is a match or mismatch with a block at an adjacent level. Mappings typically emerge from the validation of design studies. Evaluation papers may result in either kind of guideline, mappings or comparison.

The primary contribution of this paper is our proposed extension to the nested model and the resulting implications, presented in Section 2. Section 3 presents an analysis of open problems in our field illuminated by these extensions: the need to define block at the problem level, to create mid-level task taxonomies at the abstraction level and possibly refine the model itself at that level, and to establish a more complete set of mappings up from the algorithm level to the technique level. Thus, a secondary contribution is the elucidation of these gaps in our collective knowledge and a call for action to close them.
PREMISE:

CONSIDERING THE MOTIVATIONS CAN HELP US DO GOOD EVALUATION

AND GOOD EVALUATION IS REALLY OUR GOAL!
Why seek good evaluations?

because they can:

GUIDE

and PERSUADE

and other things we won't talk about now
GUIDE

How well do you inform the AUDIENCE to do some THING?

non-prescriptive \[\rightarrow\] ACTIONABLE
PERSUADE

How well do you convince the Audience to believe some Thing

vacuous assertion $\rightarrow$ Convincing argument
Two Independent Axes

Non-Prescriptive → Actionable

Guide

Persuade

Vacuous assertion

Compelling argument
Convince audience and tell them what to do

Audience won't believe wouldn't know what to do even if they did

Non-prescriptive → Actionable

Guide
Convince audience
But they can't do
Anything about it

Tell audience what to do
Doesn't convince them to do it

Non-Prescriptive ➔ Actionable

Guide
GUIDE
How well do you inform the audience to do something?

PERSUADE
How well do you convince the audience of something?
GUIDE
How well do you inform the audience to do some thing?
PERSUADE
How well do you convince the audience of some thing
CONTEXT
how ACTIONABLE and how PERSUASIVE depends on CONTEXT
Scientists using Tool A make more discoveries than using Tool B

**Not-Actionable**  **Actionable**
Scientists using tool A make more discoveries than using tool B

**Not-Actionable**

So what? How does this help me make better tools?

**Actionable**

Cool! I'll buy 10 copies of tool A for my lab!

**VIS Researcher**

**Biology Lab Director**
A PUNDIT ASSERTS MINIMALISM IS GOOD IN A SELF-PUBLISHED BOOK

NOT PERSUASIVE PERSUASIVE
A PUNDIT ASSERTS MINIMALISM IS GOOD IN A SELF-PUBLISHED BOOK

NOT PERSUASIVE

Show me some evidence!
Do a study - give me stats!

PERSUASIVE

Wow! He's famous - he must know what he's talking about
And he's a good writer too!

VIS RESEARCHER

NORMAL PERSON
Evaluating Evaluations

How well do they guide?

How well do they persuade?
Making **GOOD** Evaluations

Make it **ACTIONABLE**

Make it **PERSUASIVE**

To the **TARGET AUDIENCE**
How to make **Persuasive** Evaluations?

- Measure the right things
- Design good experiments
- Report it in a way that convinces the audience

Sophisticated methods may be harder to report
How to make Actionable Evaluations?
NESTED MODEL

- Domain problem characterization
- Data/task abstraction design
- Encoding/interaction technique design
- Algorithm design

Munzner 2009
NESTED MODEL

domain problem characterization

data/task abstraction design

encoding/interaction technique design

algorithm design

EXTENSION

domain problem

abstraction

technique

algorithm

blocks
outcome of a design decision

Munzner 2009
NESTED MODEL

domain problem characterization

data/task abstraction design

encoding/interaction technique design

algorithm design

Munzner 2009

EXTENSION

domain problem

abstraction

encoding/interaction technique

directed graph

node-link diagram

force-directed layout

blocks

outcome of a design decision
NESTED MODEL

domain problem characterization

data/task abstraction design

encoding/interaction technique design

algorithm design

blocks

outcome of a design decision

EXTENSION

domain problem

abstraction

technique

algorithm

Munzner 2009
**NESTED MODEL**

- Domain problem characterization
- Data/task abstraction design
- Encoding/interaction technique design
- Algorithm design

Munzner 2009

**EXTENSION**

**blocks**
**guidelines**

Statement about relationship between blocks
NESTED MODEL

- domain problem characterization
- data/task abstraction design
- encoding/interaction technique design
- algorithm design

Munzner 2009

blocks guidelines

categorical data

good for categorical data

hue colormap

hue colormap appropriate

EXTENSION

domain problem

abstraction

technique

algorithm
NESTED MODEL

- domain problem characterization
- data/task abstraction design
- encoding/interaction technique design
- algorithm design

Munzner 2009

blocks guidelines

faster Voronoi treemap

EXTENSION

- domain problem
- abstraction
- technique
- algorithm

Nocaj 2012
Balzer 2005
NESTED MODEL

domain problem characterization

- data/task abstraction design

- encoding/interaction technique design

- algorithm design

Munzner 2009

blocks

guidelines

between-level mapping

within-level comparison

EXTENSION

blocks

guidelines

between-level mapping

within-level comparison
implications
mapping problems to abstractions
mapping algorithms to techniques
reporting algorithm comparisons

blocks

guidelines

between-level mapping

within-level comparison
persuade & guide
blocks & guidelines
vocabulary

persuade & guide

actionable & convincing

blocks

guidelines

mappings & comparisons
DISCUSSION IDEAS

Are these useful concepts? Are these the right words for them? Are there other useful axes to describe evaluations?

How do we make evaluations actionable? Is this the right goal? Are there other types of actionability besides guidelines?

What evaluation methods to use for which contexts? What does the diversity of contexts mean for method development? Are all goals and contexts being served?

Is the visualization community too focused on itself as a context? Do our evaluations help anyone other than us? Should they?