Exploring human cognition through multivariate data visualization

by

Caitlyn M. McColeman

M.A., Simon Fraser University, 2012

B.A. (Hons), York University, 2010

Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in the

Department of Psychology

Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY

Fall 2017

Approval

Name:	**Caitlyn M. McColeman**
Degree:	**Doctor of Philosophy (Psychology)**
Title:	**Exploring Human Cognition Through Multivariate Data Visualization**
Examining Committee:	**Chair: Dr. Timothy Racine** Professor **Dr. Mark Blair** Senior Supervisor Associate Professor **Dr. Thomas Spalek** Supervisor Professor **Dr. Steve DiPaola** Supervisor Professor **Dr. Ashley Farris-Trimble** Internal Examiner Assistant Professor Department of Linguistics **Dr. John Paul Minda** External Examiner Associate Professor Department of Psychology The University of Western Ontario
Date Defended/Approved:	November 3, 2017

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

Entire disciplines are dedicated to separately exploring the relationship between sensation and perception; attention and learning; and information access and decision making. This work aims to bridge these fields though studies of data visualizations and decision making. A data visualization communicates information about synthesized data points for an observer. For graphical communication to work, all parties involved must understand regularities in the representations that are being used. Extracting regularities from observations is in the category learning wheelhouse, and so methods and findings from categorization literature are used to inform this work. Through the following experiments, the perception of multivariate data via visualization is explored. The framework for this exploration is an extension of existing proposals for a science of data visualization. The present work extends existing proposals by adding decision making as a critical element for a science of visualization. It's great to understand how people can read a graph, but it's even more informative to understand how that reading influences their actions.

Keywords: Cognition; decision making; perception; salience; data visualization

Dedication

I dedicate this work to my grandparents, who taught me, among other things, that a good investment is a piece of land and a good work ethic. They have always taught by example the value of honesty, hard work and integrity. They don't need the results of this research, but they made it all possible. Thanks Gramma and Poppa.

Acknowledgements

I would not be writing an acknowledgement section for a Ph.D. dissertation without the strength of a small army at my back. Probably 100 research assistants have touched my work in some way, shape or form in the last seven-and-a-bit years. The experiments in this paper were supported by Erica Mulder, Yue Chen, Cal Woodruff, Ruilin, Judi Azmand, David Zhang, and Joe Thompson who all came in one rainy Sunday to help me review all of the experiment code. Robin Barrett, Leif Swanson, and Jordan Barnes also contributed significantly to code review. Rajan Hayre, Katerina Dolguikh, Killian Kleffner, Andrea Smit, Sarah Power, Mateusz Michalik, Christian Peterson, and Sylvie Couture-Nowak were all very helpful in organizing data collection.

This project owes a large debt of gratitude to the hundreds of participants who spent an hour in a small room, staring at a bunch of graphs to help me test some hypotheses I had. Those participants were made available to me with help from the staff in the Psychology Department, namely Deb Jopling who organized the system and always offered help in getting participants in the room.

Other staff in the Psychology Department assisted me with getting from data to defence. In particular, thanks goes to Anita Turner for scheduling and organizing everything to make the defence day possible. My committee members, and external examiners who took time away from their work to assess mine are also appreciated. Especially, thanks to Dr. Tom Spalek and Dr. Steve DiPaola for their ideas, suggestions, critiques and efforts to help me improve the work presented here. Thanks also to Margaret, Bev, Iris and Ellen for their support with organizing finances, classes, administration and various other mission-critical steps through the years.

Jordan Barnes and Joe Thompson have been consistently helpful with every step of the research process over the course of my graduate students. Their help, support and friendship will always be cherished. I can only try to reciprocate what I've received in that regard.

Outside of the lab, many friends and family have been a big help to me. Hisham, for introducing me to these wacky graphs, and being a force for good in my life, is high among them. Holly, Marieve, Angela, Samuel, Heather, Mo, Julia, Kim, Matt, Rach, Hilary, Tam, Jessie, Ramsey: you've done more for me than could ever possibly fit in a

second dissertation. They are among many more who offered the laughs and kindness that makes everything turn out okay. Mom, Dad and my brother Matt (who all actually get what I do!), made it safe to take the risk of going to graduate school and taught me everything I know outside of science.

Kim Meier, Bill Chen, Aaron Ancell and Calen Walshe were there the earliest days of my scientific career, and started me off on the right track when I first got to the Cognitive Science Lab. My supervisor, Mark Blair, of course taught me the rest. I appreciate the guidance, and instruction I received from him during my time in the lab. I will carry many of those lessons for years to come. He made it safe to try new ideas in science and think genuinely outside of the box. I hope never to look back.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

People increasingly rely on data to make effective decisions. This is an advantage in business (Brynjolfsson & McElheran, 2016), and an effective strategy in a personal context, such as using food and calorie tracking data to improve dietary habits (Kruger, Blanck & Gillespie, 2006). Although unsurprising that data supports better decisions, the processes linking raw data to an observer's interpretation remain poorly understood. Testing an observer's interpretation of real data can be challenging because there's no pre-established objectively correct interpretation; however, to test data driven decision making, one option is to graph generated data that mimics the properties of real-world data, and test how different data visualizations help observers recreate the model from which the data were generated. The studies herein do just that: they explore the relationship between perception and decision making by manipulating the visual properties of graphs and measuring the impact those manipulations have on interpreting and using the information that the data visualizations convey.

Data interpretation is most accurate when working directly with numbers or single dimensional visualizations (Spence, 1990). That's increasingly unrealistic. Datasets are getting bigger and of higher dimensionality. As such, to make data-informed decisions, we need ways to simplify and aggregate datasets while losing as little meaningful information as possible. In many cases, a data visualization (e.g. a graph or table) is one stage between the raw data and an observer's decision. A data visualization, by its most inclusive definition, is the representation of information in the visual modality. Functionally, it's an attempt at efficient information communication of synthesized data points for easy and clear interpretation by an observer. Bar graphs presently reign supreme for their ease of use and clarity, but their time may soon be up (Weissgerber, Milic, Winham, & Garovic, 2015) since their clarity comes at the cost of obscuring all but one property of the data. Though the bar graph usually does well in simple cases where the only property of interest is the mean (or median, or count, etc.) of a dataset, there are many circumstances where it's less effective, as is the case with multidimensional, interrelated data.

Finance professionals use such multidimensional data. Many futures traders use candlestick charts to represent six dimensions of a stock, all of which can be correlated.

Furthermore, these correlations are thought to be genuinely meaningful, so dimension reduction is not often a helpful pursuit. The candlestick chart shows five continuous dimensions (reflecting opening, closing, highest and lowest prices and time), and one binary one (colour) that reinforces the relationship between open and close price. One strategy to gather information from a candlestick chart is to determine the colour of the binary dimension, and use that to say if the closing price was higher than the opening price that day. If so, the colour will suggest that the observer look at top of the candle to determine the closing price. This manner of serial information access has been identified before as stimulus-responsive attention (Blair, Watson, Walshe, & Maj, 2009).

Stimulus-responsive attention is traditionally elicited by a particular category type, wherein an observer learns that one part of the stimulus provides high level information about the next best place to look for more information to make the right decision. It is part of reading a candlestick chart too; in practice, much of graph perception is related to categorization (Greensmith, 2016). “Glyphs” are multidimensional objects that represent data across their different dimensions. In the context of the candlestick plots, five dimensions are shown in a single glyph, and six are shown in candlestick time series. Learning to read glyphs, like learning to identify category members, involves learning where to deploy attention and learning the relationship between the physical reflection of a dimension and its abstract meaning.

Considering the relationship between extracting information from visualizations and category learning, the knowledge accumulated over decades of categorization research can be extended and applied to data visualization. In doing so, this study aims to apply decades of productive research on human cognition to support effective data driven decision making. While the application of this work is to improve visualization practice, the intellectual contribution of the work is to better understand how abstract information is understood by human observers. The studies connect manipulations of visual perception to overt performance, and contribute to our understanding of basic cognition. The method by which people gather information from their environment is important to understand fundamental cognitive processes. In this respect, this series of experiments increments upon earlier work in attention, memory, and cognitive offloading, all of which share some interest in humans’ internal representations of external sources of information.

Specifically, the research investigated herein explores how different instantiations of multidimensional representation can impact decision making while moving toward a scientific framework for data visualization.

1.1. Data-Driven Decision Making

Understanding how an agent comes to a conclusion in the midst of uncertainty is at the heart of behavioural science (Kahneman & Tversky, 1979). Data driven decision making is basing those decisions more on data analysis and less on intuition (Provost & Fawcett, 2013). Such decisions are commonplace in manufacturing. In a survey of manufacturers, respondents were asked to indicate, on a scale from one to five, 1) the availability of data to support decision making and 2) whether their organization used data to make decisions. These two questions, asked of 34,000 respondents, were used to quantify the extent to which organizations used data to make decisions. Controlling for other predictors of success the authors found that, for the average manufacturer, employing data driven decision making adds as much value as investing $5 million (USD) in information technology would (Brynjolfsson & McElheran, 2016). Understanding how people interpret data from graphs, then, is beyond a theoretically interesting pursuit. Knowing this visualization problem space well and maximizing data interpretability can save millions of dollars in business operations at scale.

The phrase “data-driven decision making” typically refers to educational (Park & Datnow, 2009), governmental (Jetzek, Avital, & Bjorn-Andersen, 2014) and commercial organizations’ (Brynjolfsson & McElheran, 2016) management. This is in part due to the capacity of such large organizations and networks to collect meaningful data to assess their performance and the requirements of the people they serve. Provost and Fawcett (2013) describe this stage of data development as Big Data 1.0, where most of the work with data is preparing infrastructure and capacity, and advancements are largely centralized to larger firms with the capacity to develop and store unprecedented masses of information. This Big Data 1.0 capacity is extending beyond large organizations, and more individuals are using personal and public data to support their own decisions. Just as Web 2.0 marked a shift from traditional websites to more user-generated content, Big Data 2.0 is projected to mark a shift from technology companies focusing on data capacity and infrastructure development to more user-focused analytics (Provost and Fawcett, 2013). Although existing early cases of Big Data 2.0 manifest as data driven

product recommendations and advertisements (as of 2017 at the time of writing), other industries are marching toward personalization (Boguski, Arnaout & Hill, 2009), exhibiting not only user-focused analytics but user-oriented data displays to support data driven decisions.

Of course, not all decisions are motivated by graphs and formal statistics. Earlier decision science examines how an observer will make a decision regardless of how the pertinent information is communicated. Centuries of productive work rest on understanding expected value, which is calculated by the value of an outcome multiplied by the probability of observing that outcome. A illustrative example is playing a lottery, where the prize is $1 million and the probability of winning that prize is one-in-a-million. The expected value of each play is $1, and so a rational agent should easily choose not to buy a $2.50 ticket. The calculation of expected value is unaffected by the context or manner by which the information is conveyed: it could be written out in text, described verbally, or perhaps symbolically. The expected value relies solely on the information necessary to perform the calculation. While behavioural economics has improved upon expected value as a predictor of real world behaviour, the role that the manner of information presentation has in decision making remains poorly understood.

The functional definition for data visualization used through this paper is a visual design that provides synthesized data to an observer. The interpretation of the data precedes a decision, and in the case of data-driven decision making that interpretation of the data should have a causal role in the decision. Realistically, biases and prior knowledge will add to the interpretation of the data in the decision making process. Critically, for this work, the visual system will also play a role in the decision (Lurie & Mason, 2007). Structural choices (such as the type of graph to use, and the data to include) and aesthetic selections (such as the colour, size, shape, object ratios) all impact how the data is conveyed to an observer (Cleveland & McGill, 1984). The creator of a graph, then, is influencing the observer's decision before they even have a chance to consider the data, and the connection between perception and decision making is built up through the experiments herein.

1.2. Models of Cognition and Data Visualization

Data visualization practitioners are influenced by advances in cognitive science (Healey & Enns, 2011; Hegarty, 2011; Crapo, Waisel, Wallace & Willemain, 2000) and in many cases, ideas about data visualization are brought forth by cognitive scientists themselves (e.g. Kosslyn, 2006; Green, Ribarsky & Fisher, 2009; Pinker, 1990; Lewandowsky & Spence, 1989). Models of data visualization, however, are less for explaining the mechanisms underlying cognition in response to data visualizations and more for elucidating the relationship between the users and the display. I provide a brief survey of models of data visualizations that make claims about the cognitive system, in part to identify regularities, but also to showcase the level of abstraction currently considered in understanding the relationship between the observer and the data: most models that connect the observer to data in the context of data visualization fail to account for important properties of visual cognition earlier uncovered by basic visual cognition research. The studies to follow focus on how different design decisions can impact decision making based on multivariate data, but by manipulating particular low level visual properties that are not explicitly baked into existing models of data visualization. The connection between perception and cognition is conspicuous in its absence in many of the models below.

1.2.1. Mental models

Models of data visualization are typically high level, process-level accounts of how the user connects to the graphic. For example, Waisal, Wallace & Willemain (2000) propose a theory of data visualization, assuming that mental models are the method with which observers understand the world. Mental models are internal representations of the world thought by some to be the basis of relational reasoning, (Johnson-Laird, 1983; Byrne & Johnson-Laird, 1989). Mental models are said to have three components: comprehension, description and validation. The theory of data visualization posited by Waisal and colleagues aligns with the description phase of a mental model, wherein information is encoded and the cognitive system prepares for further elaboration. Waisal *et al*'s model outlines a set of steps required for an observer to align their own model of the information in the display with the actual display. It takes an observer's comprehension of the data as input, and outputs a response that could be validated.

Taking as input the observers' comprehension of the data visualization, the first stage of the Waisal *et al* model is to build a mental model, after which an observer extracts one view from that mental model and transcribes that view back to the observed graph. To advance, the observer must determine 1) whether the visualization appropriately matches the model view, 2) whether the visualization aligns with the mental model and 3) whether the view and visualization together are viable when considering external facts. As long as all three of the criteria are met, the observer advances to the validation stage of the mental model. If one or more of the criteria are not met, then the mental model is updated and the view is extracted using the updated mental model.

Data collected through a think-aloud protocol and mouse-tracking suggest that observers do use some of these steps in developing an understanding of the data visualization (Waisel et al., 1999). While it was difficult to discern the sequence of steps from the observed data, the authors suggest that proposing discrete steps was helpful for isolating sub-processes and better understanding the description element of building a mental model.

Pinker's theory (1990) shares properties of the above theory in that his is a relatively high level account of the human factors required for graph comprehension. In his account, there is a matching process (that invokes a graph schema of the appropriate type), a message assembly (that converts visual information to conceptual information), an interrogation (where additional information is extracted from the graph if required) and then an inferential process (where decisions are made or insight is generated). These four processes each encompass massively complex issues. Converting visual information to something conceptually meaningful is far from trivial, but is captured in a single process in this model. That isn't necessarily a short-coming, though: so long as this theory is thought of as a method by which the problem of data visualization can be meaningfully broken down into manageable parts, then it is productive to separate an impossibly big problem (how people understand graphs) into slightly more meaningful problems: one of which is to understand how people turn visual information into meaningful information. What Pinker identifies as the matching process is encompassed by "cognitive fit", explored by Vessey and colleagues as its own discrete problem.

The manner with which information is conveyed is important for participants to understand the underlying data: scatterplots are great for visualizing the relationship between two variables, bar graphs are effective for communicating a critical property of multiple levels of a single variable, and histograms are excellent for communicating univariate distributions. Cognitive fit is achieved when the designer uses the right tool (graph) for the job and when the observer understands the relationship between their task and the tool (Vessey & Galletta, 1991). Part of designing a visualization for cognitive fit means using congruent visual representations for the appropriate mental representation of data (e.g. to say a value is higher for one observation than another, indicate that difference visually on the Y-axis of a graph) (Bertin, 1983).

Achieving good cognitive fit reduces the cognitive burden of the observer in completing the task at hand. The basis of the cognitive fit idea is information processing theory (Newell & Simon, 1972) and the observation that observers trade-off error and effort in performing their tasks (Beach & Mitchell, 1978; Payne, 1982; Vessey, 1994; Wickelgren, 1977). Cognitive fit is among a number of models building off of an information processing framework.

1.2.2. Information Processing Models

Models of data visualization that make explicit the generation of the visualization (data to image) and the perception of the data (image to observer) are helpful in thinking about a scientific framework for data visualization. There are two high-level, cognitive actors at play: the designer and the observer. Wijk's model (2005) is an example of how the connection between components can be nicely explained in an information processing manner. The system (comprised of the data and the observer) changes as a function of operations applied between input (data) and output. Knowledge is updated in response to information extracted from the image, where the amount of information depends on the cognitive and perceptual properties of the observer. It is humbling to see an entire domain of study reduced to a couple of element in an equation. All of cognitive psychology, really, is in Wijk's model as one of three steps for information extraction from the environment.

That said, graphic design, and much of computer science are similarly reduced. The simple model glosses over endlessly fascinating nuance, but in a productive

manner. Without zooming out and looking generally at gross connections between data and people to get the big picture, it'd be nearly impossible to make meaningful progress toward its components. Additionally, considering the human and the data as a system together forces theory to effectively capture the relationship between the two component parts.

To move forward and toward explanations of the connections between data, visualization, people, and decisions it's valuable to consider information processors in a hierarchical manner. While Wijk's model is a very high level model that speaks to the overall system, a lower level information processing model can be used to posit cognitive processes necessary for an observer to make sense of a data visualization. Simkin and Hastie do just that (1987). In their model, they suggest that basic processes operate between visual information and the observer's understanding of the task (like the concept of cognitive fit). These basic processes include, scanning, projection, superimposition, and detection. Additionally, anchoring is a process that's thought to be important for making sense of a graph.

Visual anchoring is a process that has been identified in isolation as impacting an observers' perception of a stimulus when comparing (Johnson & Pashler, 1990) or mentally rotating objects (Xu & Franconeri, 2015). That anchoring, a process isolated in theories of data visualization, is identified as a critical process in other types of cognitive tasks (Couclelis, Gollegdge, Gale & Tobler, 1987; Pylyshyn, 1999) suggests that it is at least a pervasive property of human cognition, if not a fundamental one. This is the sort of process that is of particular interest: it makes data visualizations make sense, and knowing more about the process informs cognition more broadly.

While the models to this point are effective at naming potential processes that allow an observer to make sense of the information in their environment, they are missing a critical element: output. A more traditional information processing model of the form input, memory processes and output (Neisser, 1967) is adapted for the operation of making a decision in response to a data visualization (Patterson, Blaha, Georges, Grinstein, Liggett, Kaveney, Sheldon, Havig & Moore, 2014). In addition to actually making output explicit, the authors wisely included a process to allow for attentional capture. This is the first model of data visualization that allows for bottom-up processes to impact observers' understanding of graphs in their own right. In their model, the

stimulus is the input, the first stage is encoding (which may be modulated by attentional capture), and encoding then allows for long term memory and working memory to modify the information. Long term memory can be modified by working memory and pattern recognition processes, and working memory can modify the encoding step. Working memory and pattern recognition together inform the decision, upon which the response is based. While it does not make explicit the multiple components of interest in the generation of the stimulus (where the stimulus is the data visualization), this model does identify the general clusters of processes that must be understood to develop a science of data visualization.

1.2.3. Empirical Studies of Visual Cognition and Implications for Data Visualization

Reading a data visualization is necessarily a visual experience. In the models discussed above, visual input is in the form of a graph or plot which is subsequently encoded into memory for elaboration or consolidation into long term memory. Earlier work in basic cognition has isolated manipulations of the visual environment to determine their impact in simpler tasks. Indeed, a scientific account of visual cognition does require vision to be studied in near-isolation, and recent work in the perception of data visualization is building upon the strong foundation established by basic vision research.

This research has shown that while many aspects of the environment require attentive processing, there are some visual features that can be processed pre-attentively. Whereas “attentive” processing is the slower, more deliberate processing of a subset of the visual environment, “pre-attentive” processing is the faster, more reflexive processing of the visual environment more globally. Therefore, designing visualizations that support pre-attentive processing should free up the observer’s cognitive resources to be more effectively used in interpreting the meaning of the data (Healey, Booth & Enns, 1995), and result in a more effective method for communicating data.

Based upon a survey of basic research in attention, Healey and Enns (2011) identified clusters of tasks that are supported by pre-attentive processing: target detection (finding a goal-relevant unique item in an array), boundary detection

(identifying distinct spatially organized, texture-based groups based upon how they differ from each other), region tracking (or multiple object tracking) and counting (or subitizing). These findings clearly provide an evidence based manner for making design decisions when generating data visualizations. If the goal of the data designer is to show the difference between groups on two different variables, for example, the designer may use a set of features, such as colour hues, that allows an observer to pre-attentively discern boundaries between the two groups on a scatter plot (Callaghan, 1990).

In the above particular example of two groups on a single scatter plot, an effective design decision to communicate the values of the two groups is to invoke the percept of two distinct textures. That is, one set of co-ordinate values belongs to one group or one perceived texture, while a second set of co-ordinate values belongs to another. Very large datasets and data of a spatial nature lend themselves well to texture-based visualizations (Enns & Rensink, 1990; Julesz, 1984; Rensink, 2006; Ware, 2004) wherein perceived surface values are used to communicate properties of the data. Heat maps, multivariate scatter plots, and cloropleth maps are all good examples of visualization strategies that rely on texture information to convey values of data. For textures to effectively communicate differences between groups, it is helpful if the boundaries between textures is easily observed.

Boundaries that define categories and texture segmentation have been phenomena of interest for cognitive science generally, extending beyond data visualization applications (Neisser, 1967; Triesman & Gelade, 1980). Boundaries are essentially what is of interest in category learning research, where the capacity of an observer to learn the difference between in-group and out-group items is based on their capacity to develop a group boundary — a point after which the observed features are deemed to be a member of the group summarized by a category label. Boundaries are foundational for concept development. Knowing the start and end of a phoneme is critical for humans to communicate verbally, and knowing the difference between a lion and a house cat has clear implications for survival. Texture segmentation is a special type of visual boundary utilization, where two surfaces can be preattentively (Triesman & Gelade, 1980; Julesz & Bergen, 1983) separated. Work by Callaghan (1989; 1990) shows that some features are more supportive of rapid boundary detection.

If an observer is shown three columns of circles and three columns of squares all of the same hue, the observer will perceive a boundary between the groups of shapes. In her 1989 work, Callaghan found that boundaries formed by different shapes (the percepts of a boundary between a set of circles and a set of squares) were disrupted if hue discontinuities were introduced. If the groups of circles and squares are heterogenous in colour, the group boundary is lost. However, boundaries formed by hue were robust to disruptions of shapes; a group comprised of shapes is perceived as a group so long as the group members are of good form and similar hue. Therefore, it appears that hue better forms texture boundaries than form (consistent with Pashler, 1988). The implication for data visualization, then, is that colour is a more effective channel by which to communicate group differences for data plotted in two [pseudo-]continuous dimensions. Colour can form boundaries that are pre-attentively processed. Boundaries, at least colour-formed boundaries, elicit one type of pre-attentive processing.

Other visual features that support pre-attentive processing include basic elements that might be observed in a basic saliency map model (Itti, Koch, & Nieber, 1998) like orientation, hue and length and more complex Gestalt-like features that emerge from basic components such as density, boundary crossings/intersections and closure (Healey & Enns, 2011; Triesman & Gormican, 1988). Using such pre-attentive features to develop easier data visualizations is a practiced skill for accomplished data designers. Practically, most designers know how to effectively leverage pre-attentive features; theoretically, data visualization provides an opportunity for cognitive scientists to study stimuli that elicit pre-attentive processing to observe its operation in applied settings. Building upon the earlier example, a scatter plot that uses hue to distinguish two groups would more effectively support pre-attentive processing than scatter plots that rely on shape to distinguish two groups. This design practice could easily be data-driven, in that it builds right off of Callaghan's (1989) work investigating visual cognition in the perception of stimulus items in an array. However, grouping by colour is a practice already employed by most data designers who have not been exposed to the empirical work. In this respect, data visualization is rich with opportunities to validate ideas about visual cognition. Designers are using effective practices that they have not formalized, and scientists are looking for opportunities to test their theories of visual cognition outside of the laboratory.

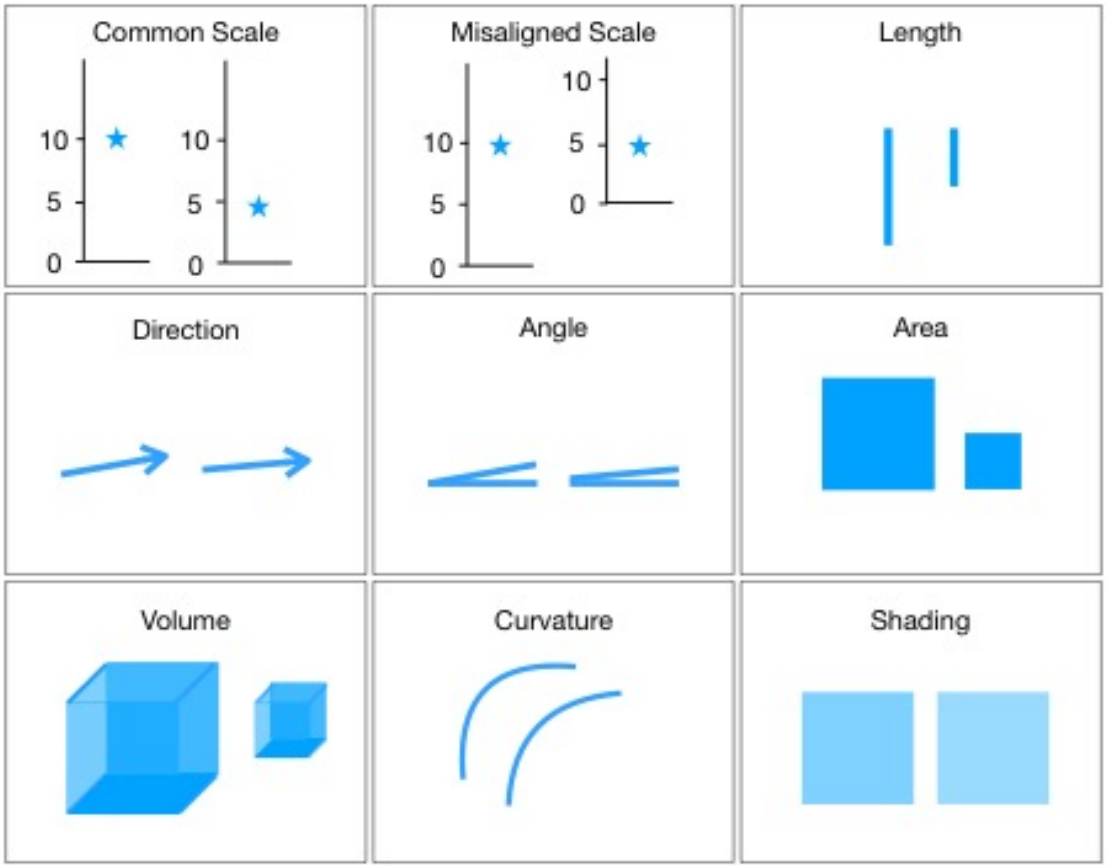


Figure 1.1. Adapted representation of Cleveland and McGill's elementary perceptual tasks.

Quantitative information can be extracted from data visualizations. The figure shows, in descending order, typical performance accuracy. Judgements of position and length correspond to better accuracy than judgements of volume and shading.

Cleveland and McGill (1984) tested different visualization strategies in developing their theory of graphical perception. They did not focus on pre-attentive processing, per se, but they did encourage use of graphical elements that maximized performance on elementary perceptual tasks. Elementary perceptual tasks are the simple information extraction tasks for which there is a one-to-one mapping between graphical properties and some underlying data (for example, a line graph's increasing height represents its increasing value). The elementary perceptual task is what Julesz (1984) would call the basic unit in their theory of data visualization. Building toward a science of data visualization benefits from having basic units of analysis to build upon, and the elementary perceptual task is befitting of the role as the atomic, most fundamental phenomenon upon which data visualization is made possible. To operationalize the efficacy of different elementary perceptual tasks in data visualization,

Cleveland and McGill presented different types of graphs to observers, and ranked the elementary perceptual tasks on how closely observers responded with the actual data for each type of visualization. It's important to note that the elementary perceptual task is to identify a single value, and so precision is rewarded. They are not particularly concerned with relativistic interpretations of data (as may be the case with heatmaps) or interpretability of findings beyond extracting that particular value (as is valued for geopolitical cloropleths). The most accurate information extraction performance was for graphs that varied position (height) on a common scale (y-axis). Performance was less accurate for scales that were misaligned, for data represented using changing angles (pie charts), and for curvature. Interestingly, the worst overall performance was for variables communicated through colour.

That colour poorly communicates the value of data is surprising at first glance: empirical work in basic visual cognition suggests that colour is processed before shape (Theewues, 1992) and Callaghan's finding about colour as an effective grouping variable would perhaps suggest that colour should be the *best* method to communicate data. At this point, the role of task, prior knowledge, expectation and other top down effects should be considered: colour is an effective grouping variable but it is a poor method to communicate continuous variables (Cleveland & McGill, 1984). While grades of change on a heat map can convey the relative values of a continuous variable in space, such visualizations do not support precise estimates of the plotted value. Going back again to the scatter plot example, the colour is effective for discerning the two groups, but the data that is presented on the scatter plot within groups is presented by position on the X and Y axis. Lessons from visual cognition are clearly important, but must be considered with nuance when applied to more complex task environments.

One method to best use lessons from visual cognition is to consider which basic research translates into applied visual cognition (Fisher, Green, & Arias-Hernández, 2010). Patterson *et al*. (2014) identify opportunities to apply basic research to human computer interaction design as "leverage points" because they are meant to take advantage of the existing knowledge of the cognitive system to improve data visualizations and design. A leverage point gets its name from the opportunity gain "leverage" in applied environments from basic research. It is a point, in that there are certain steps in human computer interaction design that are better suited to benefit from lessons in human cognition than others. There are six leverage points, increasing in their

level of abstraction. The first, lowest-level leverage point is to 1) capture exogenous attention (bottom-up processing) by using salient visual cues for important parts of the visualization. Designers can take advantage of the basic science to, for example, better capture observers' attention to help observers find important parts of a data display. The second is to 2) guide endogenous attention (top-down processing) to help the observer choose meaningful information in the visualization. The remaining leverage points build upon visual cognition research investigating 3) chunking, 4) mental models, 5) analogical reasoning and 6) procedural learning. In this respect, these six leverage points are the closest we have to an existing framework for integrating multiple levels of analysis for data visualizations. While multiple levels are effectively identified, there is still more development required before predictions about how those levels are integrated. Thus, Patterson *et al.*'s leverage points are not a theory, so much as they are ideas to support a translational cognitive science. They do, however, offer a great way to anticipate which findings in basic visual cognition would be observed in data visualizations. In lieu of top-down pressures and expectations, salient cues in data visualization will attract attention. Of course, in actual practice, no one of top-down or bottom-up types of process exist in isolation, which is explored in the next section.

As such, there is good precedent for connecting lessons from basic vision research and data visualization. Pre-attentive, bottom-up type processes have been better explored for texture segmentation and boundary detection in the context of data visualization than other visual processes. While the present studies do not rely on boundary detection and textures, there are important considerations drawn from this survey. For one, colour is an effective way to communicate group differences, but less effective for communicating precise values. Additionally, considering the motion toward a scientific framework for data visualization, it's important to note the value added to data visualization design from basic visual cognition research in segmentation and boundary detection.

1.3. Attention, Cue Use and Decision Making

Prior knowledge can mediate the influence of a cue on attention. Athletes playing team sports need to track their movements, their allies' movements and their opponents' movements to know what to do next: tracking behaviour in sports exemplifies how the task or prior knowledge can mediate the influence of a cue. In a study of how people

integrate this kind of information, observers were shown video frames of a basketball game up to the point where the player with the ball may choose to shoot. The videos are coded for five cues that indicate the viability of a shot. For example, if the player in possession of the ball is closely guarded by an opponent, it would be unwise for them to take the shot. One half of participants in the study were asked just to watch the videos and say whether they thought taking a shot would make sense. The second half were coached to pay attention to four of the five cues. The fifth cue, distance to the basket, was not mentioned for either group. The researchers found that instruction hampered observers' ability to integrate meaningful information, in that the first group relied on the distance cue more than the coached group (Cadenas, Cárdenas & Delgado, 2015). Since the observers differed only in whether they were told to focus on a subset of the available information, it appears that the coaching was so effective in getting people to rely on particular cues that they ignored additional, helpful information.

Further evidence for the importance of top-down information in driving attention is shown in differences between novice and expert observers. There is an effect of expertise and domain knowledge on how people scan their environments (Koide, *et al*, 2015; Reingold, Charness, Pomplun, & Stampe, 2001). It appears that exogenous salience is less predictive of eye movements if an observer is familiar with the content of their environment (Schütz, Trommershäuser, & Gegenfurtner, 2012). In a recognition task, when participants were shown images relevant to their domain (engineering or American studies), domain experts were less likely to look at salient regions of photos (Humphrey & Underwood, 2009). They were also better at recognizing images that were domain-relevant in photos than recognizing photos that were not domain-relevant. Experts scanned pictures more broadly, exhibited longer eye movements, and made fewer fixations than novices.

Salient cues also impact information processing (Hammer, Sloutsky & Grill-Spector, 2015): the physical properties of an object matter. In earlier work exploring attention to category features (or cues) with varying salience and varying task relevance, we found that eye movements to irrelevant cues were faster than eye movements to relevant cues when each were of similar salience. However, when the irrelevant cues were brighter, and were more visually conspicuous than the relevant cues, saccades directed to the salient irrelevant cues were the same velocity as saccades directed to less salient relevant cues (McColeman & Blair, 2015). Finding that saccades were faster

to irrelevant cues was consistent with earlier theories of visual attention (Munoz, Broughton, Goldring, Armstrong, 1998; van Zoest, Donk & Theeuwes, 2004; Xu-Wilson, Zee, Shadmehr, 2009). However, it was not anticipated by existing theories that visually salient distractors would draw slower eye movements than when the distractor was less salient. It seems that there is an interplay between top-down and bottom-up attention in controlling the eyes. Top-down, endogenous value judgements (to look at the relevant features) appear to drive attention toward relevant cues instead of distractors. If all cues (relevant and irrelevant) are of similar salience, fixations to irrelevant dimensions are rare, and fast — eye movements appear to escape the distractor suppression (Caputo & Guerra, 1998; Gaspar & McDonald, 2014) and quickly query the irrelevant dimension. When the irrelevant dimensions are salient, however, eye movements are slower and more common (Walker, Walker, Hussein & Kannard, 2000), perhaps owing in part to the observer choosing to look at the irrelevant cue rather than ignoring it.

Salience appears to influence performance on simpler tasks a little differently. In visual search, the participant is asked simply to find an object in a search array. A visually conspicuous icon in an array of otherwise homogenous icons stands out as salient and appears to capture attention (at least early on, van Zoest & Donk, 2007; Theeuwes, 2004). Exogenous salience enacts greater influence on early visual processing while top-down information more strongly influences later visual processing (van Zoest, Donk & Theeuwes, 2004). Salience draws attention early, and so rapid responses to the environment are more likely to be salience driven. In the present studies, participants' responses are self-timed, and so salience is expected not to entirely interrupt more top-down processing of the input. However, the influence of salience on early attention may proliferate to higher level processes.

The extent to which properties of the environment influence decisions remains poorly understood. Clearly, the environment does impact behaviour, but the connection between salience and decisions still requires further investigation. One major line of query is the relative role of endogenous versus exogenous predictors of how information is processed. An object with high endogenous value might be a puzzle piece that completes the image, while an object of high exogenous value might be something brightly coloured that appears through rapid onset. In the context of decision making, visual salience appears to mediate the influence of endogenously defined value in evaluating stimuli (Louie, 2013; Schütz, Trommershäuser & Gegenfurtner, 2012). Recent

attempts to quantify the importance of exogenous and endogenous factors in attention during a goal-directed task suggest that endogenous/top-down factors outweigh exogenous factors about two-to-one (Towal, Mormann & Koch, 2013). While the balance of power favours top-down attention in driving performance, it is not so one-sided that exogenous attention can be ignored. Understanding the relative contributions of top-down and bottom-up factors and the interplay between them is critically important to developing a robust model of how people interpret data visualizations. Given that the purpose of the experiments below is to explore how different visual manifestations of a data visualization can impact decision making to build the foundation for a framework for data visualization, it's important that there's some consideration for both exogenous and endogenous factors.

A visual cue is any sort of visual input that might inform a decision or response. Regardless of whether a visual cue is an an error bar that reflects uncertainty in a graph, the jersey colour to represent the team of a basketball player, or a category dimension that is predictive of category membership of an experimental stimulus, the physical manifestation of that cue does have some impact on the observer's decision. An observer cannot make a decision until they have information; information cannot be elaborated upon unless it has been attended to; attention is impacted by physical properties of the environment. An effective model of decision making based on data visualizations should then extend beyond traditional decision-making parameters (e.g. subjective utility, prior knowledge) to include parameters to capture attention and properties of the environment (Healey & Enns, 2011). This is a good general goal for improving theories of decision making, but is particularly valuable for predicting effective data visualization, given that data visualizations are designed to be an interface between data and people. Knowing how the visual properties of a data visualization impact a human's interpretation of data is valuable for both designers and for researchers interested in cognition.

1.4. Data Visualization and Categorization

Data visualization communicates abstract properties of data to an observer. Rensink (2014) defines it as the "transforming of a problem into graphical form, so as to engage the *visual intelligence* of a human viewer" (pp. 151; my italics) which implicitly acknowledges the need to abstract information into its human-friendly form. To

understand the observer's experience, it is valuable to consider the process by which the graph comes to be in the first place. To generate a data visualization, there's an encoding process to move from data to graph (Cleveland & McGill, 1984). This is like categorization: to learn a category, one observes multiple instances of a single kind, and identifies their common group membership, encoding abstract properties of those single instances into categories (abstraction, Figure 1). After learning the category, a single label can then be applied to these multiple observations extracting underlying properties that members of the category share. After the category is established, determining what constitutes that category is a decoding or extraction problem.

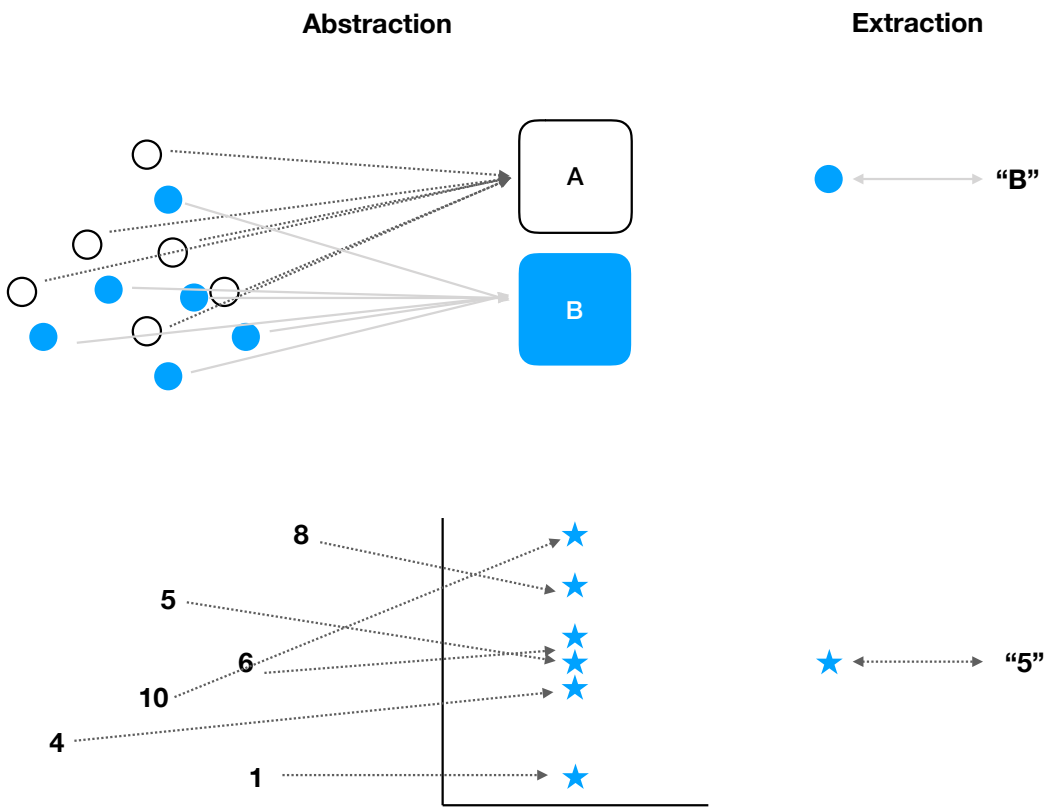


Figure 1.2 A simplification of the connection between categorization (top) and data visualization.

To develop the abstraction for types of problem tasks is to connect specific items to general concepts. To then use the category/visualization as an observer, an observer must extract meaning and apply it to specific exemplars. In the top panel, category extraction is shown, by using a quality of the stimulus to assign a label to it. In the bottom panel, data visualization extraction is shown where some meaning is gathered from the spatial information conveyed by the visualization.

This requires identifying features from each instance and establishing their congruence across multiple observations as well as some indication of the category membership of each observation (Smith, Shoben and Rips, 1974). In the task of learning to identify birds, we must have received sufficient feedback or information to learn that the presence of feathers is a good predictor of “birdship”, size is not bad (though chickadees and ostriches may disagree) and that the direction of motion is nearly useless (save for migrating flocks). The sheer number of features that may once have been in contention but were ultimately discarded is impressive in itself, but the ability to abstract away from physical representations of these properties (a brown feather is as much a feather as a blue one) is where the cognitive complexity arises. The generalizability and abstraction of concepts is a critical skill, underlying cognition as we know it. Reading a data visualization invites the same type of process: over repeated exposure and instruction, the graph reader learns that the vertical position of a line graph is more informative about the value the graph conveys than the colour of the line.

In psychometrics, we also select data for a measure that we're interested in (Shultz & Whitney, 2005) such as reaction time and response accuracy. Reaction time is meant to reflect processing time, where a higher reaction time indicates a greater processing load for the participant. Accuracy is determined by comparing the participants’ responses to an objectively correct response, and is important for understanding whether participants were able to meet the goals of the task (for example, finding a target in an array of distractors). After deciding it’d be meaningful if the speed and accuracy change together in some sort of consistent way, the scientist may decide to represent each data point with a black dot, where the position of the dot horizontally is scaled to correspond to reaction time and the extent of the dot vertically represents accuracy in a scatterplot. Should the scientist have collaborators that they’d like to communicate with, this plot can be helpful, but the processes necessary to allow people to communicate with graphs is complex. Choosing to investigate the accuracy and reaction time in a two-dimensional scatter plot requires the audience to follow the scientist’s abstraction process from real world event, to recording, to digitization and the complex conceptual jump from data to graph. Even so, with prior knowledge, labels and context, a single black dot on a computer screen can be interpreted by multiple observers as (for example) describing a reaction time of 448 ms and an accuracy of 84%. This is a phenomenal information density for such a simple image.

Like the birdwatcher, the scatter plot observer expects a few features of the image to be meaningful, and some to be less meaningful. The birdwatcher looks for feathers. The plot observer looks for axes. Where the birdwatcher looks for size, the plot observer looks for icons. In both tasks, any single feature alone is insufficient to make sense of the stimulus, and there are a great number of features that don't matter to complete the task. In a data visualization for example, the font type is unimportant; in birdwatching, the presence of a neighbourhood cat is unimportant to determining the category of the bird. Categorization and data visualization, then, both require understanding of relevant features of the stimulus to make sense of the environment (Rehder & Hoffman, 2005; Blair, Watson & Meier, 2009). In category learning, we know that people will look at irrelevant dimensions until they have sufficient practice to learn to ignore them (Blair, Watson, Walshe & Maj, 2009; McColeman, Barnes, Chen, Meier, Walshe, & Blair, 2014). In data visualization, practitioners generally aim to represent one variable with one feature and avoid uninformative changes in the display (Bertin, 1983): "The Principle of Relevance" is one of Kosslyn's psychological principles of effective graphics. It suggests that the reader expects to see only and all relevant information in the graph. The impact of including irrelevant information on a graph observer (a violation of Kosslyn's principle) is explored within the present work (Experiments Two and Four).

1.5. Candlestick plots

Four centuries ago, Honna established a plotting strategy to visualize multiple inter-related, auto-correlated variables (Nison, 1994). The variables of interest were the opening, closing, high and low prices of stocks — specifically rice futures — and the manner by which he represented these variables was with candlestick charts. Recall, the candlestick chart captures four unique dimensions (high, low, open and close prices), and a fifth dimension that summarizes the relationship between two of them (whether the stock closed higher than it opened, see Figure 1.3). It also allows for a sixth dimension (time) by presenting the candlesticks as a series (Figure 1.4).

The candlestick includes a body and its shadows, where the upper and lower edge of the body reflect the opening and closing prices (not necessarily in that order) and the upper and lower extents of the shadows represent the high and the low price of the time step (necessarily in that order). Interpreting the high and the low price of the time step is pretty straightforward: look at how far the shadow goes and its vertical

extent marks the value indicated on the y-axis. That's an easy interpretation: lower shadows indicate lower prices; higher shadows indicate higher prices. The oddity of the candlestick plot is in interpreting the open and the close price of the stock. The opening price is either the top part of the body or the bottom of the body, and to know which, the observer must use the colour of the body. Standard candle plots use green to indicate that the stock closed higher than it opened for that time step, and red to indicate that it opened higher than it closed. Variations of this scheme include using white for closing high and black for closing low, or empty boxes for closing high and filled boxes for closing low in different colours. In all, to read a candlestick plot, the observer must first look at the colour of the body and then use that information to determine whether the top part of the body represents the opening or the closing price, with the bottom representing whichever price is left over.

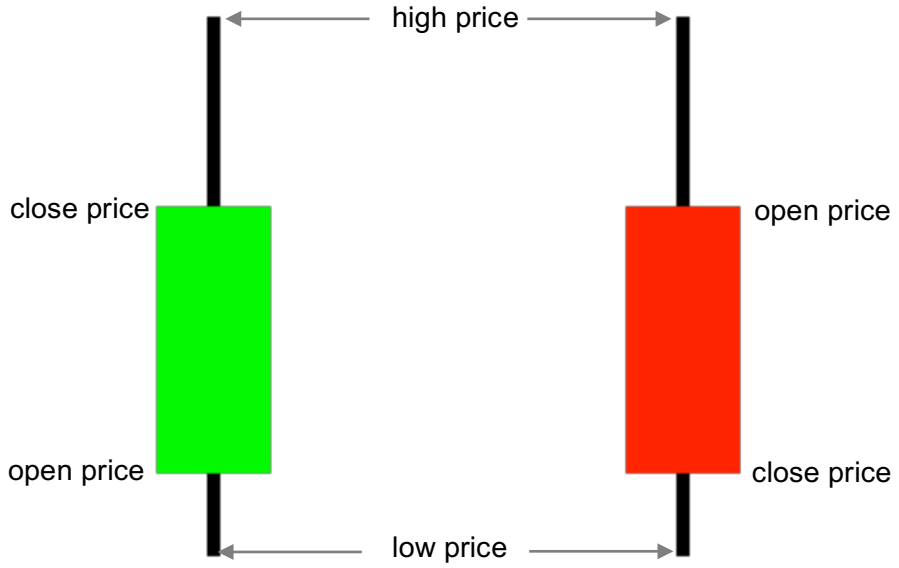


Figure 1.3. A sample of candlestick chart components.

The larger boxes are the candle bodies. Notice that the opening and closing are represented by different locations. The colour indicates whether the top or the bottom of the body reflect the opening or the closing price.

Their standard use does have unnecessary barriers to entry beyond the complexity of reading opening and closing prices. Roughly 4% of males (and some females) have some difficulty distinguishing colours in the red-green spectrum due to retinal cone anomalies (Modarres, Mirsamadi & Peyman, 1996), which makes the traditional candlestick plot tricky to use. Beyond accessibility concerns there is reason to believe that these plots mislead the observer about the relative importance of the

dimensions the candlesticks represent. Despite their issues, they are still helpful graphs to many people in finance, and so, evidently, proficient users have found their way around the shortcomings of the candlestick plots.

Figure 1.4. A sample of a full candlestick plot.

The Y axis represents the price of gold, while the X axis represents the year (1968-2008). Reprinted from "Gold as Investment", In *Wikipedia*, n.d., Retrieved December 2, 2016, from https://commons.wikimedia.org/wiki/File:Gold_Price_(1968-2008).gif. Public domain.

The proficiency shown by candlestick users is, like most other skills, learned. Through training and education, practitioners accumulate knowledge about how to use them. Previous work has shown that prior knowledge and top-down factors are often better predictors of where people will look and what decisions they will make than visual salience and bottom-up factors (Baluch & Itti, 2011; McColeman & Blair, 2015; Tatler, Hayhoe, Land & Ballard, 2011). The advantage of prior knowledge is even clearer in experts, who appear to see patterns more wholly than novices.

Experts' capacity for pattern perception is so good, and can be so automatic that it impedes their ability to see the patterns' component parts. In the case of athletes viewing a scene unfolding, experts struggle to identify mismatched frames if the contents of a frame display a meaningful story (Gorman, Abernethy, & Farrow, 2011) because they are focused more on the event than the irrelevant contents of the mismatched frames. When they're shown the mismatched frames, they are so aware of the more cohesive whole movie that they cannot isolate mismatches. There's reason to believe, then, that sufficient experience with plots that would mislead novices would not negatively impact experienced observers. However, even if expert traders and finance professionals are knowledgeable enough to use their prior knowledge to attend to all four (and not just the more salient) dimensions, it's an inefficient use of plotting space and observer effort to require an observer to perform additional unnecessary operations to extract information. While it's good that the icons are salient, it may be maladaptive to have two of the dimensions (open and close price) represented as more salient than the remaining two dimensions (high and low price), given that there's no evidence to prefer one dimension over the others in predicting stock performance.

The candlestick plot is not widely used outside of finance, and so most undergraduate psychology students have not learned how to read them. If a study is conducted about line graphs, most of the prior knowledge about the task has already been developed through a combination of previous exposure, education and experience with using lines to represent change/motion/similar constructs. This is great for data literacy, but it's challenging to study how people learn how to use graphs if they already know how to. In contrast, if a study is conducted on the perception of candlestick plots, the participants have to learn what the icons mean generally (that is, the body reflects the open/close price; the upper shadow represents the high price, the lower shadow the low price), how they relate to the axes, and that the location of the opening and closing prices are indicated by the colour of the icon. The relative difficulty of each of these subtasks can be inferred by the order in which they are learned. It's rare in the study of graph perception that there is a standard plot that requires learning to use, which makes interpretability of familiar graphs such as bar and line plots more difficult to study, and the study of candlestick plots particularly appealing for studying people as they learn to read a data visualization. Additionally, due in part to its clear connection to economics and the opportunity to build upon earlier findings in behavioural economics, the

candlestick plot is an interesting object of study: it invites connections between perception, abstraction, information integration and decision making.

Given the idiosyncrasies of the plot outlined above —- the need to learn how to read the candlesticks, the potential for improvement, and the connection to established attentional learning findings —- the candlestick plot is a great starting place for the overarching goal of this work: building toward a science of data visualization to include decision making.

1.6. Summary of Experiments

Seven experiments are conducted to build up an overview of user performance error from basic information extraction through to complex risky decision making when using a candlestick plot. Inspired in part by Cleveland and McGill's theory, the experiments are organized such that the first experiments are about elementary perceptual tasks (extracting information; Cleveland and McGill, 1984). The tasks in later experiments move from information extraction to pattern perception to using the information in the graphs to make a risky decision. The goal of all seven studies generally is to test which manipulations of visual properties affect participants' performance while using a candlestick plot to move toward bridging perception and decision making in a scientific framework of data visualization.

Leverage points have been identified before as an opportunity to use knowledge from basic cognition to improve data visualization (Patterson et al., 2014). Instead of using known properties of basic cognition to make a better graph, however, this series of experiments studies properties of visualization to better understand basic cognition. For example, in all experiments, a consistent question is whether the colour or the shape of the glyph supports better understanding of the graph. Colour is known to be processed quickly, but in the right task environment, shape can be more informative and drive reflexive attention. While the measures discussed herein pertain to accuracy and cannot probe the order of processing to query reflexive attention directly, the role of reflexive attention might improve the encoding of dimensions that the observer is reflexively oriented to (namely, the close price in the triangle candle body conditions). The test of colour versus shape is a test that helps theories about applied cognition generally. The

results suggest that while colour might be processed faster, shape supports better overall pattern perception.

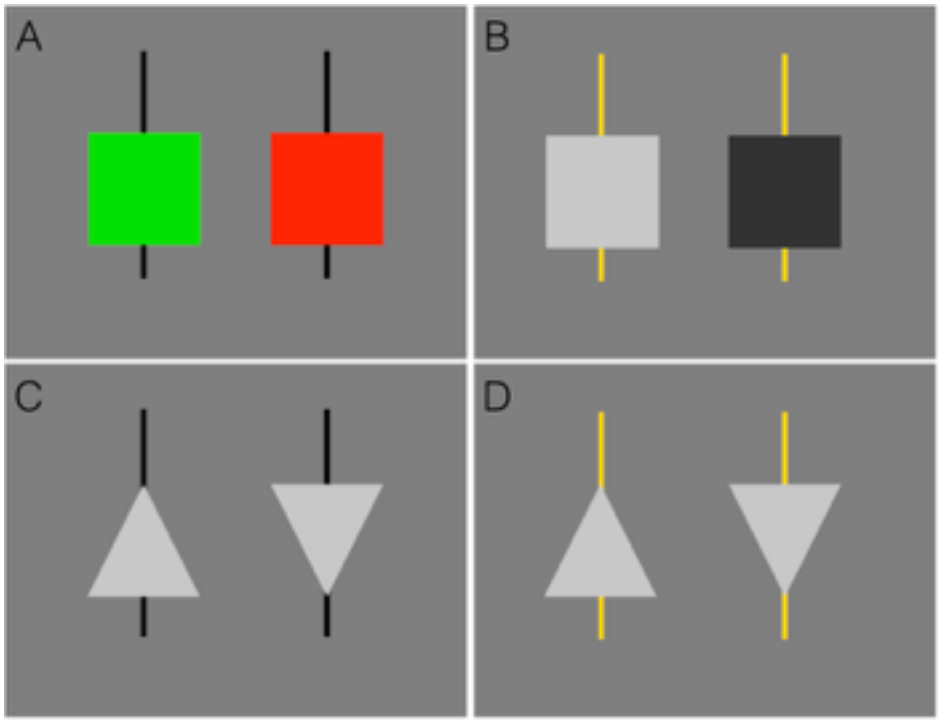


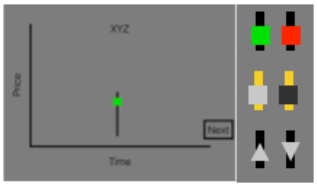
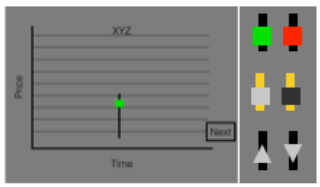
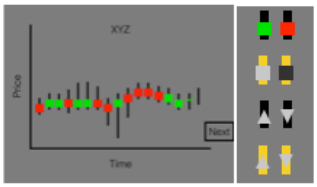
Figure 1.5. An example of the possible stimuli used in the experiments. Each tile (A-D) represents one experiment condition. In this example, the left icon in each pair represents a higher closing price than opening; the right represents a lower closing price than opening. The large portion of the icon is the "candle body" (red/green in Condition A; up/down in Conditions C and D) and the sticks outside of the candle body are the shadows (black in Conditions A and C; yellow in Conditions B and D). Note that the candle body in Conditions (C and D) "point to" the closing price.

The seven experiments in this study are summarized in the following tables, where each table corresponds to the experiments contained in a chapter. The subsequent chapters themselves provide more detail and rationale, while this section is meant to serve as a quick reference with experiment summary tables and a brief overview of the results for each chapter. Variables associated with an effect are marked with asterisks. The general goal of the seven experiments is to test the impact of manipulating visual representation on participants' accuracy while they use a candlestick plot.

The second chapter in this dissertation (Table 1.1) moves from simple information extraction (in Experiments One and Two) to more complex forecasting (in Experiment Three). Experiments One and Two examine the simplest case of reading a

candlestick plot: extracting specific information from a single glyph. On each trial, Experiment One's stimuli are a single icon and axis labels. The performance error was collected to test whether salient dimensions (candle bodies in Condition A) attract additional processing to reduce user performance error, or whether more symbolic representations (triangle candle bodies) were better at supporting user performance. There was no main effect of condition in Experiment One. Experiment Two introduces the inclusion of grid lines, a design practice for which there was little empirical evidence. It was expected that participants would better be able to extract information from candlestick glyphs when grid lines guide perception. However, participants actually performed worse when grid lines were introduced. Possible explanations for this are identified in Chapter Two. In Experiment One and Two, participants were asked to report the particular value of a single dimension. Practically, this serves to establish what user error associated with just a single icon looks like. Theoretically, this is a test of stimulus responsive attention — the flexible access of information in response to properties of the available external information — and whether it's observable at the level of user

Table 1.1. Perception of candlestick plots and the efficacy of arrow candle innovation (Experiments 1-3) Overview

	Experiment Title	Manipulations				Sample Image	Explanation
		Candle body	Noise	Spread	Grid lines		
1	Extracting plotting information from candlestick icons	3	No	No	No		A single glyph is presented to participants and they answer questions of the form, "what is the opening price?" by typing in a response. Candle bodies are manipulated between subjects to create the possible conditions (shown to the right of the image)
2	The impact of grid lines on accurately extracting information from a candlestick plot	3*	No	No	Yes*		Same as Experiment 1, but with the addition of grid lines.
3	The influence of salience on forecasting	4*	No	No	No		A more realistic candlestick plot, shows multiple glyphs simultaneously to indicate a time series. There are four possible conditions based on the four possible candle bodies; the candle bodies are a between-subjects manipulation.

performance error. Stimulus responsive attention is further discussed in the next chapter.

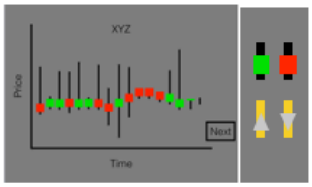
Experiment Three is the foundational work for this series. It is a study of different types of candlestick iconography under more standard use in visualizing financial time series. There was an effect of candle body in Experiment Three, where triangle candle bodies supported better forecasting performance. Experiment Three is the first forecasting task, where the stimulus is many icons rather than one. Given evidence from Experiments One, Two and Three, it appears that using a stimulus responsive information access rule is possible under all three tested conditions when there's only a single icon, but, given the poorer performance in Condition A, it appears the rule is more challenging to use when the stimulus set is richer — the more symbolic candle bodies

render the stimulus-responsive rule learning trivial, but the rectangle candle bodies are more challenging.

All experiments up to, and including, Experiment Three tested the impact of relative salience for dimension representations in candlestick plots. To avoid under-powered research beyond that, though, Experiments Four, Five and Six used only the candle body types shown in Figure 1.5A and 1.5D. Condition A and D were selected to be maximally different from each other, and to set the foundation for additional experiments exploring their component parts (beyond the scope of this dissertation). Of course, it'd be preferred to continue with all four conditions throughout the entire seven experiments, but resource constraints rendered that implausible. Choosing the two maximally different conditions, however, offer an opportunity to test for differences in circumstances where differences are most likely observable. If such differences are observed, then diving into the nuances that support performance differences in a separate set of studies is a clear next step. For example, the relative salience of the dimensions may be manipulated to test separability constraints in time series pattern perception.

Experiment Four (Table 1.2) tested the impact of an irrelevant dimension on the perception of a time series. While in the other experiments, there was a relationship between the four dimensions, in Experiment Four one of those dimensions was wholly uninformative. Conceptually, the question is whether time series pattern perception is integrated when reading candlestick plots, or whether glyph dimensions are encoded into working memory separately. In Experiment Four, the test was whether rendering one dimension irrelevant negatively impacted performance beyond that dimension.

Table 1.2. The impact of irrelevant dimensions on multivariate forecasting

Experiment Title	Manipulations				Sample Image	Explanation
	Candle body	Noise	Spread	Grid lines		
4 Irrelevant dimensions in forecasting	2	4* (open, close, high or low)	No	No		Each participant has one of the four dimensions (open,close,high or low) replaced with noise. This example shows noise in the "high" dimension. There are eight between subjects conditions: four noise conditions x two candle body conditions (shown to the right of the graph).

This manipulation was to test the capacity of an observer to develop an understanding of the informative dimensions with an irrelevant distracting dimension, building on earlier work in category learning in which participants learn to ignore irrelevant distractors to effectively make their decisions. Learning relevant from irrelevant features in making category judgements is a critical skill for gaining knowledge about the world. In data visualization, a designer often works to ensure that they include only relevant information (consistent with Principle of Relevance, discussed below). Experiments Two and Four offer empirical validation of that practice. More critically, however, is a study of how including irrelevant information impacts perceivers' understanding of the plotted data. If participants can ignore irrelevant dimensions in category learning, it was expected that they should be able to ignore irrelevant dimensions in other types of problems too. When one candle dimension was replaced with random noise, there was no noticeable difference between the candle conditions in forecasting suggesting that the advantage of symbolic, triangular candle bodies is lost when an irrelevant dimension is introduced. However, the performance did differ depending on which dimension was replaced with random noise. When the close dimension was irrelevant, participants performed better.

The close dimension is represented by the candle body. It was expected that the candle body, when appearing more erratic and oftentimes larger than normal, would have negatively impacted performance and so further investigation was required to better understand why the noisy close dimension was better for participants. One possibility is that the close dimension simply subtended more degrees of visual angle,

and the larger size of candle bodies allowed for more precise recall. To test whether the performance differences are due to the size differences of the random icons, Experiments Five and Six test the impact of size through manipulations of the icon spread (Table 1.3). In Chapter Four, theories of spatial representation are also explicitly explored, and are tested in the context of each experiments' manipulations.

Experiments Five and Six tested perceptual distance and salience on participants' ability to extract data from plots and forecast financial patterns, building upon theories of spatial representation. Size generally has minimal impact on the ability of participants to extract information. Experiment Five is more similar to Experiment One, in that a single icon is on the screen at a time and the participants' task is to extract a single value from that multivariate icon. Regardless of the sizes tested, the different Conditions all supported similarly good information extraction. In Experiment Six, it was found that more extreme sizes correspond to better performance while forecasting, suggesting that either an overview "zooming out" on the data and using symbolic icons is easier for pattern perception in time series data plotted on candlestick charts, or "zooming in" on the data to better extract details from the symbolic icons is better for user performance. A possible explanation is that there is a trade off between "zooming in" (to support detail extraction) and "zooming out" (to support general pattern perception) that is balanced overall in user performance error observed in this Experiment Six.

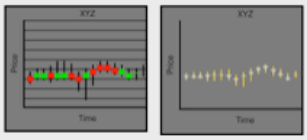
Table 1.3. The distance between glyphs and its role in extracting information from candlestick plots

	Experiment Title	Manipulations				Sample Image	Explanation
		Candle body	Noise	Spread	Grid lines		
5	The influence relative glyph spacing on forecasting multivariate trends	2	No	3 (close, standard, far)	No		The three horizontal images show the three different spread conditions (close, standard, far) where "close" displays smaller distance between the four dimensions. There are six between-subjects conditions:
6	The influence of relative graph spacing on interpreting data	2*	No	3* (close, standard, far)	No		Same as Experiment Five, but with a full time series represented. There are six between-subjects conditions: three spread conditions x two candle body conditions.

Experiment Seven is a test of the applicability of findings when participants are asked to genuinely make decisions based on the graphs with which they are presented (Table 1.4). In Experiment Seven, the stimulus set is the same as in Experiment Three, but the task is not just to forecast what comes next in the financial time series. Rather, it is to bet on the stock closing higher after the next time point. The experiment tests whether the factors that improved performance in the earlier experiments (triangle candle body, no grid lines) can be added together to make a better overall plot compared to the factors that corresponded to worse performance in the earlier experiments (rectangular candle body, red/green representation, grid lines all of which also correspond to standard practice). The best and worst reasonable design choices are contrasted to see whether design impacts risky decision making — additional experiments can be performed to isolate each of the design differences, but the omnibus test is a good first step to discern whether differences in design impact performance at all. Experiment Seven is a test of the manipulations that corresponded to better performance in Experiments One through Six. This bridges the earlier experiments to a more applicable use case of candlestick charts in the context of risk and reward. There was no difference between groups. The best manipulations from the earlier experiments do not add up cleanly to make a better overall plot. Given this, and related work, it appears that cognition cannot be understood by adding up component knowledge from particular domains, but needs to be understood in context of task, environment and experience allowing for dynamic interactions between all three.

Overall, this work serves as an example of how to build up a scientific understanding of data visualization, bridging perception (all experiments, manipulations of plots' visual properties), abstraction (all experiments, learning easy representation-to-meaning abstractions in shadows versus harder abstractions in arbitrary symbols), information integration (observed in Experiment Three and specifically tested Experiment Four) and decision making (specifically Experiment Seven). As the experiments build in complexity, user performance is differently impacted by the visual properties of the environment and the task at hand. Simple information extraction is robust to minor visual manipulations, whereas better information integration is supported by using more symbolic data representations. The advantage of symbolic representations is only effective under certain task conditions; it does not impact participants' performance during risky decision making.

Table 1.4. Making data-driven decisions with candlestick plots

Experiment Title	Manipulations				Sample Image	Explanation
	Candle body	Noise	Spread	Grid lines		
7 Testing perceptual economics	2	No	Yes	Yes		Participants make bets about the values at the end of the next time point rather than completing the pattern like in the previous experiments. The two conditions represent standard practice (left) and data-driven best practice (right).

While there was no overall impact of Condition on user performance error during information extraction (Experiment One), there was a difference in how user performance error was distributed. The more salient candle body (Condition A) might have reflexively attracted more early processing (van Rullen & Thorpe, 2001, Theeuwes, 2010; van Zoest & Donk, 2007) but it did not invite better performance. Implications for theories of attention are discussed in Chapter Two, and connections between visual salience and more complex tasks are elaborated upon throughout. Generally, visual salience is a poorer predictor of user performance than higher level effects as tasks build in complexity, suggesting that effective theories of visual attention may be better served to prioritize endogenous factors to predict real world performance (McColeman & Blair, 2015; Navalpakkam & Itti, 2015; Tatler, Hayhoe, Land & Ballard, 2011; Wolfe, 2006; Theeuwes, 2010; Thompson, Blair, Chen & Henrey, 2013).

Chapter 2. The Foundation of A Science of Economic Data Visualization

The most basic method with which to respond to external stimuli is through reflexes or instincts, as is the case when the reflex arc is activated after grabbing a hot pan. As neural circuitry become more complex, so too do behaviours in response to the environment. The dung beetle, for example, is capable of using the polarity of moonbeams to guide its movements in a straight line (Dacke, Nilsson, Scholtz, Byrne and Warrant, 2003). The capacity to respond to the environment develops on evolutionary scales for lower animals, but animals with some memory capacity can learn how to use rules or integrate external information based on feedback and experience well within their lifespans (Blair, Watson, Walshe & Maj, 2009). Reactions to regularities in the environment are category responses.

Learning how to respond appropriately to categories is not a singular process. Two major types of categorization are rule-based and information-integration (Maddox & Ashby, 2010). They are learned differently, and have different optimal applications. Integrating information in an associative manner to make a category decision is valuable, but constructing rules in response to the environment (when possible) may be more powerful. A nuanced and complex representation of the world can be built upon the foundation of rules and logic to make sense of the environment in an analytical manner. Further, using rules also allows for formal updates to category representations or developing category flexibility.

Having flexible categorical representations requires continued responsiveness to input from the environment. In some cases, through apparently error-driven learning, responsiveness to the environment is diminished once an association rule is built. For example, rats learned to associate the presentation of a predictive light and a subsequent shock. When the light was paired with a tone, rats still displayed a learned association between the stimulus and the shock, but then when the tone was presented independently the association significantly diminished (Kamin, 1969) which is thought to be due to the lack of error in the system (Rescorla & Wagner, 1972): the light is as good of a predictor of shock as the combination of light and tone, so there's no need to integrate that additional information, and association between tone and shock is not

built. This is an example of the phenomenon “blocking”, which is observed during conditioning.

During category learning, though, additional information can be acquired. The process of categorization differs from the standard conditioning process in that additional information is helpful for building more robust category knowledge. Blocking[1] seems not to occur in category acquisition, following a couple lines of evidence.

First, there is a lack of bias to change information access patterns any more after error than after correct trials (McColeman *et al.*, 2014), which suggests that learning is not error driven (at least where error is defined by response accuracy). Following Rescorla and Wagner (1972), it would be expected that error reduction would be a primary motivator for learning. However, given that participants’ eye movements were the same following both correct and incorrect trials, it appears that information access patterns are not (just) responsive to error in the system. Additionally, because participants are willing to gather information after a sufficiently predictive cue is available, it appears that blocking does not occur in categorization (Bott, Hoffman, & Murphy, 2007). Since information access is not shaped by error, and since information access extends beyond the sufficient information, blocking appears not to apply to categorization.

Category flexibility, and the capacity for stimulus responsive attention, suggest that people can update their information access strategies. The development of stimulus-responsive attention over learning was previously observed in a task where category features were separated in space. Eye-trackers recorded where participants looked (Blair, Watson, Walshe & Maj, 2009). As people learned how to correctly categorize stimuli, they generally learned to look at the most informative features and then to look at the second-most informative feature.

---

[1] In standard conditioning, a conditioned stimulus is paired with an unconditioned stimulus, such that a natural (unconditioned) response is eventually elicited by the conditioned stimulus. Famously, Pavlov’s dogs would salivate to a bell/buzz if the sound had been previously paired with food enough times. Blocking occurs when a second conditioned stimulus fails to elicit the same unconditioned response. That is, if Pavlov’s dogs were already responsive to the bell, it’d be difficult to elicit a similar response to a second, simultaneously presented stimulus such as a flashing light.

Stimulus-responsive attention, ultimately, allows an observer to flexibly respond to a subset of the information in the environment to best deploy their finite cognitive resources in achieving their goals. In earlier studies, the relationship between stimulus responsive attention and decision-making is measured using a categorical outcome: whether or not the participant made the correct category decision (Aha & Goldstone, 1992; Blair, Watson, Walshe & Maj, 2009). Tautologically, category learning studies use category decisions to measure participants' understanding of the task. In the interest of both generalizing concepts from categorization literature and for moving forward an ecologically valid data visualization framework, the present studies use continuous responses (dollar estimates) to capture the same construct: the attention paid to visual information that informs decisions.

Candlestick plots are interesting stimuli for exploring perception and decision making. One reason for this is the responsive manner in which the reader must extract information from the graph: when the icon is red, the opening price is indicated at the top of the candle body, but when the icon is green the opening price is indicated by the bottom of the candle body. This is stimulus-responsive, in that the property of one stimulus dimension indicates where best to gather subsequent information. In category learning, stimulus-responsive attention is seen when one feature indicates the next best feature at which to look.

Experiment One builds upon earlier work on stimulus-responsive attention (Blair, Watson, Walshe & Maj, 2009). Rather than using spatially separated features, however, this experiment uses spatially integrated features. Because the features of interest are within a couple of degrees of visual angle from each other, eye tracking cannot offer evidence about the order to which features are attended. Performance errors, however, can indicate the order in which the dimensions (open, close, high and low) are learned, in that a sharper learning curve for one dimension would indicate a stronger understanding of that dimension earlier than a shallower learning curve observed for another dimension would. The practical question of this study is to determine how difficult it is to learn how to extract information from data represented on a standard candlestick plot (Figure 1.5)).

Experiment One also introduces a second contribution: a design innovation for a better user experience, by offering a more intuitive design for financial data visualization

in line with suggested practice in data visualization (Ware, 2004) and earlier work in visual attention that shows the early and profound influence of arrows on attention (Tipples, 2002; Ristic, Friesen & Kingstone, 2002). The standard candlestick plot (Figure 1.5) uses an arbitrary symbol; the colour of the candle body indicates the position of the open and the close price, but nothing intrinsic to that representation suggests that should be the case. Our modified candlestick plot simplifies the representation of the same data, by using what Ware would call a “sensory symbol” (Figure 1.5C). The expectation, building upon existing work, is that increased attention to task-relevant items (graphed elements to indicate the prices) should correspond to improved processing of that information relative to task irrelevant items, and as such to better decisions; when attention is exogenously driven by stimulus properties (such as an arrow pointing to the close price) then the dimensions represented by those properties will be better understood.

Building on the issue of the visual representation of data, the third contribution of Experiment One is a test of the effect of salience on participants’ performance: it’s expected that more salient dimensions should draw attention and lead to more accurate responses. This extends earlier work investigating stimulus-responsive attention by manipulating the salience of dimensions. Existing research does not manipulate the salience of category dimensions during stimulus responsive attention, and so testing the impact of visual salience offers some insight about the relative importance of external properties of the environment versus top-down influences on driving stimulus responsive attention. In Condition A, the open/close dimensions are more visually salient (brighter, more colourful and larger than other dimensions and brighter and more colourful than the background) while in Condition B, the high/low dimensions are relatively more salient than the remaining dimensions when compared to Condition A. Participants’ performance associated with each of these dimensions in the different conditions were compared and contrasted, with the expectation that salient dimensions would improve performance relative to less salient dimensions. The main research question, as such, is whether salient dimensions attract additional processing which corresponds to better accuracy than symbolic sensory representations.

2.1. Experiment One: Extracting Plotted Information from Candlestick Icons

An important step toward understanding how the perception of candlestick plots influences decision making is to understand how difficult the elementary perceptual task is. This experiment simplifies the task of making decisions based on candlestick charts by focusing just on the ease/difficulty with which participants are able to extract particular dimensions from individual candlestick icons. The location representing the open and closing price is indicated by the colour of the rectangular candle body. The standard candlestick graph uses red and green rectangular icons: it is an arbitrary symbol, in that it requires learning or training to interpret correctly (Ware, 2004).

It was expected that the arbitrary symbols (Conditions A and B) would be more challenging, and would take longer to learn how to interpret than the triangular candle bodies. In Condition A, because the candle body is more salient, it was expected that the performance of the high and low dimensions would be relatively worse than the performance of the open and close dimensions represented by the candle body. This hypothesis was built upon earlier findings, where salience was found to draw attention away from less salient items. In candlestick plots, of the sort in Condition A, the salient dimension is the candle body (the open and close price) and the less salient dimension is the candle shadow (the high and the low price). Salience of candlestick dimensions is manipulated to test if the visual properties of the graph impact the ability of observers to learn them.

2.1.1. Experiment One: Methods

Participants.

Participants were recruited from the research participation system at Simon Fraser University and received partial course credit in exchange for their time in this study. Participants were randomly assigned one of three primary conditions: red/green candle body (Figure 1.5, Panel A), light grey/dark grey candle body (Panel B), and the triangle with less salient shadows (Panel C). Of all participants, 30 were in the red/green candle body condition; 27 were in the light grey/dark grey condition; 31 were in the triangle body condition.

At the beginning of the experiment, some general demographic information was collected. Participants were prompted to optionally report their gender, experience with math and economics, and whether they were colourblind.

Apparatus and materials.

The experiment was written using a series of MATLAB scripts based upon the Psychtoolbox (Brainard, 1997; Kleiner, *et al.,* 2007). The stimuli were candlestick plots representing simulated financial data. The data were constructed using vector autoregression models built with the MATLAB economics toolbox. Forty unique vector autoregression models were each simulated ten times; each simulation returns a value for four dimensions (open, close, high and low prices) over 100 time points. These data were then exported as the information necessary to construct the stimuli: each vector represented one dimension per trial. The corresponding candlestick plot was constructed with shape drawing functions in the Psychtoolbox to represent the four dimensions.

Figure 2.1 A schematic representing a trial procedure.
Each trial has a fixation cross, a stimulus, a response phase where participants type in their answer (the example response here: 43.75), and then they receive spatial feedback. In the experiment, there were tick marks and value labels on the Y axis that are too small to show here.

Depending on the condition, the candlestick icons were drawn as shown in Figure 1.5: A) red boxes if the closing price was lower than the opening price; green boxes if the closing price was higher than opening price; with black shadows reflecting the high and low prices B) dark grey boxes if the close price was lower than the opening price; light grey boxes if the close price was higher than the opening price with yellow shadows reflecting the high and low prices C) grey triangles pointed down if the closing price was lower than the opening price; grey triangles pointed up if the closing price was higher than the opening price and black shadows reflecting the high and low prices

The difference in luminance between red and green candle bodies was not the same as the difference in luminance between the dark grey and the light grey candle

bodies; and so each condition had its own predefined shade of background grey. The goal of this was to ensure that candles that could be contrasted by luminance were not systematically easier or harder to identify relative to the background. For Conditions A and B, the grey background luminance was exactly the average of the luminance of the candle bodies. The Condition A background was 44.3% black. Condition B candle bodies were 25% or 75% black, and so the background was 50% black. For Condition C, the candle bodies cannot be distinguished by luminance, and so the background was simply 50% black (the same as Condition B). The percentage black values were determined in RGB colour space, and the images were presented to participants on 20" 2005 iMacs' LCD screens using MATLAB psychtoolbox software.

The Y-axis of the candlestick plot represented the price. There were 25 ticks on the Y-axis, each of which represents $0.50. Each whole number was labelled, while the $0.50 marks were not. The maximum Y-value was between $30 and $100 and varied trial-to-trial. In all cases, the Y-axis spanned 600 pixels, and an increase in one dollar was 48 pixels. The X-axis of the candlestick plot represented the time. There were no tick marks on the X-axis, but there was an axis label that said "time". Each trial stimulus consisted of one glyph selected randomly from 99 possible locations across the X-axis. In all cases, the X-axis spanned 81% of the total horizontal screen size (1360 pixels for a standard 1050x1680 display). The glyphs were 9.69 pixels wide on a standard 1050x1680 display.

Procedure.

The data were collected in private rooms, and all participant responses (including the small survey about experience with math and economics) were entered in privacy. The participants were randomly assigned to a condition by the experiment program ahead of time, so the experiment was double-blind. All participants, regardless of condition, received the same set of automated task instructions that differed only in the sample images which were responsive to the condition that the participants were assigned to.

The study went for 25 minutes or until the participant completed 40 trials. Each trial began with a central fixation cross (each line segment was 3% of the height of the screen) which stayed on the screen for roughly one second, then the stimulus was shown. The stimulus included everything that would be expected from a standard graph:

an X axis (labelled “time”), a Y axis (labelled “price”), a title (a fictitious stock code), and a single candlestick glyph (see Figure 1.5). A “next” button was to the right of the graph to allow the participant to advance to the next stage where they were presented with a question and allowed to make a response. The questions were written above the next button as “what is the [opening/closing/highest/lowest] price?” and the participants type their answer in the response box.

Every twenty trials, the participant was encouraged to take a break, told how many points they had accumulated and challenged to score the number of points equivalent to answering perfectly on five trials over the next set of twenty. The error value was a function of the number of pixels between the correct answer and the participants’ response if it was going to be graphed. If, in the example, the participants’ response converted to pixels would be 25 pixels higher than the bottom of the candle body (the correct answer), then the participant would lose 25 points from a possible 125 on that trial. The maximum number of points is 125, and the minimum number of points is -500. To ensure participants received positive feedback for good performance, it was important that close estimates corresponded to higher values. The value of 125 was chosen to allow participants some room for error while still receiving positive feedback. The minimum value was selected to keep scores within reasonable bounds. It was possible to lose points on a trial, but there was a lower bound to mitigate the negative impact of typos. To be more than 500 points off of the correct answer almost always corresponds to a point that was beyond the range of the plotted y values, as the average vertical position of the icons is half way up the screen (525 pixels), and so it’s assumed that a response corresponding to more than 500 points off is not a genuine reflection of the participants’ estimate.

2.1.2. Experiment One: Results

The User Performance (UP) is of primary interest throughout these analyses: the difference between the participants’ responses and the models’ predictions (the user performance error or UPE) should be small if the participants were able accurately assess the values that they were asked to report. However, if participants struggled to extract information from the presented icon, then the UP error would be larger. The first step is to examine the overall error by candle condition (Fig. 3) to see if the physical representation of information impacts observer performance.

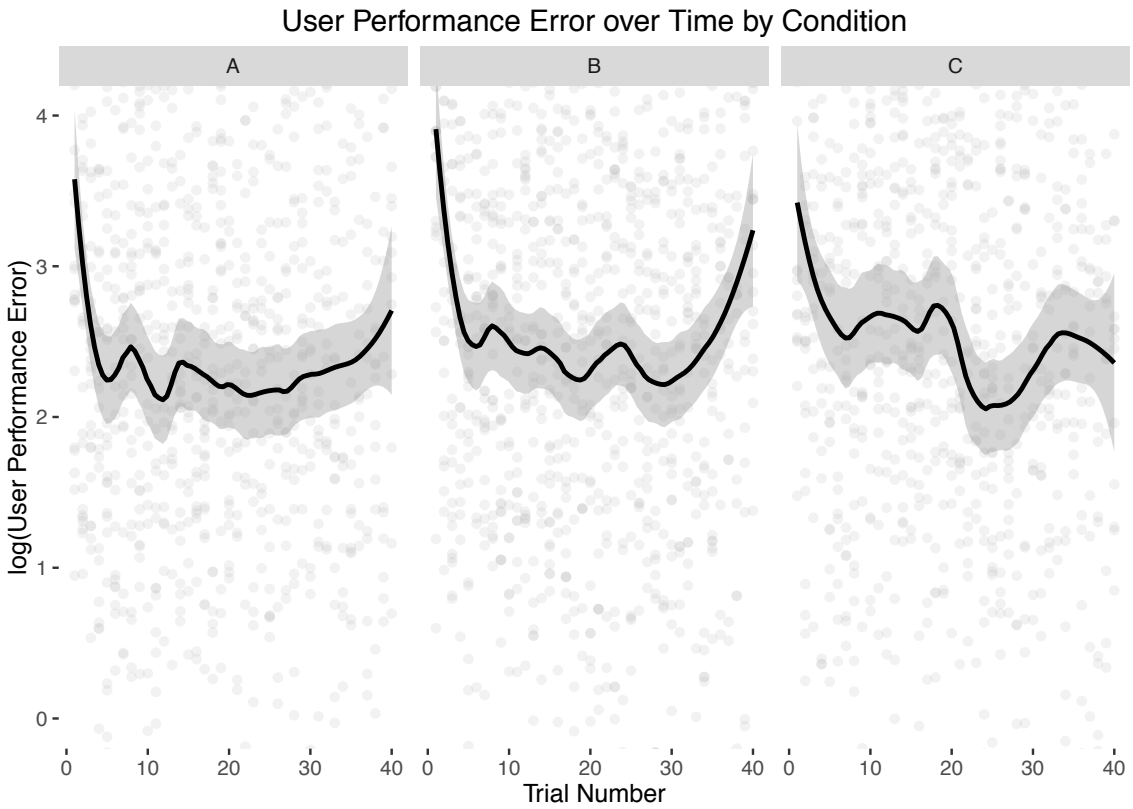


Figure 2.2 User Performance Error in Experiment One.
Each panel of this plot represents one Condition. There is no difference overall by Condition, but there is an effect of trial on User Performance Error. Each data point in the experiment is shown (grey dots), and the dark line is a LOESS[2] fit to the data for each condition. The grey shadows reflect the standard error surrounding the LOESS curve.

User performance error (UPE) is the distance between the participants' response and the actual value of the queried dimension. If the participant was asked to report the high price, for example, and responds as if it was higher than the icon indicated, then the error associated with the high price was the difference between the participants' estimated "high price" (if it were to be drawn on the graph) and the actual icon's "high price" in pixels. If participants over estimated the high price by $0.25, the UPE for that trial is .25 x 48 = 12, because $0.25 is .25 of a dollar, and a dollar corresponds to 48

[2] The LOESS (local regression) curve is a non-parametric fit of the observed data to its neighboring points. These particular curves were produced in R, using the GGPlot package, specifically, the geom_smooth argument. The "span" argument passes the alpha value to the LOESS function, where alpha determines the number of neighbouring data points that inform the local fit. These data were fit with span =1, and as such all data played a role in fitting the curve. LOESS fit assumes no pre-specified shape, and so it can more effectively respond to the genuine properties of the data in contrast with a linear regression which would assume the data to change linearly. Given that there was no prior knowledge to anticipate a linear relationship between these variables, the LOESS curve is a preferable method to summarize the change in user performance error over time.

pixels. The participant would then receive 125 - 12 points on the trial, which they would be told during the feedback phase. To generalize; the UPE is the error associated with participants' estimates for the queried dimension, where error is the distance (in pixels) between the model generated correct values and the participant's estimate if it were to be drawn on the graph. The higher the UPE score, the worse the performance: the goal of the respondent is to match the presented values as closely as possible. The points awarded to participants is calculated by 125 - UPE.

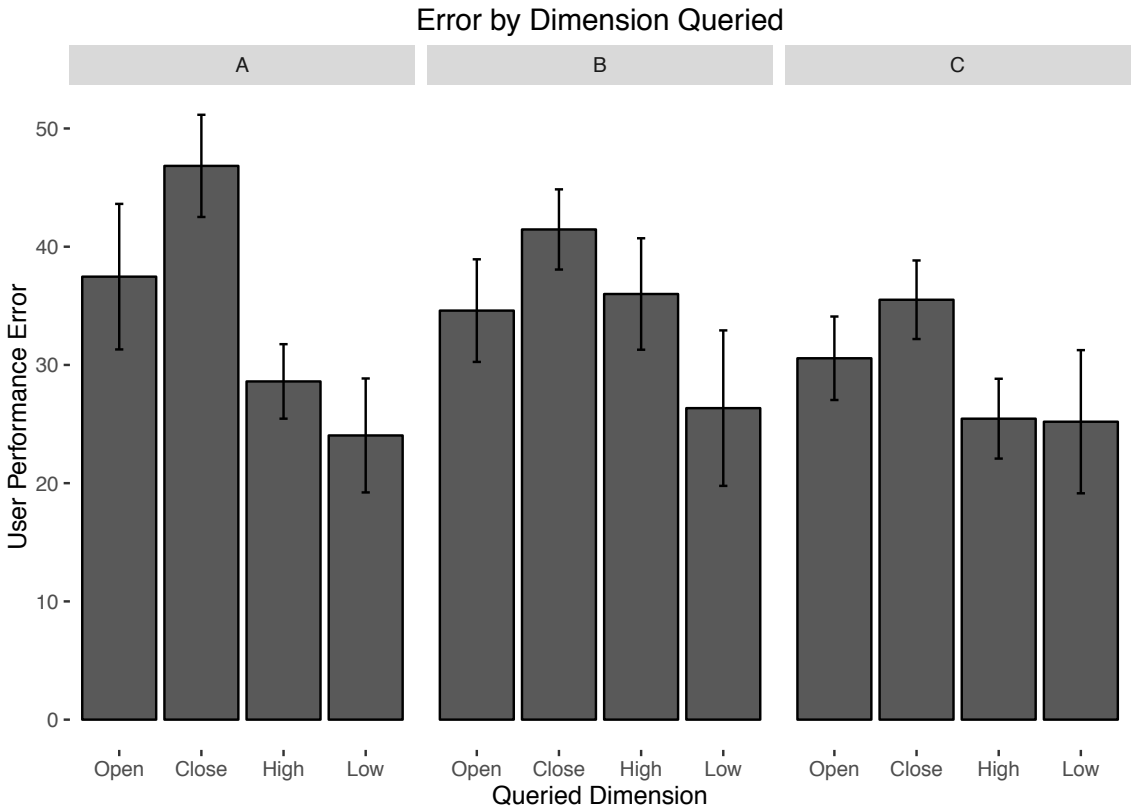


Figure 2.3. User Performance Error in Experiment One by Dimension.
Each panel of this plot represents one Condition. For each trial, the participant is asked to provide the value associated with a single dimension. The queried dimension is on the X-axis. Notice generally that error associated with Open and Close dimensions is higher. The error bars reflect the standard error of the mean.

A linear mixed model was constructed to analyze differences between groups after accounting for individual variation (Woltman, Feldstain, Mackay & Rocchi, 2012), and to reflect the hierarchical nature of the study design: response error is nested within trial, which is nested within subject. The linear mixed effects model controls for variance associated with the number of trials for each participant, too, by having Trial as a predictor. The queried dimension (open, close, high or low) was included in the model to

test whether the bodies associated with salient portions of the icon improved performance. The R package lme4 was used to build the model and fit the data (Bates, Maechler, Bolker, & Walker, 2010). Because the red/green candlestick plot is the standard version of a candlestick plot, it was the reference category in the analysis model for this study. The other conditions are analyzed relative to Condition A.

Table 2.1. Coefficient estimates from linear mixed effect model predicting log(UPE) in Experiment One over all dimensions

	Coefficient Estimate (β)	Standard Error	*T*	*p*
(Intercept)	2.8987	0.1514	19.1486	0.0000
TrialID	-0.0124	0.0030	-4.1180	0.0000
Close	0.3294	0.1463	2.2523	0.0243
High	-0.3056	0.1450	-2.1078	0.0350
Low	-0.5410	0.1735	-3.1178	0.0018
Condition B	-0.0415	0.1897	-0.2187	0.8269
Condition C	-0.1147	0.1960	-0.5854	0.5583
Close x Condition B	-0.1024	0.2000	-0.5119	0.6087
High x Condition B	-0.1612	0.1984	-0.8122	0.4167
Low x Condition B	-0.2632	0.2355	-1.1178	0.2637
Close x Condition C	-0.1620	0.2072	-0.7822	0.4341
High x Condition C	-0.1568	0.2061	-0.7607	0.4469
Low x Condition C	-0.1688	0.2436	-0.6929	0.4884

The linear mixed effects model revealed a main effect of Trial on user performance error. The combination of this finding and visual inspection of the LOESS fit suggest some decrease in user performance error over time. There was no discernible effect of Condition on User Performance Error (estimated β magnitudes <.115, *p*s > .57) or interactions between Condition and Trial. There was, however, a surprising difference between queried dimensions. In all three conditions, the close dimension yield the worst performance. Contrary to earlier predictions, the low dimension yielded the best performance. It is worth noting, however, that the difference between the relatively bad performance on the close dimension and the good performance on the low dimension diminishes in the Conditions B and C.

2.1.3. Experiment One: Discussion

There were three primary goals from this first experiment. One was to discern how difficult it is to learn data represented on a standard candlestick plot. The results indicate an effect of Trial, and, upon visual inspection of the plotted data over trial, it suggests that some learning has occurred. Ideally, when possible, graphs should be easy to understand and since it requires seeing a graph multiple times to understand its contents, then there is probably room for a design innovation. While it may be that apparent learning effects were due to participants adjusting to the task and the environment, further work will test whether some of this adjustment phase can be reduced by better glyph design (e.g. triangle candle bodies in a more complex task) or whether improved performance over time is simply due to the new circumstance in which participants find themselves.

In preparing to step toward a science of visualization, this experiment also acts as a first look at the influence of salience on information extraction in the context of an economic problem. In Condition A, the difference in salience between the candle body and the shadows is larger than in Condition B. Earlier work has shown that visual salience impacts early visual processing (van Rullen & Thorpe, 2001, Theeuwes, 2010; van Zoest & Donk, 2007) but relatively little is known about the lasting impact of visual salience on decision making. Connecting early visual perception, attention and cognition (Gottlieb, 2007) is a challenging, but critical pursuit; manipulating the visual properties of data represented in a graph and collecting participants' response is one way to tackle it. If the candle body is more salient, it is expected that it would draw additional attentional resources (Jarvenpaa, 1990); if those attentional resources support better understanding, then the same salient candle body should positively impact participants' understanding. The manipulation of the candle's visual properties did not impact performance overall, and so salience did not appear to influence responses.

Finally, this experiment is the first of a series of experiments that test user performance when reading graphs exhibiting data by "sensory" symbols (Ware, 2004) instead of arbitrary symbols. Ware's sensory symbols are those that can be understood without training. Sensory symbols generally work well across cultures and may be perceived rapidly or in parallel in the context of many sensory symbols. It is difficult to train sensory symbols to mean something opposite their natural interpretation; a line

chart that represents sea depth is challenging to interpret if the increasing line height corresponds to deeper values, for example.

Arbitrary symbols are those that require training or prior knowledge to understand. They are more challenging to learn, easy to forget, and subject to cultural differences due in part to their abstract nature. There is an advantage of using arbitrary symbols in data visualizations, however, in that their abstract properties allow them to represent complex ideas. A good example of arbitrary symbols in action is mathematical notation, where a single symbol "√" represents a relatively complex concept of a number that can be multiplied by itself to equal the input value. The square root takes some instruction and effort to learn, but once that training is complete, the symbol can be used in even more complicated equations to efficiently communicate the concept.

In this experiment, the sensory symbols are the triangular candle bodies and the arbitrary symbols are the rectangular candle bodies (Conditions A and B). Ware suggests that sensory symbols should be easier for the reader, and so it was expected that the triangular candle bodies would exhibit less User Performance Error on the open and close conditions than Conditions A and B. However, this was not found in Experiment One — at least not in the simplest sense. For simple visualization tasks, then, arbitrary symbols, like the square root sign, might be well suited to conveying concepts. The task at hand is important to consider. In simple information extraction from glyph visualizations, perhaps arbitrary symbols are preferable to sensory symbols because they offer more opportunities to communicate complex concepts to the reader. As will become evident in later experiments, sensory symbols do appear to better communicate their meaning in more complex tasks.

In all, Experiment One offers insight into a) whether the candlestick plot requires learning to use; b) the influence of salience on graphical information extraction generally and in stimulus responsive attention; and c) the relative advantage of sensory symbols. It is found that performance does improve in the start of the experiments, indicating some learning effects. The sensory symbol (Condition C) does not appear to be much better than the arbitrary symbols (Conditions A, B), at least in the simple task of extracting values from graphed data.

2.2. Experiment Two: The impact of grid lines on accurately extracting information from a candlestick plot

The findings in Experiment One may be due to the relative difficulty of creating a spatially accurate visual memory trace under different stimulus conditions. The memory trace is necessary because the stimulus is not available when the participant is asked to respond with the value for a particular dimension. Maintaining a sense of position without grid lines is challenging. Experiment Two is a test of the presence of perceptual “hints” in the form of grid lines and whether they make recalling information contained within the candle glyphs any easier. The visual meta-data (Stone & Bartram, 2010) or perceptual anchoring that grid lines afford is thought to make it easier for an observer to establish an object of interest in visual space.

Reports of the efficacy of gridlines are sparse, and suggest that they are helpful only in particular conditions. Gridlines may be a violation of the Principle of Relevance (Kosslyn, 2006) which states that observers expect only to see meaningful information when they observe a graph. While some visualization experts suggest that “grid lines in graphs are rarely useful” (Few, 2005), more assume that they are helpful as long as they are not too light (Stone & Bartram, 2010) or too dark (Kosslyn, 2006). Earlier work shows that grid lines are more valuable in line graphs than in bar graphs (Lohse, 1991), which is intuitively plausible: they help in some task conditions but not others. This experiment offers a couple of critical data points to contribute some understanding about their value. It is the first known exploration of the value of grid lines in reading information from candlestick plots, so it has immediate practical benefit for users of the candlestick plot.

Beyond the basic use case, though, this study offers a theoretically valuable dataset: if gridlines are simply “chartjunk” (Tufte, 1983), or if they violate the Principle of Relevance (Kosslyn, 2006) then they should detract from participants’ performance. Grid lines are irrelevant in that they add no meaning to the visualization. While they may help observers in their spatial navigation, they do violate the strict read of the Principle of Relevance. To empirically validate suggested design practice, Experiment Two tests the impact of gridlines. Based upon the experience and suggestions of expert practitioners, it’s expected that gridlines will negatively impact performance in Experiment Two when compared to data from Experiment One. Alternatively, gridlines may make it easier to extract information by having an external visual reference beyond the object which

would, at least, introduce a corollary to the Principle of Relevance which holds that observers expect to see only and all relevant data represented in a graph, and that so additional information violates expectations.

There is a question of basic visual cognition in this study, too. Given the structure of this task, where the participant views the stimulus, the stimulus is removed and the participant responds from their memory of the stimulus, the participant must rely on visual working memory to make a response. If localizing an object in visual memory is easier when there are gridlines in the task, that suggests that glyph's memory trace is supported by reference points (Sadalla, Burroughs & Staplin, 1980): it's not just where the glyph was, but where it was relative to the gridlines. The grid lines were not expected to interact with any earlier effects of Condition.

2.2.1. Experiment Two: Methods

Experiment Two is methodologically identical to Experiment One, except for the presence of grid lines. The grid lines are drawn behind the icons, on every Y tick mark and are 20% brighter than the background. The 20% value is based upon empirically tested design practice, which suggests that 20% transparency for grid lines is a good default to support observers' graph reading without being intrusive. In the Stone and Bartram study (2010), two groups were asked to set the transparency value for black grid lines. One group was asked to set the value so the grid lines were just perceptible (but still useable) and the second group was asked to set the value so the grid lines were as dark as they could be without being intrusive. There was no group asked to perform the task without grid lines. The original study found that 10% was a good cut off for as light as possible and around 35% was a good cut off for as dark as possible. A crowd-sourced replication of that study shows the same general pattern of results with a general preference for slightly darker grid lines (Heer & Bostock, 2010). The authors of the original study suggested an alpha of 0.2 as a safe default value for grid lines, further supported by the authors of the follow up study.

2.2.2. Experiment Two: Results

The analyses from Experiment One are repeated. Again, a linear mixed effects model is constructed with Trial, Condition and Queried Dimension as factors. There is a

main effect of Trial, reflecting some early improvement in performance. There is an effect of queried dimension for when participants were asked to respond with the High and Low price (compared to the open price), but not for the Close price. In this experiment, there is also an interaction between particular dimensions and Condition: when asked to report the High price, participants were better in Conditions B and C than in Condition A.

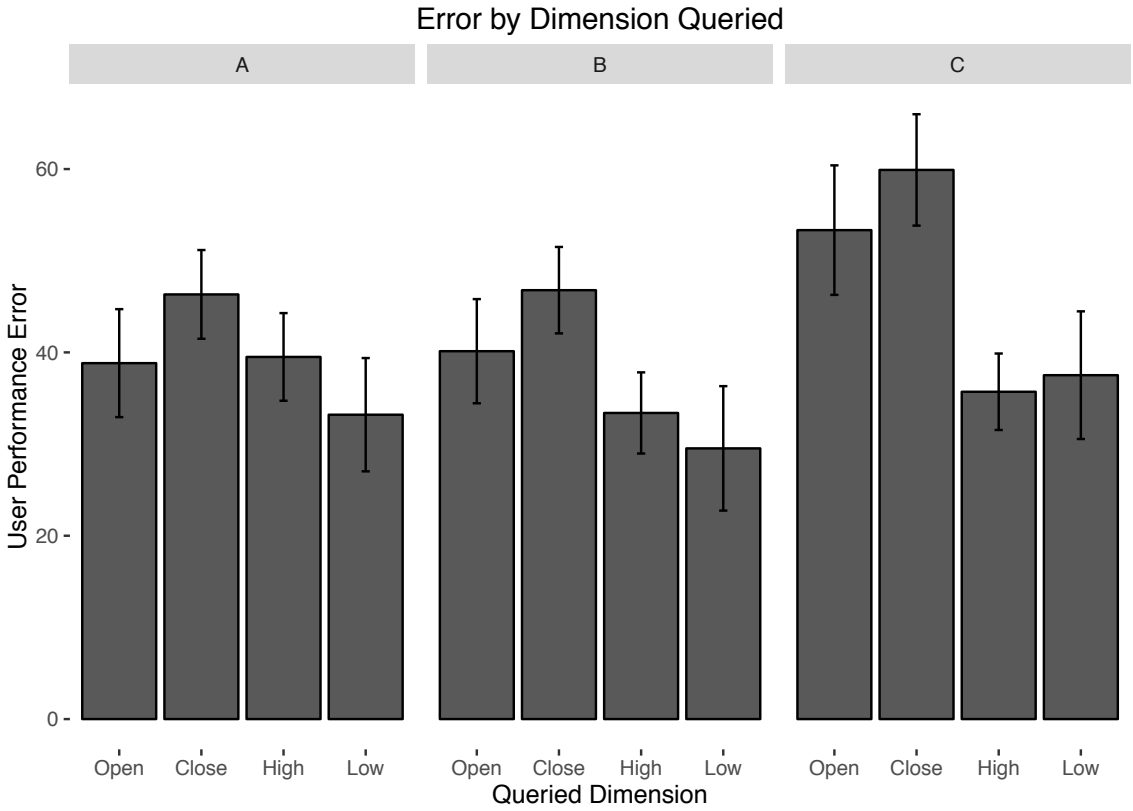


Figure 2.4. User performance error in experiment two by dimension.
Each panel of this plot represents one Condition. The queried dimension is on the X-axis. Notice generally that error associated with Open and Close dimensions is higher. The error bars reflect the standard error of the mean.

Additionally, an examination of the direct advantage of grid line is conducted here by comparing Experiment Two data to Experiment One. Overall, the presence of grid lines negatively impacts performance. The mean UPE in Experiment One is lower than the mean UPE in Experiment Two (with gridlines) as indicated by an independent samples T-Test (T = 2.549, df = 150.32, p = 0.011). To test whether the grid line difference was due to a particular condition, a family of T-Tests were subsequently conducted by Condition. Because this is a family of tests, the corrected alpha value to

reject the null hypothesis that the two groups are equal is .05/3 = 0.017. There was no overall difference for Condition A (T = 1.102, df = 50.189, p = 0.2759) or Condition B (T = 0.5378, df = 55.664, p = 0.593). Rather, much of the negative impact on user performance with the introduction of grid lines was for participants in Condition C (T = 2.5148, df = 43.775, p = 0.016). While grid lines have little impact for the rectangle body conditions, they negatively impact participants' ability to accurately extract information from the adapted candlestick plots (Condition C).

Table 2.2. Coefficient estimates from linear mixed effect model predicting log(UPE) in Experiment Two over all dimensions

	Coefficient Estimate (β)	Standard Error	*T*	*p*
(Intercept)	2.8928	0.1719	16.8287	0.0000
TrialID	-0.0225	0.0037	-6.1622	0.0000
Close	0.0033	0.1582	0.0211	0.9832
High	-0.5178	0.1561	-3.3175	0.0009
Low	-0.8199	0.1900	-4.3141	0.0000
Condition B	0.3176	0.2276	1.3951	0.1630
Condition C	0.3547	0.2130	1.6650	0.0959
Close x Condition B	-0.1343	0.2289	-0.5866	0.5575
High x Condition B	-0.5762	0.2272	-2.5358	0.0112
Low x Condition B	-0.5057	0.2729	-1.8531	0.0639
Close x Condition C	-0.3664	0.2186	-1.6764	0.0937
High x Condition C	-0.5409	0.2147	-2.5192	0.0118
Low x Condition C	0.1762	0.2613	-0.6744	0.5001

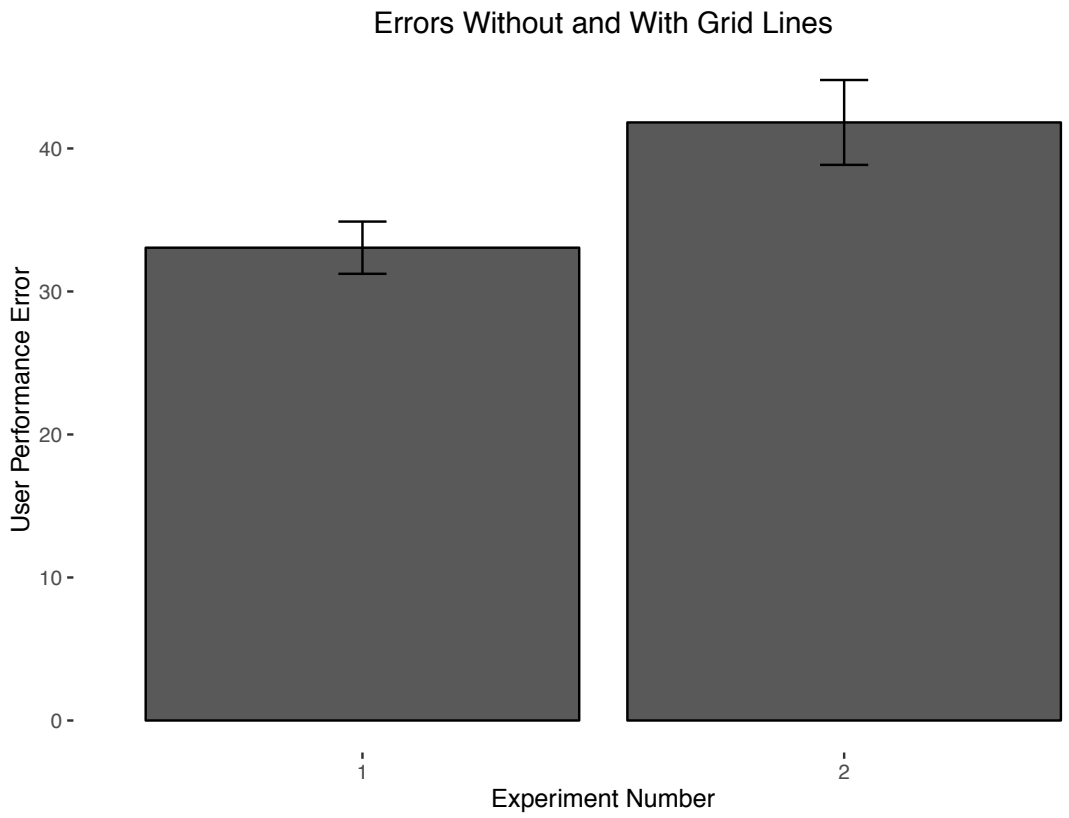


Figure 2.5. The difference in User Performance Error between Experiments One and Two.
The mean User Performance Error for each experiment overall is shown here. The error bars reflect standard error of the mean.

2.2.3. Experiment Two: Discussion

The difference between Experiment Two and Experiment One was of primary interest. There was an effect of grid lines for Condition C, in keeping with the design principles of Kosslyn and Few: gridlines negatively impact performance accuracy when extracting information from candlestick charts. While this was consistent with some design principles, it does pose some challenge to proposals about relational encoding (e.g. Brady & Alvarez, 2011).

User performance was better when participants were asked to report the high price in Conditions B and C than Condition A, which may be a function of the relative attention dedicated to the salient candle body in Condition A. If there are finite resources to allocate to information encoding, and salience draws an unfair share of those resources, then there are fewer resources to give to effectively encoding less salient

items, like the shadows that reflect the high and the low price. In Condition A, those shadows are relatively less salient (compared to the candle body) than in Condition B and C.

Experiments One and Two capture the essential properties of extracting information from icons of various forms. However, a goal of this thesis is to move toward a scientific framework of data visualization that includes decision making, and so the next experiment builds upon the previous two to test the impact of these design manipulations on information integration in preparation for decision making. Experiments One and Two establish a foundation for understanding which glyphs are more difficult to learn, and visual properties that impact performance. Generally, performance during information extraction is similar for the different glyph conditions, but information extraction for the high and low dimensions is better. Visual extent, or position, does appear to foster more precise judgements. Moving toward a more ecologically valid test, though, requires using candlestick glyphs to represent time series information, as they are intended to do. The next step is thinking about larger, integrated pattern perception.

2.3. Experiment Three: The Influence of Salience on Forecasting

This experiment explores a more complex task, where instead of simply responding with the value represented by a glyph, the multiple dimensions of the glyph are presented among many other glyphs to represent a pattern. In this Experiment the observer is completing the pattern. Experiment Three explores the extent to which shape and colour influence observers' perceptions of patterns in simulated financial data and impact decision making. This is an example of the class of studies previously identified as *User Performance* studies (or "UP") as it investigates how properties of a visualization impact an observer's performance against an objective ideal (Lam, Bertini, Isenberg, Plaisant & Carpendale, 2012). In this case, the objective ideal is the final simulated value in a time series model. Participants are asked to supply their answer based on the rest of the time series, and are awarded points for how close they are to the simulated values.

2.3.1. Forecasting

While research in human decision making and category learning typically explores the cognitive processes underlying discrete decisions, forecasting traditionally focuses on the cognitive processes underlying continuous decisions informed by time series data. An intuitively appealing way to understand forecasting is to consider weather forecasts, which can be done, but with some caution. In formal meteorological models, forecasts are informed by time series data which is important to forecasting; a layperson is less likely to use time series data to guess whether they need to bring an umbrella to work tomorrow. The distinction between the meteorological use of forecasts and the public understanding of forecasts is exhibited in survey data where respondents on various city streets interpret “a 30% chance of rain” differently than meteorologists meant them to (Gigerenzer, Hertwig, van den Broek, Fasolo, & Katsikopoulos, 2005) When meteorologists report a 30% chance of rain, they mean to say, given current conditions, 3 out of 10 times it would rain. Rather, survey respondents interpreted the same information to mean “it will rain 30% of the day”, or “it will rain in 30% of this area”. The authors attribute forecast misunderstandings in this domain to poor communication from experts to laypeople, although innumeracy or general discomfort with interpreting numbers is not explicitly ruled out.

Humans may not be great at interpreting probabilities, but we are quite good at detecting patterns. Arguably, our capacity for pattern recognition exceeds our capacity for decision making. Forecasting, or estimating an outcome based on time series data, is one part pattern recognition and one part decision making. It requires an observer to extract information from observed time series data, and then, using information from that pattern, proceed to estimate the outcome at some point after the last observation of time series data (essentially making the decision but with a continuous outcome). More research has been conducted on constructing formal models of forecasting and time series analysis than on how humans perceive forecasting problems. It’s understandable. There are benefits to having a good formal model that can effectively estimate future events in a tractable way. In some cases, though, a human is just as good (Lawrence, Edmunson & O’Connor, 1985), and so having a person (or people; Dalkey & Helmer, 1963), to extrapolate data might be the optimal approach to estimating future events. That said, given advances in machine learning and data science, if accurate estimation of continuous future outcomes was the whole problem, then formal models would

eventually best human decision makers (e.g. Grove, Zald, Lebow, Snitz & Nelson, 2000).

Even if economics (or science generally) has reached a point where formal models are better predictors of future events than humans, understanding how humans gather information and extract regularities is valuable in understanding human cognition itself. In categorization, common elements of individual observations are abstracted to indicate group membership and make it possible to use the concept of that group for higher cognition. In forecasting, common changes between observations of (sets of) time points are abstracted to yield a pattern, and make it possible to use that pattern for higher cognition such as predicting subsequent observations.

If an observer is watching a ball roll down a ramp, where part of the ramp is behind a curtain and they try to guess when it would emerge past the curtain, they are extrapolating from an observed pattern. Since it's going down hill, the ball will accelerate, and since it's partially occluded on the ramp, the observer will have to extrapolate the velocity and acceleration to correctly anticipate when they will see the ball again on the other side of the occlusion. There's little abstraction to be done here, as the ball is its own best representation and the ball itself is visible again when it's recognized on the other side of the curtain. Nonetheless, the observer is extrapolating the pattern of movement from the data they collected while the ball was rolling down the visible part of the ramp to anticipate its arrival further down the ramp. An empirical investigation of this scenario shows that observers accurately track the velocity of moving objects well, even if their view of the object is obscured. Accelerating objects can be tracked too, though tracking high acceleration is more difficult (Rosenbaum, 1975).

Whether the motion perception architecture of the human brain (Bischof & Di Lollo, 1990) that allows for observers to track a moving ball behind an occluder can be co-opted to host higher level perception of the rate of change remains to be seen. Regardless of the biological basis of this skill, Rosenbaum's (1975) study shows that human observers can track a moving and accelerating object without persistent visual cues, which requires some level of encoding (if not some simple abstraction) to estimate when the object will reach a particular point in its path. This is a primitive time series forecasting problem. The observer sees the ball for some time and some space while it moves along the visible part of the ramp, and is asked to use that information to estimate

when it will reach a later threshold (or estimate the outcome after a certain amount of time has elapsed).

Time has been touted as an important factor in understanding cognition (van Gelder & Port, 1995; Kingstone, Smilek, Risitic, Friesen & Eastwood, 2003) and for good reason: information in the environment is not static, and neither are percepts and ideas. Time series, however, can be presented in a way that simplifies some of the dynamic complexity that arises when percepts and concepts unfold over time. Time series visualizations can capture changes over time and represent those changes in a single plot. Most people have seen line graphs that represent an object's "distance over time" in high school math classes, for example. This is a two dimensional, static abstraction of a more complex scenario. Epidemiologists can look at instances of HIV over multiple years to represent time series, and traders can look at opening, closing, high and low prices of a stock over time to gather insight into whether they'd like to buy it.

2.3.2. Formal approaches to time series analysis

Time series analysis is the precursor to forecasting. It captures the properties of observed data over multiple time points. It's similar to analyzing a correlation: observing data on two dimensions (for example, closing price and day) and examining how the two variables change together. It's easy to envision a line of best fit through a scatter plot representing closing price against time and using that line to infer something like "the closing price goes up as time goes on". Simple conclusions are good for simple data, but formal models are valuable assistants for when the scenario becomes more complex.

For example, it might be the case that each observation is partially predicted by the observation from the time step before it. This phenomenon can be captured formally by an autoregressive (AR) model. If it is the case that each an observation at the current step is partially predicted by the time step prior, it can be described by an AR(1) model: that is, an autoregressive model with a lag of one. If each time step was partially predicted by observations up to two time steps prior, it'd be an AR(2) model; three time steps prior, an AR(3) model and so on.

Real data are rarely so clean that a single variable changing over time is enough to paint a whole picture. Vector autoregression captures the same change over time as

univariate autoregression, but also considers how multiple variables impact each other (Sims, 1980; Sargent & Sims, 1977). Using a vector autoregression model to generate data is one way in which we can assess realistic multivariate data visualizations, wherein the variables have an impact on each other and many variables need to be visually represented for an observer to make sense of the data. This Experiment Three, as with all previous experiments, uses vector autoregression models to generate financial data patterns that act as the stimuli.

2.3.3. Experiment Three: Methods

Participants in this study were shown simulated financial-type data patterns visualized using candlestick plots. Each stimulus was comprised of 99 candlestick glyphs. After viewing the stimulus, participants were asked to use the information they gathered from those plots to draw in the glyph that came next in the series. The stimulus is not on screen while the participant makes a response. This is in order to query the participants' working memory of the graphed dimensions, rather than how effectively they can read information directly off of the graph. While this is at some cost for ecological validity (users usually have graphs in front of them when they make decisions) it does offer some theoretical advantages in that lessons from the experiment can be used to update cognitive models of data visualization with working memory isolated from visual sensory memory.

Participants

Participants were recruited from the research participation system at Simon Fraser University and received partial course credit in exchange for their time in this study. They were undergraduate students enrolled in first or second year psychology classes. Participants were randomly assigned one of three primary conditions: red/green candle body (Figure 1.5A; n = 38), light grey/dark grey candle body (Figure 1.5B, n = 46), triangle with less salient shadows (Figure 1.5C, n = 55), and a triangle with more salient shadows (Figure 1.5D, n = 43).

At the beginning of the experiment, some general demographic information was collected. Participants were prompted to optionally report their gender, experience with math and economics, and whether they were colourblind.

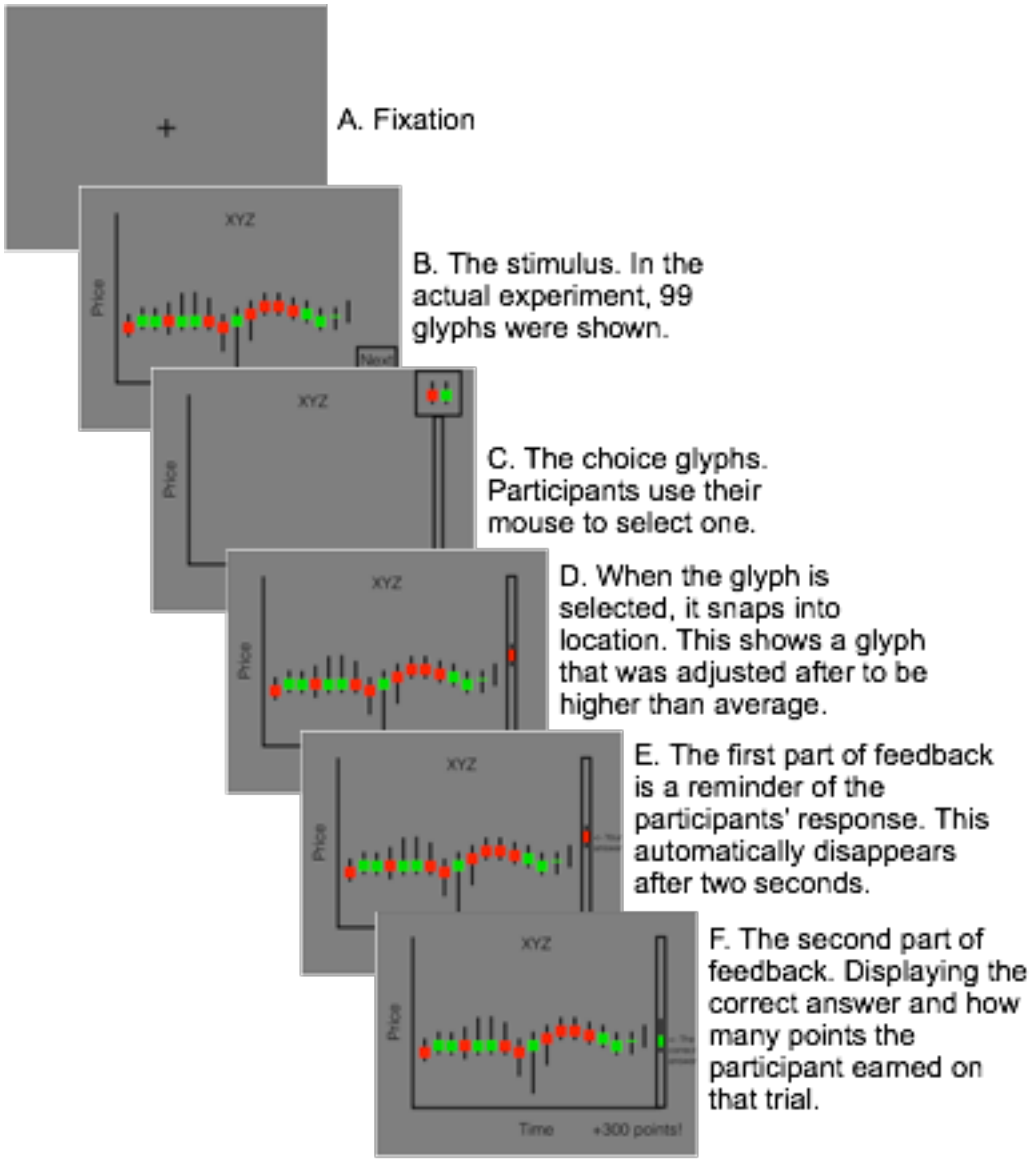


Figure 2.6. A schematic representing a trial.
Each has a fixation cross, a stimulus, a response phase where participants select and adjust their answer, and then they receive spatial feedback.

Apparatus and Materials

As with the earlier experiments, the Y-axis of the candlestick plot represented the price. The X-axis of the candlestick plot represented the time. There were no tick marks on the X-axis, but there was an axis label that said “time”. The stimuli consisted of 99 glyphs presented in equal intervals plus one placeholder for the response. In all cases, the X-axis spanned 81% of the total horizontal screen size (1360 pixels for a standard 1050x1680 display). The glyphs were drawn with 25% spacing between them and 150

pixels from the leftmost edge of the screen, so the width of each glyph was (X-axis span - 150)/(number of glyphs*1.25) or 9.69 pixels on a standard 1050x1680 display.

Procedure

Again, all data (including a short survey) were collected digitally in private rooms. Participants were randomly assigned to a condition by the experiment program ahead of time, and the experimenter does not know who is in what condition. This experiment is double-blind. At the start of the experiment, after completing the brief survey, all participants receive the same set of automated task instructions. The instructions differ in the sample images which are responsive to the condition that the participants are assigned to.

The study runs for 55 minutes. Each trial begins with a central fixation cross (each line segment is 3% of the height of the screen) which stays on the screen for roughly one second, then the stimulus is shown. The stimulus includes everything that would be expected from a standard graph: an X axis (labelled “time”), a Y axis (labelled “price”), a title (a fictitious three-letter stock code), and a series of candlestick glyphs (see Figure 2.6). When the participant clicks “next”, the graph disappears and two glyphs are available to select. The participant uses the selectable choice glyphs to complete the pattern exhibited by the stimulus (which, in turn, was established by a vector autoregression model in MATLAB). The choice glyphs differ only in colour. The length of the candle body and the shadows is the length of the average of all candle bodies and shadows in the trial, respectively. The choice glyphs are centred around the average of the location of the stimulus glyphs in the trial. When one is selected, it snaps into the average position, and it can be adjusted: the candle body height can be changed, the shadow heights can be adjusted separately and the whole unit can be moved up and down. One viable strategy is then just to pick a glyph and leave it at the same height and skip adjusting it. People are encouraged to take the time to adjust the glyph they chose so it looks like it’s likely to occur next. When they are satisfied with their adjustments, participants click “next” to advance to the next phase where they receive feedback.

Every twenty trials, the participant is encouraged to take a break, told how many points they have accumulated and challenged to score the number of points equivalent to answering perfectly on five trials over the next set of twenty trials. The error value is a function of the number of pixels between the correct answer and the participants’

response. The maximum number of points is 500, and the minimum number of points is -2500. It is possible to lose points on a trial, but there is a lower bound to mitigate the negative impact of typos. To be more than 2500 points off of the correct answer almost always corresponds to a point that is beyond the range of the plotted y values, or a massively stretched out candle body which isn't observed in any of the stimuli anyways.

2.3.4. Experiment Three: Results

The User Performance (UP) was of primary interest throughout these analyses: the difference between the participants' responses and the models' predictions (the UPE) should be small if the participants were able to discern the underlying model from the representation of the data it generated. However, if the model was more difficult for participants to understand, then the observed error would be larger. The first step is examine the overall error by candle condition (Fig. 7) to see if the physical representation of information impacts observer performance.

User performance error (UPE) is the distance between the four dimensions that were simulated by the underlying time series model and the dimensions that were predicted by the participant. For example, if the participant overestimated the high price by adjusting the glyph so it was 10 pixels higher than the model's prediction, then the error associated with the high price is ten. If they underestimated the low price by 5 pixels, the open price by 7 pixels and the close price by 20 pixels, the UPE for that trial is 10+5+7+20 = 42 (a very good answer!) To generalize; the total UPE is the sum of the error associated with each of the four dimensions, where error is the distance (in pixels) between the model generated correct values and the participant's estimate. Distance is an absolute value, and so the direction of the error is not factored into the user performance error score. The higher the UPE score, the worse the performance: the goal of the respondent is to match the simulated values as closely as possible.

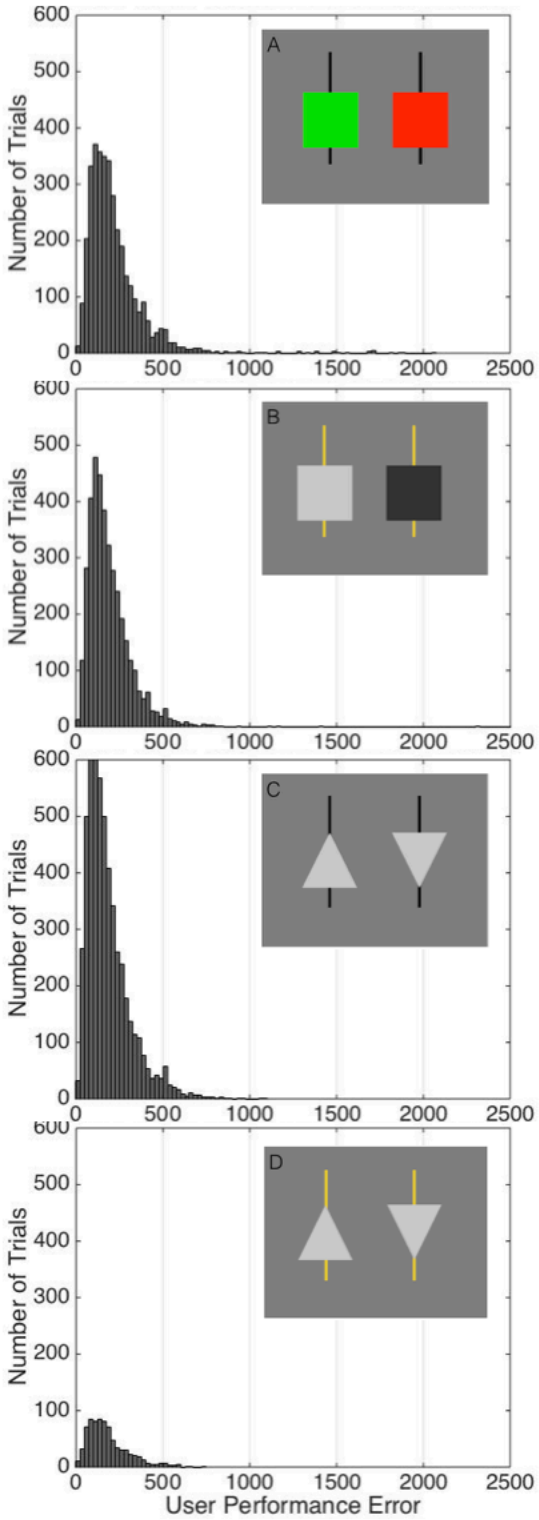


Figure 2.7. Histograms displaying the distributions of User Performance Error for each trial in the experiment by condition

The total User Performance Error is sum of the error associated with each of the four dimensions (open, close, high and low price) on each trial.

The average total UPE for all trials from participants in Condition A was 219.46 (standard error of the mean +/- 2.83, median = 180.49), for Condition B was 191.35 (+/- 2.20, median = 161.08), for Condition C was 181.95 (+/- 1.83, median = 149.01), and for Condition D was 182.4857 (+/- 4.17, median = 156.15) suggesting that Condition A supported less accurate decision making than Conditions B through D. A formal test of this hypothesis was conducted by using a hierarchical linear model. Note that the distributions of UPE were positively skewed (Figure 2.8). To aid in model fitting, we transformed the UPE in analysis and instead used the natural logarithm of UPE.

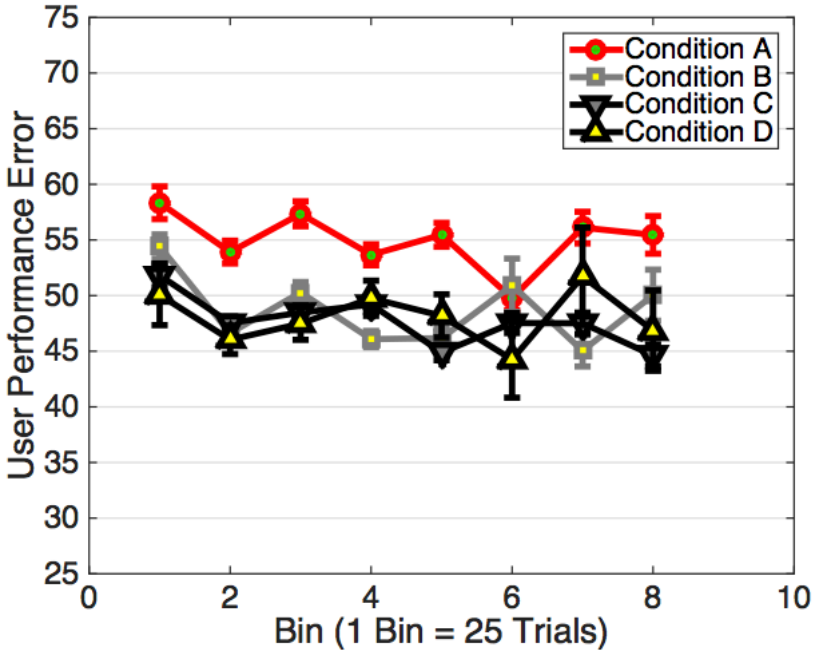


Figure 2.8. The average user performance error associated with each of the four dimensions over time

One bin is 25 trials, so the marker reflects the average of four dimensions, *n* participants, over 25 trials. Error bars reflect standard error of the mean. Note that this plot reflects the average error associated with each of the dimensions (open, close, high and low) which may differ from the overall UPE tested in the LME model.

A linear mixed model was constructed to analyze differences between groups after accounting for individual variation, and to reflect the hierarchical nature of the study design: response error is nested within trial, which is nested within subject. The linear mixed effects model controls for variance associated with the number of trials for each participant, too, by having Trial as a predictor. The R package lme4 was used to build the model and fit the data (Bates *et al.*, 2015). Because the red/green candlestick plot is

the standard version of a candlestick plot, it was the reference category in the analysis model for this study. The other conditions were analyzed relative to Condition A.

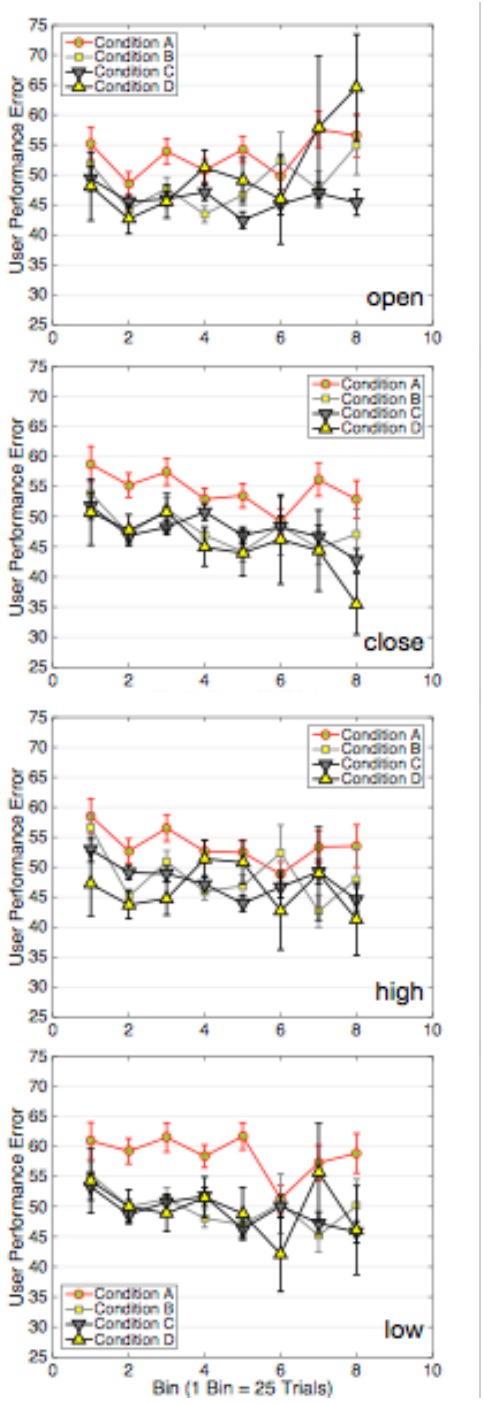


Figure 2.9. The average user performance error associated with each of the four dimensions over time

One bin is 25 trials, so the marker reflects the average of four dimensions, *n* participants, over 25 trials. Error bars reflect standard error of the mean.

In the hierarchical linear model, the unique subject identifier was included as a random effect, while the fixed effects were the trial number and the condition. Table 2.3 shows the estimated coefficients' *p* values estimated through normal approximation (Barr *et al.*, 2013; Mirman, 2014). Since Condition A is the reference category, coefficients were in relation to the red/green candle body Condition. Negative values, then, indicate a lower error rate relative to Condition A. Of particular interest was the significantly lower error value associated with Condition C suggesting that the triangular candle bodies support better user performance.

Table 2.3. Coefficient estimates from linear mixed effect model predicting log(UPE) Experiment Three over all dimensions

	Coefficient Estimate (β)	Standard Error	*T*	*p*
Intercept	5.1632	0.0411	125.5133	0.0000
Trial	-0.0002	0.0003	-0.6498	0.5158
Condition B	-0.0376	0.0549	-0.6859	0.4928
Condition C	-0.1532	0.0567	-2.7045	0.0068
Condition D	-0.1824	0.0955	-1.9093	0.0562
Trial x Condition B	-0.0006	0.0005	-1.2289	0.2191
Trial x Condition C	-0.0004	0.0005	-0.8717	0.3834
Trial x Condition D	0.0012	0.0009	1.3908	0.1643

Post-hoc comparisons were tested to discern differences between conditions (while also including the reference Category A). While Condition C clearly had an advantage over the standard red/green candlestick plot (Condition A) in supporting better performance ($T = 3.013$, $df = 43.26$, $p = 0.004$), it had a more modest advantage over the light/dark grey candle bodies (Condition B). The advantage exists for Condition C relative to Condition B (T=2.49, df=60.26 p=0.016), in that UPE in Condition B was typically worse than Condition C. Condition D shows no difference compared to the other conditions (Ts > 1.946, ps > 0.058). Of Conditions A, B and C, the best way to represent financial data in a candlestick plot for accurate forecasting is with the new triangle candle body (Condition C).

Responses for the low price show a particular advantage in performance for Condition C relative to Condition A (T = - 3.495, df = 44.46, p = 0.001). Participants also

made fewer errors to the close dimension in Condition C compared to Condition A ($T = -2.735$, $df = 58.292$, $p = 0.008$), but performance was not statistically distinguished between the two conditions on open or high price dimensions ($Ts < 1.85$, $ps > 0.073$).

2.3.5. Experiment Three: Discussion

The first prediction — that the sensory symbol (the triangle) would support better decision making — was supported overall (Figure 2.9). User Performance in Condition C was consistently better than participants using the standard candlestick plot (Condition A). User Performance was better also in Condition C than in Condition B, but to a lesser extent.

We were deliberate in naming the difference between Condition C and Conditions A and B after Ware's distinction between "sensory" and "arbitrary" symbols (2004). He describes sensory symbols as those that correspond to earlier visual processing, in contrast to arbitrary symbols that require learning to use and understand. The arrow (in Condition C) is a powerful cue. Indeed, even when an arrow is not predictive of an event, and even when and the observer knows it doesn't help, the arrow draws attention toward the direction to which its lines converge (Kingstone, Smilek, Ristic, Friesen & Eastwood, 2003). Because the arrow makes the directionality in the candlestick plot easier to intuit, it eliminate one unknown element of reading a candlestick plot. If the arrow is pointing up, then the stock closed higher that day. In the rectangle candle bodies, the user must use the colour to choose the side of the body that the close price is represented by. While a simple enough step, it is still an additional stage of processing. It would be expected, then, that the triangle candle body helps people make better forecasting estimates by reducing unnecessary cognitive load.

To our knowledge, no one has used candlestick plots with triangular candle bodies in actual trading practice. Evidence from this study suggests that it might be time for that to change. At very least, it's worth more investigation. The UPE was measured in pixels, because participants were essentially drawing in the glyph that they thought completed the time series. In this study, 48 pixels corresponds to one dollar. Participants were typically off by 190 pixels (see Figure 1.5), and across 200 trials that corresponds to a cumulative error of $791.67.

While the actual use of a candlestick plot does not require exact prediction of dimension values, it is concerning that the standard candlestick plot design negatively impacts participants' ability to predict dimension changes over time in the standard case (Condition A) relative to simple improvements (Conditions B and C). Earlier evidence exploring time series forecasting suggests that forecasting performance is better when it's based on tables than line graphs (Lawrence, Edmunson & O'Connor, 1985); and so it may be that there is a more fundamental issue in graphing time series that goes beyond any peculiarities of the candlestick plot and impacts graphical representations of time series data more generally. Even if time series cannot be represented well graphically, there are fascinating idiosyncrasies of the candlestick plot identified in these results.

Surprisingly, the Condition C advantage was clearest for the low price. Errors associated with the low price were higher for Condition A. This was a surprising finding in two respects: a) low price wasn't expected to differ much from the error associated with the high price but it did, and b) the low price was represented by the same symbol (a black stick; Figure 1.5) in Conditions A and C, and so a performance advantage (while expected in the aggregate) was not anticipated in this dimension.

There are some potential explanations for this result; we developed these post-hoc and further work is required to better discern which (if any) of the explanations are the most accurate. One is that the salient candle body relative to the shadow in Condition A draws attention to the extent that the lower shadow is not properly processed. Salience is known to predict eye movements in natural environments (Parkhurst, Law & Neiber, 2002) and salient components of the candlestick display are similarly expected to draw attention. If salience were the only source of difference, though, we would expect the performance in estimating both the high price and the low price to be equally poor. That is not the case. There is an imbalance between high and low error: especially for Condition A. This suggests the increased error in the low price for Condition A is not simply a question of salience.

Alternatively, it may be that the arbitrary representation of information requires an extra processing step in working memory. Recall that participants were selecting and adjusting the glyph they select without the stimulus on screen, and so their adjustments are guided by their memory of the stimulus rather than the stimulus itself. It's been shown that there is a performance cost when graphs are shown sequentially compared

to when they are presented side-by-side (Ware, 2004) which is attributable to having to rely on working memory instead of the veridical perception of the graph. Working memory is a brief, low-capacity system that can typically store only 3-7 “chunks” or groups of units (Miller, 1956; Ericsson *et al.*, 1980; Vogel *et al.*, 2001). Kosslyn (2006) acknowledged this as an important concept for creating effective graphics when he introduced the “Principle of Capacity Limitations”. The “Principle of Capacity Limitations” suggests that effective graphics must be built with the user in mind, and that humans are beings with finite working memories who cannot maintain too much information.

Working memory limitations can help explain the imbalance between performance in the high and low dimensions if each dimension (open, close, high, low, and colour) is a “chunk” and if the dimensions are consistently recalled using a strategy that leaves the low price last. Said differently, one possible explanation for the worse performance in the low dimension is that working memory cannot hold all of the dimensions in time to adjust the glyph to accurately reflect the low price. If capacity limits of working memory were the only source of difference, though, we would expect performance to be worse for the low price in all conditions. However, the imbalance between high and low price estimation performance was worse for Condition A. Thus, the increased imbalance between high and low price estimation performance cannot be attributed only to working memory capacity limits.

The particularly bad UPE for the low dimension in Condition A is probably due to interactions between higher-level processing (working memory) and lower-level processing (salience). In making categorical decisions about glyphs, participants who use a problem-solving strategy in which they prioritize the general picture outperform those who use a strategy when they focus on details (Greensmith, 2016). If that generalizes to the continuous case, as in the present study, it may be that Conditions B and C invoke a more general perception than Condition A does. If Conditions B and C support a more integrated perception, then there less of a load on working memory. Working memory is taxed more by a detailed dimension-level memory test in Condition A than the general object-level memory in Condition C because there are four dimensions for each object. Insofar as determining the relative influence of low-level salience and higher level memory and problem solving factors, scientifically investigating human perception of candlestick plots is quite like investigating visual categorization (Hammer, 2015; Hammer et al., 2012; McColeman & Blair, 2015), speech perception

(Zekveld, Heslenfeld, Festen & Schoonhoven, 2006) and visual search (Theeuwes, 2010; van Zoest, Donk & Theeuwes, 2004; Gaspar & McDonald, 2014; van Zoest & Donk, 2007): among the biggest questions in cognitive science are the identification, integration and dissociation of top-down and bottom-up processes underlying phenomena of interest.

2.3.6. Chapter Two: Discussion

Results from this chapter show that visual salience does not globally impact information extraction, but it does impact performance during forecasting. In Experiments One and Two (information extraction), participants were asked to extract the value for a single dimension based upon a previously presented candlestick glyph. Globally, there were no differences between conditions, but there were interesting differences between dimensions: the dimensions associated with the candle bodies (open, close) were more difficult for participants to accurately recall than the dimensions associated with the extending candle shadows (high, low) as long as there were no grid lines. The relatively poor recall for the candle bodies in Experiment One might offer some lessons generally for spatial memory.

Colour has been shown to be more discriminable than shape (Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1991), and so it could be expected that candlestick values represented by colour should be easier for participants to recall. The results show that this was not the case, and there are some valuable considerations as to why colour is associated with poorer spatial recall in Experiment Three. Earlier work has shown selective attention increases activation to neural circuitry specializing in processing that feature (e.g. colour; Corbetta *et al.* 1991). Given that the open/close performance is worst in Condition A, in which the candle body is more perceptually distinct from the candle shadows, the glyph may be eliciting a more dissociated object percept than the other conditions. That is, the candle body and the candle shadows are more different from each other in Condition A, and so each engage their own distinct processing. Conceptually, this is in line with the theory of feature binding posited by Wheeler and Triesman (2002) wherein features compete for limited working memory within their own property (colour, shape) but not between. The same features that compete for working memory are better integrated (Triesman & Gelade, 1980), and so the percept of an

object comprised of those integrated features will be more cohesive than the percept of an object comprised of distinct features.

The candlestick plot adds an additional process on top of object perception, though: while the percept of the glyph is important, localizing its position relative to the graph axis is also critical for making a response about the dimension values. Colour and shape are both properties of the candlestick glyph (in Condition A), but the shadows and candle bodies' position in space is what actually provides the information to the observer. Evidence from Experiment Three suggests the colour may be maladaptive for spatial localization in visual working memory, or at least less helpful than simple line position in space (consistent with Cleveland & McGill, 1984). Further, considering visual working memory capacity as defined by "slots" or discrete opportunities to store information is fundamentally problematic for continuous recall in space, where error is in degrees (as in Experiments One and Two) rather than binary. Proposals to conceptualize working memory as a topographical map instead of a slot-based approach (Franconeri, Alvarez & Cavenagh, 2013; Barnes *et al,* under revision) may be better suited for the next generation of visual working memory problems. Topographical models also allow for additional representation of nuance in the environment itself, and can model problems of interference.

Even in the simple glyph-based experiment, data visualization is suited to productively challenging existing accounts of visual cognition. Parallels between data visualization and cognitive science have been drawn before (Pinker, 1990; Green; 1998; Rensink, 2014) and for good reason: the relationship between the observer's beliefs and the data that can update those beliefs is dictated in large part by the manner in which those data are presented.

Chapter 3. Irrelevant dimensions in forecasting

To study data visualization is to study the complex relationship between organizing data for easier understanding (or abstraction, on the generation end, Figure 1.2) and distilling information from graphic displays (extraction, on the receptive end). Generating a graph means simplifying complex datasets in meaningful ways to best communicate the essence of the data to an observer. Perceiving a graph means extracting meaningful information from abstract representation. Generation and reception of data visualizations shares many qualities with language (Ware, 2004): there are some arbitrary signs and symbols, and some rules that are culturally normed for the the graph to effectively communicate information. For example, in Western culture, time is typically represented on the X axis of a graph, where rightmost points represent later dates. Similarly, height on the Y axis typically indicates increasing magnitude.

Like understanding spoken or written language, to extract information from a data visualization, the observer must attend to the appropriate elements of the stimulus and associate them with meaningful regularities in conceptual space. Researchers in categorization have been studying how this is done for decades. Recent advances in computer vision are inspired by the cognitive and neural architecture of humans (LeCun, Bottou, Bengio & Haffner, 1998), in part because we are capable of extracting regularities from complex input. Hand-written numbers manifest in diverse ways, but for most adults, reading any one of the MNIST data set and assigning a meaning to the input is trivial: such ability to assign deeper conceptual meaning to a massive set of inconsistent input is foundational to communication and intelligence.

Some of the capacity to identify regularities in the environment arises from the ability to disentangle relevant from irrelevant stimulus features. To build on the MNIST number reading example, symmetry is not a critical feature of “0”, but perhaps containing a single loop is. While learning categories in lab-based tasks, it’s known that participants eventually stop sampling the part of the visual display that contains irrelevant information as they get better at making category decisions (Blair, Watson & Meier, 2009; Blair, Watson, Walshe, & Maj, 2009; Rehder & Hoffman, 2005). Earlier in learning, participants gather more information than is necessary to make a category choice, and learn over time that some features are not informative and that it’s more efficient to ignore them.

One proposal of this thesis is that data visualizations and categorization share the same general process: extracting lower-level regularities from the environment to inform higher level decisions. Building on parallels between categorization and data visualization, Experiment Four investigates the impact of an irrelevant dimension in a pattern completion task. Practically, this experiment is an example of how data visualization can act as a traditional stimulus to query basic psychology. While earlier work about learned attentional optimization has used category features, this work uses visualized data dimensions, but the purpose is the same: to understand how information from the environment informs human cognition.

Efficiently deploying cognitive resources is adaptive, given that the cognitive system is limited and only so much information can be processed at once. Indeed, a massive part of cognitive science research is dedicated to the study of selective attention. The ability to selectively suppress distractors is correlated with working memory (Gaspar, Christie, Prime, Jolicœr, & McDonald, 2016), and the ability to integrate meaningful features is thought to be fundamental for object perception (Triesman & Gelade, 1980). Categorization and data visualization both require feature integration to make meaningful decisions based upon visual input.

It was expected that adding a noisy dimension to candlestick plots would (trivially) increase error associated with that dimension because there is no trend to discern. If the performance on the remaining three dimensions is also worse than for candlestick plots without noisy dimensions (in Experiment Three), that suggests that the perception of the data are integrated at least to some degree. Objectively speaking, participants should integrate information while viewing a candlestick plot to most effectively use the information they convey: the values of the open, close, high and low dimensions are related (except for the noisy dimension in the present experiment).

Information integration is one type of categorization process often presented in contrast with rule-based categorization (Maddox & Ashby, 2004). Information-integration and rule-based categorization are dissociable in terms of neural mechanisms (Nomura, Maddox, Filoteo, Ing, Gitelman, Parrish, Mesulam & Reber, 2007), and behaviourally, information-integration categories are harder for participants to learn (Ashby & Maddox, 2010; McColeman *et al*., 2014) when controlling for the number of dimensions to integrate. One particularly appealing quality of using data visualizations as stimuli to explore

cognition is that more complex, realistic problem spaces invite more complex, realistic tests of cognitive theory. In candlestick plots, rule-based and information-integration processes are both optimal at different levels. In extracting information from a single glyph, participants can rely on rule-based strategies to learn how the different features of the glyph correspond to values of the stock. Experiments One and Two isolate single-glyph learning. In Experiments Three and Four, participants can first learn the rule (how to read a single glyph) and then integrate the output of the rule across multiple icons to generate a pattern percept. Given prior work indicating that dual-task environments impair rule-based but not information-integration category learning (Waldron & Ashby, 2001; Zeithamova & Maddox, 2006), and the procedural nature of information integration (Ashby, Alfonso-Reese, Turken & Waldron, 1998), information integration is the more effective process to rely on for complicated pattern perception.

In addition to testing the impact of an irrelevant dimension in making forecasting judgements from data visualizations, Experiment Four also offers another way to test the impact of salience. While the experiments discussed thus far test salience as defined by relative colour changes, another consideration is the “perceptual weight” of an icon — its size relative to the whole scene (Hoffman & Singh, 1997). Because the open/close dimensions are represented using more pixels compared to the high/low dimension, it was expected that noise applied to the open or close dimension will more negatively impact performance than noise added to the high or low dimension. While data visualization designers already tend to exclude irrelevant dimensions/information in their plots, this is the first known empirical investigation into the effect that including uninformative values has on reading the visualization as a whole.

3.1. Method

The data that were graphed to create the stimuli were generated by a vector autoregression model, as was the case for Experiments One through Three. The output of the vector autoregression model was scaled to be between 0 and 1 for all experiments. In Experiment Four, to generate a “noisy channel”, one of the four dimensions (open, close, high or low) was overwritten: what was originally the output of the vector autoregression simulation was replaced with a random set of values selected from a uniform distribution between 0 and 1 for the noisy channel. The remaining three

dimensions were determined by the the output of the vector autoregression simulation.

Table 3.1. Summary of experimental conditions in Experiment Four

Candle Condition (CC)	Noisy Channel (NC)
red/green (Figure 1.5A)	open
red/green	close
red/green	high
red/green	low
up/down (Figure 1.5D)	open
up/down	close
up/down	high
up/down	low

To avoid the proliferation of conditions and retain sufficient experimental power, this experiment is constrained to two candle-types: the standard red/green candle (Figure 1.5A) and the updated up/down candle (Figure 1.5D). The noisy channel is consistent within subjects, and so there are eight between subjects conditions in this study (Table 3.1).

3.1.1. Participants

Fifty-eight participants were recruited from the research participation system at Simon Fraser University and received partial course credit in exchange for their time in this study. Participants were randomly assigned to one of two possible candle conditions (Figure 1.5A or Figure 1.5D) and one of four noisy dimension conditions. There were 14 assigned to noisy open, 16 assigned to noisy close, 14 assigned to noisy high and 14 assigned to noisy low. As with earlier experiments, some general demographic information was collected.

3.1.2. Apparatus and Materials

The experiment was the same as Experiment Three, except for the noisy dimension. The stimuli were constructed with values determined by vector autoregression models, but one of the four dimensions (open, close, high or low) was replaced with random values to make it a noisy channel. The graphs had the same visual properties as Experiment Three.

3.1.3. Procedure

This experiment, like the others, was double-blind. The experiment was 55 minutes, or 200 trials long (whichever came first terminated the program). Figure 2.6 displays the trial procedure: a participant sees a fixation cross, then a stimulus (examples in Figure 3.1), and then makes adjustments to the possible glyphs to complete the pattern that they were shown. Upon completion of their response, participants receive feedback based on how near their adjusted glyph was to the simulated values. Like in earlier experiments, participants get a break every 20 trials, during which they are reminded to try to gather as many points as they can.

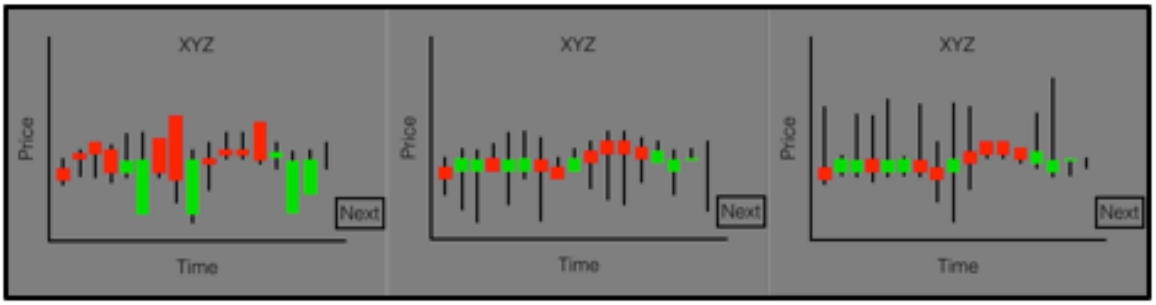


Figure 3.1. A schematic of how noise applied to different dimensions would appear on the same dataset.
The leftmost panel shows noise applied to either the open/close dimension, the centre shows noise applied to the low dimension and the right shows noise applied to the high dimension.

The leftmost panel shows noise applied to either the open/close dimension, the centre shows noise applied to the low dimension and the right shows noise applied to the high dimension.

3.2. Experiment Four: Results

The average total UPE for all trials from participants in Condition A was 178.746 and for Condition D was 171.453. An independent samples T-test did not reveal any reliable advantage of Condition D (T = 1.051, df=60.65, p=.30). The experiment design is complex, however, and failure to discern a difference between groups may be attributed to the error associated with noise on particular dimensions.

A formal test of the full between subjects crossing was conducted by using a

hierarchical linear model. As in the earlier experiments, to aid in model fitting, UPE was transformed in analysis and the natural logarithm of UPE was used.

As in Experiment Three, the unique subject identifier was included as a random effect, and trial number and the conditions were fixed effects. Table 3.2 shows the estimated coefficients' *p* values estimated through normal approximation (Barr *et al.*, 2013; Mirman, 2014). In this, and all linear mixed effects models herein, Condition A is the reference category. Negative values are interpreted as an improvement relative to Condition A, given that the UPE — the error — is lower.4

There was no overall advantage of Condition D. The biggest difference in user performance error was when the noisy channel was the "close" dimension (estimated $\beta = -0.187$, $T = -3.53$, $p < 0.001$): participants were better at dealing with unpredictable close values then unpredictable open, high or low values. This was contrary to expectations that pixel-heavy or visually larger dimensions (open and close) would harm performance more than physically smaller dimensions (high and low) when they were noisy.

To probe the impact of each of the dimensions as manifested in each of conditions, a set of T-tests was conducted based on the mean user performance errors (UPE) for each participant. In no case was any condition seen to be differently impacted by the type of candle icon used. Notice, however, that the error associated with noisy Close dimension was lower than the error associated with other noisy dimensions.

Finally, to test whether the presence of a noisy dimension negatively impacted performance, the error associated with Conditions A and D in Experiment Four was compared to the overall error associated with Condition A and D in Experiment Three. Recall, Experiment Three and Four are the same except for the presence of a noisy dimension, and so the isolated difference between the two is the presence of a noisy dimension. A T-test revealed an effect, but contrary to the expected direction: in Condition A, the presence of a noisy dimension (Experiment Four) was associated with better performance ($T = 2.607$, $p = 0.012$). Similarly, error was lower for Condition D in Experiment Four with the presence of a noisy dimension ($T = 2.227$, $p = 0.030$) While it may be that the performance of the participants in these Conditions were just particularly bad in Experiment Three, it is also possible that having fewer meaningful dimensions to integrate in order to make a forecasting judgement was easier for participants.

Table 3.2. Coefficient estimates from linear mixed effect model predicting log(UPE) Experiment Four considering Candle Condition and Noised Out Channels

	Coefficient Estimate (β)	Standard Error	*T*	*p*
(Intercept)	5.1153	0.0461	111.0259	0.0000
TrialID	-0.0002	0.0003	-0.7652	0.4442
Condition D	-0.0371	0.0473	-0.7834	0.4334
Noisy Open	-0.0534	0.0532	-1.0038	0.3155
Noisy Close	-0.1865	0.0529	-3.5255	0.0004
Noisy High	-0.1050	0.0574	-1.8281	0.0675
TrialID x Condition D	-0.0002	0.0004	-0.4674	0.6402

In Experiment Four, there is an advantage for people assigned to a noisy Close dimension in reducing performance errors. To extend the earlier test, and to determine the extent to which the advantage of the noisy close dimension contributed to the lower performance error in Experiment Four relative to Experiment Three, a second T-test was conducted, but without data from participants assigned to the noisy Close condition. The difference between Experiments Three and Four was no longer significant (T = 1.9998, p = .0514), suggesting that much of the performance advantage came from the participants who were making forecasting judgements with a noisy close dimension.

Table 3.3. Results of noisy dimension differences between Conditions

Noisy Channel	Mean UPE: Condition A	Mean UPE: Condition D	*T*	*p*
Open	183.9886	169.9990	1.041	0.3157
Close	165.3392	153.1433	1.0428	0.3136
High	184.7109	172.6531	1.3241	0.2104
Low	185.7492	199.0388	-0.8716	0.4163

3.3. Discussion

There were two primary research questions addressed in this experiment. One was to test the role of the size of objects — one type of salience — in affecting participant performance. The second was to test the impact of irrelevant dimensions in forecasting performance.

The open and close price are represented using larger icon elements than the high and the low price. As such, it was expected that having unpredictive open and close values

would more negatively impact performance, based on the assumption that the visual salience of larger dimensions is higher and would more effectively get encoded into working memory (Fine & Minnery, 2009) to then influence decisions. Contrary to this expectation, the lowest error (the best performance) was observed in participants for whom the close value of the candlestick plot was replaced with noise. The more salient dimension, when made irrelevant to the forecasting judgement, apparently helped participants make better decisions.

This sort of complex and surprising connection between salience, attention and decision-making has been noted before. Attentional optimization (the time spent on relevant versus irrelevant features) was found to be better in an early study with salient distractors than with less salient distractors (McColeman, 2012). In a second category learning study, saccade velocities were shown to be slower to task-relevant parts of the visual environment. Saccades to irrelevant features were faster and rarer. However, when those irrelevant features were salient, saccade velocities to them were slower, and approximated the saccade velocities made to relevant features in that same task (McColeman & Blair, 2015). One possible explanation for those results is a sort of nested exclusionary criteria for exogenously driven attention: first, low-level visual properties can be filtered (Broadbent, 1958), then task-defined, irrelevant items can be eliminated from contention (Shepard, Hovland & Jenkins, 1961). Exogenous attention draws faster saccades (Xu-Wilson, Zee & Shadmehr, 2009), and is enacted earlier in visual processing (van Zoest, Hunt & Kingstone, 2010). After filtering for exogenous attention-capturing features (Theeuwes, 1991, 1992), what remains to be executed are deliberate, volitional eye movements, said to be slower than their exogenous counterparts (Schütz, Trommershäuser, & Gegenfurtner, 2012).

For the previous experiments in this thesis, all dimensions were task-relevant. However, in this Experiment Four, one dimension is uninformative, just as the irrelevant items were in the category learning study. When a dimension is made uninformative, performance is a bit better. This is only statistically supported, however, when the people assigned to the noisy close dimensions are included in the analysis. The best performance, then, is when the most salient dimension is the least helpful. While at first difficult to consider, it is consistent with the findings in McColeman and Blair (2015), where salience did not seem to distract at the level of overt attention (as measured by

oculomotor activity). Salient uninformative features, when sufficiently different than the relevant task set, can actually be helpful for effectively deploying attention to the features that do matter.

Chapter 4. The Influence of Relative Glyph Spacing on Interpreting Data Values

Earlier work explored the accuracy of extracting information from plots, and found that position (or, more intuitively, “height”) is the easiest property to extract from a graph (Cleveland & McGill, 1984; Simkin & Hastie, 1987). There’s good evidence that this is the case, and little reason to doubt that spacing the glyphs differently would impact the perception of the pattern, such that larger spacing would appear noisier. However the candlesticks are more complicated than the average graph. Little is known about the perception of multivariate time series data visualization generally, let alone the perception of candlestick plots under different glyph spacing conditions.

The spatial representation of information is clearly critical in the context of data visualization, given that space represents some property of the data. In most visualizations, more distance between two data points corresponds to a greater difference between the values they represent. However, spatial processing is also a critical component of cognition generally. Without knowing the spatial layout of information, for example, goal-directed visuospatial attention would be impossible. Spatial representations can be observed in the form of categories or coordinates (Kosslyn, 1987), where categorical spatial representation is more abstract (e.g. the table is *below* the lamp) and coordinate representation is more specific (e.g. the pen is *six centimetres* from the computer mouse).

In earlier work, the differences between categorical and coordinate spatial representations were queried using a simple stimulus wherein the relationship between a dot and a central cross was shown, and the participant was asked to respond with whether the dot was in a particular quadrant of the area surrounding the cross (the categorical spatial representation task) or how far away from the centre of the cross the dot was (the coordinate representation task). The results suggest a double dissociation between the categorical (Kosslyn, 1987) and co-ordinate spatial representation (van der Ham & Borst, 2011) given that there were more errors for coordinate spatial representation performance when stimuli were presented to the left than the right, and more errors for categorical spatial representation performance when stimuli were presented to the right than the left. Those findings were in keeping with theories that there is some hemispheric specialization

at play for spatial representation: the left hemisphere is thought to specialize in categorical spatial processing and the right is thought to specialize in coordinate spatial processing (Tranel & Kemmerer, 2004).

In the current chapter, the stimuli are slightly more complex, but build gradually on the problem of spatial representation. Experiment Five tests the importance of size relative to the graph axis space in supporting precise judgements about particular stimulus values. Experiment Five is similar to Experiment One, but with the added manipulation of glyph spacing, which (with just a single glyph) manifests as a manipulation of size relative to the graph axis. In Experiment Six, the full series is presented, and the glyph spacing is manipulated to test the impact of icon size in making forecasting judgements. Because the tasks differ, the more effective strategy in Experiment Five is to employ coordinate spatial representation — to judge the distance from the X axis. In Experiment Six, however, the more efficient strategy is to employ categorical spatial relationships, and to make judgements about the relative spacing of the final icon compared to the original series.

In Experiment Five, it is expected that greater spread will correspond to more accurate estimates, while in Experiment Six it is expected that a lower spread will correspond to more accurate forecasting judgements. Because Experiment Five is a simple information extraction task, more space representing the same variation in graphed values allows more room for error. Said differently, erring by 25 pixels counts for more User Performance Error in the standard spread condition than it does in the high spread condition, since the increase of a pixel corresponds to the increase in relatively more value in the standard spread than the high spread. In Experiment Six, however, there's a trade-off between precision in estimating the value of a single glyph and perceiving a whole pattern. In Experiment Six, pattern perception is expected to be easier as the perceptual variation in the low spread condition is more supportive of perceiving an integrated series (via Gestalt principles of similarity and proximity) than the high spread condition which appears relatively erratic.

Additionally, the importance of the visual properties of the objects that are presented in space are manipulated (as with the previous experiments) to test the impact of visual representation on spatial encoding. The following experiments manipulate the organization of glyphs to explore the impact of spacing on the perception of graphed

multivariate data. Results from these two experiments can serve better design practice, but also can provide evidence for the manner by which the cognitive system supports pattern perception, and how it emerges from complex icon sets.

4.1. Experiment Five: The influence of glyph spacing on extracting information from candlestick icons

Stephen Few, a thought leader in the field of data visualization, suggests that the difference between tables and graphs is that tables are best for communicating specific data points, and that graphs are best for communicating relationships between data points (2004). His sentiment is partially supported by earlier work wherein data presented in tables does offer the best precision (Carter, 1947; Vessey, 1994; in contrast, Spence & Lewandowsky, 1991) but at the cost of increased time and effort (Spence, 1990). In the context of the spatial processing theories discussed above, Few essentially says that graphs are good for categorical (propositional) representation, while tables are preferable for coordinate representation. The encoding and processing of spatial relationships is critically important to effectively understanding and navigating the environment and so lessons connecting visual manipulations to the efficacy of spatial encoding are informative generally.

Of particular interest is whether the relative size (Experiment Six) or absolute size of (Experiment Five) glyph icons impact participants' understanding of the values that they convey. The phenomenon of spatial perception has been separated into two dissociable processes: co-ordinate and categorical spatial relations (Kosslyn, Chabris, Marsolek & Koenig, 1992). While co-ordinate spatial representation is the representation of a point in space, categorical representation is the representation of items relative to each other. These concepts, of course, are critically important to consider for data visualization. Given that position in space so frequently communicates information about the value of a graphed dimension, knowing how the observer understands position is important for knowing how the observer understands data. Experiment Five, in presenting just a single glyph, is designed to elicit co-ordinate spatial representation whereas Experiment Six is designed to elicit categorical representation.

4.2. Experiment Five: Methods

114 participants were each randomly assigned to one of six conditions: Where Spread refers to the relative distance from the centre of the y-axis that the glyphs' elements were located, and the Candle Condition refers to the type of candlestick body (Figure 1.5).

4.2.1. Participants

Participants in this study received partial course credit for their participation through Simon Fraser University's research participation system.

4.2.2. Apparatus and Materials

In earlier experiments, the spacing was determined by multiplying the scaled output of the vector autoregression model by forty, so if the model output a vector of values [.2 .9 -.1 .01] for open, close, high and low price respectively, the icon would be [8, 36, -4, .4]. To plot the data, the values were interpreted as "distance from the centre of the Y axis in pixels". Since 36 was greater than 8, in this example, the closing price was higher than the opening price. The candle body represented that higher closing price (either upward triangle or green body depending on the condition). The bottom of the candle would have been 8 pixels above the centre of the Y axis, and the top of the candle would have been 36 pixels above the Y axis. If there is a negative value for the high and the low price, that shadow is plotted with 0 pixels, as negative distance is not possible.

The same strategy was used for the standard spread conditions in this experiment. The low spread was half as diffuse as the standard (the model output was multiplied by 20, rather than 40), and the high spread was twice as diffuse (multiply the model output by 80 to get the pixel values). The models that generate the data were the same between all three of the Spread conditions, so it is just the scale that the model output is multiplied by that differs between conditions. This experiment, like Experiment One, included a labelled Y axis to indicate the stock prices. Because the variable of interest here was the position of the glyph and not the value it represented, the range of values on the Y axis were scaled (Table 4.1) to correspond to the range of values in the scaling multiplier (20, 40, 80). Functionally, the Spread conditions were a matter of zooming out (low spread) or

zooming in (high spread) on the data.

Table 4.1. Summary of experimental conditions in Experiment Five

Candle Condition (CC)	Spread (S)	Multiplier on VAR output
red/green (Figure 4A)	low (0.5)	20
red/green	standard (1)	40
red/green	high (2)	80
up/down (Figure 4D)	low (0.5)	20
up/down	standard (1)	40
up/down	high (2)	80

4.2.3. Procedure

For each participant, Experiment Five was terminated after 25 minutes or 40 trials. Figure 2.6 displays the trial procedure. Like in Experiment One, each trial begins with fixation cross, then the presentation of stimulus. Participants view the stimulus for as long as they'd like to before clicking "next" to advance. After they click next, they are asked to report the value of one of four dimensions: open, close, high or low price of the stock indicated by the stimulus glyph. After entering their response participants are awarded points based on how close they were to the actual presented value. The experiment was the same as Experiment One, Conditions A and D, but with the addition of Spread as a between-subjects factor. It was expected that the high spread condition would support more precise responses (McGill & Cleveland, 1984) because the visual distance between values was larger than in the lower spread conditions (Wever & Zener, 1928) and thereby easier to observe.

4.3. Experiment Five: Results

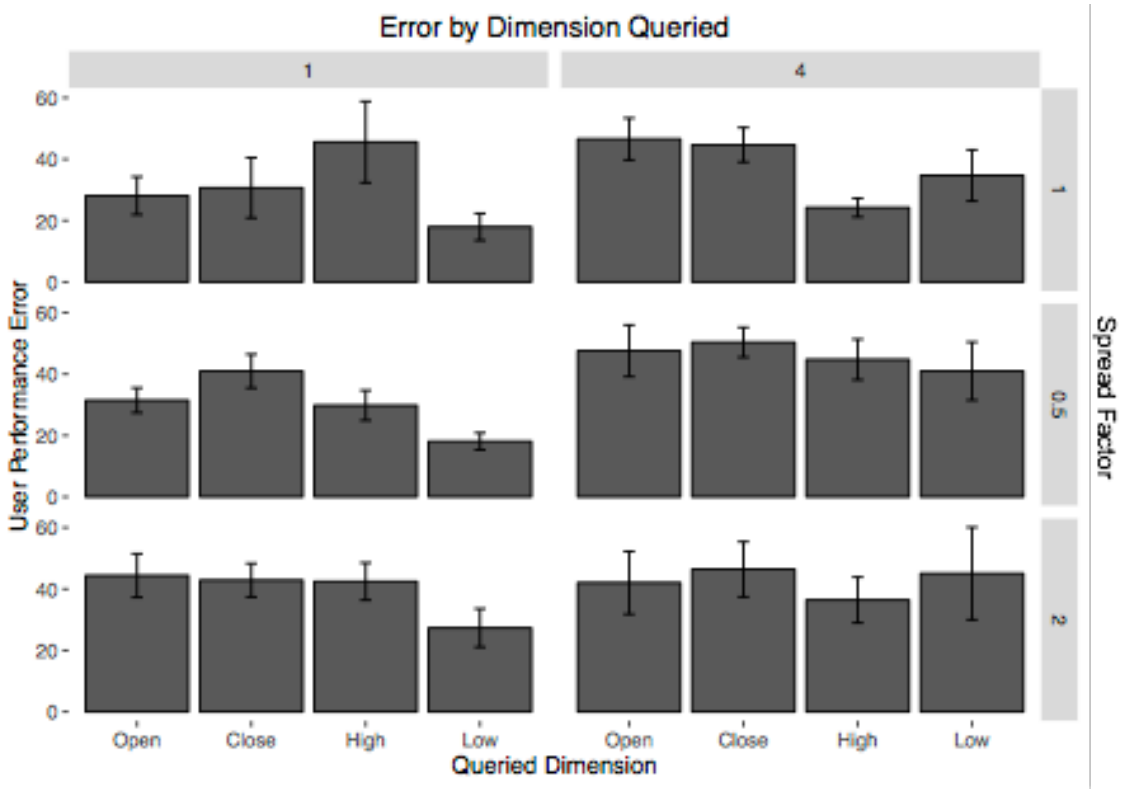


Figure 4.1. User performance error by queried dimension and spread and Condition

The mean user performance is indicated by the bar, the standard error is represented by the error bars.

The analyses were the same as the analyses from Experiment One, but with the addition of Spread as a factor and the removal of dimension to retain power. A linear mixed effects model was constructed with Trial, Candle Condition, and Spread and their interactions to predict the user performance error. As in Experiment One, the user performance error is the difference between the participants' response (as it converts to the spatial representation) and the actual stimulus value for the queried dimension. The linear mixed effects model estimates are displayed in Table 4.2.

Table 4.2. Coefficient estimates from linear mixed effect model predicting log(UPE) Experiment Five considering Candle Condition and Spread

	Coefficient Estimate (β)	Standard Error	*T*	*p*
(Intercept)	2.7444	0.3700	7.4175	0.0000
TrialID	-0.0108	0.0131	-0.8223	0.4109
Condition D	-0.0344	0.4074	-0.0844	0.9327

Low Spread	-0.2031	0.4144	-0.4902	0.6240
High Spread	0.1880	0.4080	0.4606	0.6451
TrialID x Condition D	-0.0003	0.0144	-0.0187	0.9851
TrialID x Low Spread	0.0030	0.0148	0.2011	0.8406
TrialID x High Spread	-0.0060	0.0146	-0.4121	0.6803
Condition D x Low Spread	0.6021	0.4835	1.2452	0.2131
Condition D x High Spread	-0.4852	0.4936	-0.9830	0.3256
TrialID x Condition D x Low Spread	-0.0159	0.0174	-0.9153	0.3600
TrialID x Condition D x High Spread	0.0146	0.0181	0.8074	0.4194

There was no effect of Condition (β = -.034, T = -.084, p = .933) or Spread (magnitudes of βs ≤ .203, Ts ≤ .490, ps > .624) and insufficient evidence to suggest that participants' responses improved over time (β = -.011, T = -.822, p = .411). Two-way interactions were not shown to be significant predictors of user performance error (magnitude of βs ≤ .602, Ts ≤ 1.245, ps>0.213). Overall, the manipulations were ineffective at influencing user performance in the simple information extraction task.

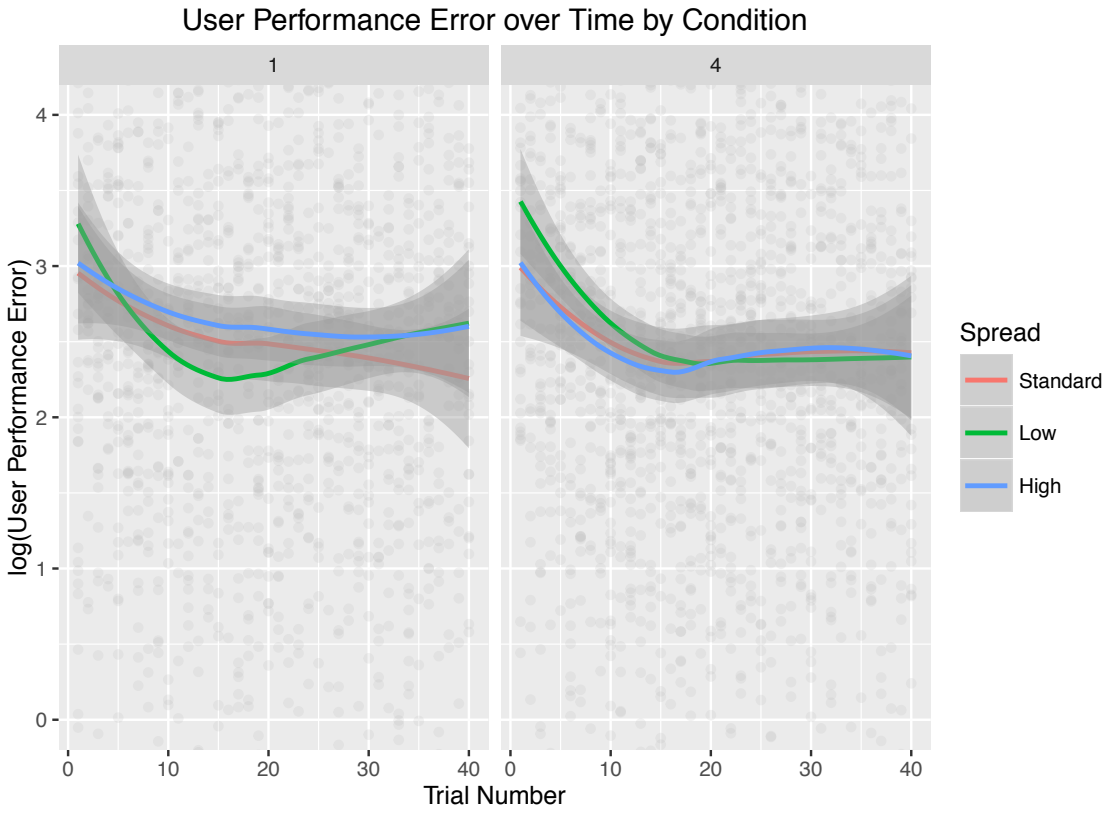


Figure 4.2. User performance error for Experiment Five
Each panel represents one Condition. The lines represent LOESS curves for each of the spread conditions, and the grey shadows represent the standard error of each. There is no significant difference between condition on user performance error.

When examining user performance error by queried dimension, there appears to be a slight advantage for Condition A. Notice, however, that the Candle D UPE is more consistent across the multiple types of spread factors — regardless of whether the candlestick plot is zoomed out (low spread) or zoomed in (high spread), performance is relatively consistent. Overall, in the context of simple information extraction from single candlestick icons, there appears to be no real advantage of larger (higher spread) icons (Figure 4.2). If the observer of the graphs desires a graph that encourages attention to be more consistently distributed across multiple dimensions, evidence from this experiment suggests that zooming in on the graph (high Spread) or using a triangle candle body may support those goals.

4.4. Experiment Five: Discussion

In the context of spatial representation (Ruotolo, van Der Ham, Iachini, & Postma, 2011), this experiment tests the influence of dimension position relative to the graph axes on decision precision. User performance error did not vary with Spread, suggesting that manipulations of coordinate spatial representation do not influence the ability of participants to link the position of an icon to an abstract value such as a dollar amount. Additionally, there was no advantage of the Close dimension in the triangle candle body condition.

The cuing effect of arrow-like shapes (Ristic, Friesen & Kingstone, 2002) does not impact performance. Experiment Five tests the importance of size relative to the graph axis space in supporting precise judgements about particular stimulus values. This was true too for Experiment One (also where participants were making point estimates about the value represented by a single dimension from a single candlestick icon). At some point in processing, between the low-level visual input and the high level estimate about the positional representation of a point in space relative to an abstract concept, the attentional advantage of arrows is lost. That is, while in low level vision, we know that arrows direct attention, that attention does not result in differences in cue utilization in decision making.

Effectively ignoring attention-grabbing, but irrelevant dimensions has been explored above in Experiment Four and in earlier work (McColeman & Blair, 2015). In Experiments One and Five, however, the arrow points to a relevant dimension: the closing value. In these experiments, ignoring the closing value is maladaptive, and given that the performance in the close dimension is also no worse than the other dimensions, participants do appear to consider the close value when viewing the stimulus. Rather, there must be a point at which the attention advantage of the arrow pointing to Close is ameliorated by other effects. It is possible that these studies are simply underpowered. However, I suggest that there is an effect of attention, but it is so minuscule and so rapid (Theeuwes, 2010) compared to the higher level cognitive effects (learning, expectation, estimation) that is not observable in high level response even in a very large sample size. A valuable next step in understanding how low level visual orienting might be “washed out” by higher level processes would be to use an eye tracker on an enlarged version of candlestick icons: in keeping with earlier work (McColeman & Blair, 2015), it may be that participants make fewer, but slower eye movements to the arrow head Close dimension

to offset the peripheral orienting effects of arrow cues and maintain a more egalitarian sampling strategy for the four relevant dimensions. It may also be that attention is drawn to the arrow head, but that the rest of the stimulus space is revisited so often that the single fixation is ineffective at anchoring the Close dimension as a perceptual cue to inform the participants' price estimate. For the same reason, it is expected that employing a rapid stimulus presentation phase to limit the number of fixations would foster a performance advantage for the Close dimension in the arrow body condition.

Overall Experiment Five tested the impact of changing spatial properties of the candlestick icon relative to a consistent graph space. Regardless of the Spread condition (zoomed in, zoomed out, or standard), the user performance error (UPE) was consistent. Extracting information from a single candlestick icon, then, appears not to be effected by the Spread extent manipulated in the experiment. The spacing in Experiment Five is not relative (or categorical, as per van der Ham, van Wezel, Oleksiak & Postma, 2007), though, in that there is only one glyph presented each trial. Though necessary to test for any influence of having glyphs that appear further from the Y axis, the real relative spacing experiment is Experiment Six.

4.5. The Influence of Relative of Glyph Spacing on Forecasting Multivariate Trends

While in the above Experiment Five, the candle body condition and the glyph spacing appeared not to influence the interpretation of values associated with specific queried dimensions, it is not necessarily the case that they play a negligible role in forecasting. In Experiments One and Two (where the task was also to extract the value from a single queried dimension) the impact of candle condition was minimal, but then in Experiment Three (a forecasting task) the candle body did influence performance. It appears, then, that simple information extraction is less impacted by manipulations of the physical manifestation of graphic icons than more complex information integration as is the case with forecasting.

Building on these earlier findings, Experiment Six tests whether manipulations of glyph spacing impact user performance in a forecasting task. Additionally, Experiment Six tests the crossing of candle body condition and glyph spacing to test whether any impact of glyph space is mediated by the candle body type (Table 4.1). It was expected that low

Spread would support better pattern perception. While not found in Condition A, such results were observed for participants in Condition D. Surprisingly, though, such results were also observed high Spread: it appears that for Condition D, the worst performance is in the default spacing condition.

4.6. Experiment Six: Methods

The experimental conditions are the same as in Experiment Five (Table 4.1), but otherwise the structure of Experiment Six is the same as Experiments Three and Four. Participants are presented with a pattern of time series data resembling four dimensions of a stock price over time and then are asked to complete the pattern by adjusting the icon to best fit with the previously presented stimulus.

4.6.1. Participants

Twenty-six participants were assigned to Condition A (n=16) or Condition D (n=10). Within those groups, participants were then assigned to High, Low or Standard spread as per the Spread definition in Experiment Five.

4.6.2. Apparatus and Materials

As in earlier experiments, graphs were defined by a set of vector autoregression models' outputs, where the values were then multiplied by a constant to determine how far off of the centre of the Y-axis the graph elements were drawn. In the High Spread condition, the output was multiplied by 80 pixels. In the Low Spread condition the output was multiplied by 20 pixels. In the Standard Spread condition, the output was multiplied by 40 pixels. As an observer, the High Spread condition appears more erratic than the low spread condition.

4.6.3. Procedure

The experiment proceeded just as Experiment Three: the participants received instructions, then began the experiment. During each trial (to a maximum of 200 or 55 minutes — whichever came first) the participant sees a fixation cross, then a stimulus,

then a simple interface where they could adjust and manipulate the glyph that they believed best completed the series presented in the stimulus and then finally they received points to indicate how close they were. The points acted as feedback. More points were awarded for better performance.

4.7. Experiment Six: Results

The analytic strategy from Experiment Three was used again here, but with modification. The linear mixed effects model for Experiment Six contains Spread as an additional categorical predictor, and so the variance in performance error was estimated using three predictors: Trial, Condition and Spread. The was a main effect of condition on performance, in that Condition D was worse than A (β = .944, T = 2.315, p = 0.021). There was also a main effect of Trial (β = -0.001, T = -2.677, p = 0.007). Of particular interest, though, is the interaction between Condition and Spread, specifically between Condition D and Low Spread. It was expected to be the condition that corresponded to the best performance, given earlier findings suggesting a triangle candle body advantage (e.g. in Experiment 3) and the hypothesis that gathering an overview of the pattern would be easier if the relative change between icons was reduced. Indeed this is the case, as there is a significant interaction between Condition D and Low Spread (β = -1.411, T = -2.681, p = .007). However, there is also an interaction between Condition D and High Spread (β = -1.494, T = -2.99, p = 0.003).

Table 4.3 Coefficient estimates from linear mixed effect model predicting log(UPE) Experiment Six considering Candle Condition and Spread

	Coefficient Estimate (β)	Standard Error	*T*	*p*
(Intercept)	5.0687	0.1186	42.7477	0.0000
TrialID	-0.0012	0.0004	-2.6769	0.0074
Condition D	0.9441	0.4079	2.3146	0.0206
Low Spread	0.3179	0.2993	1.0624	0.2880
High Spread	0.2407	0.2230	1.0792	0.2805
TrialID x Condition D	-0.0054	0.0038	-1.4395	0.1500
TrialID x Low Spread	-0.0007	0.0028	-0.2556	0.7983
TrialID x High Spread	0.0010	0.0010	0.9776	0.3283
Condition D x Low Spread	-1.4119	0.5265	-2.6818	0.0073

Condition D x High Spread	-1.4940	0.5004	-2.9858	0.0028
TrialID x Condition D x Low Spread	0.0062	0.0047	1.3149	0.1885
TrialID x Condition D x High Spread	0.0064	0.0041	1.5649	0.1176

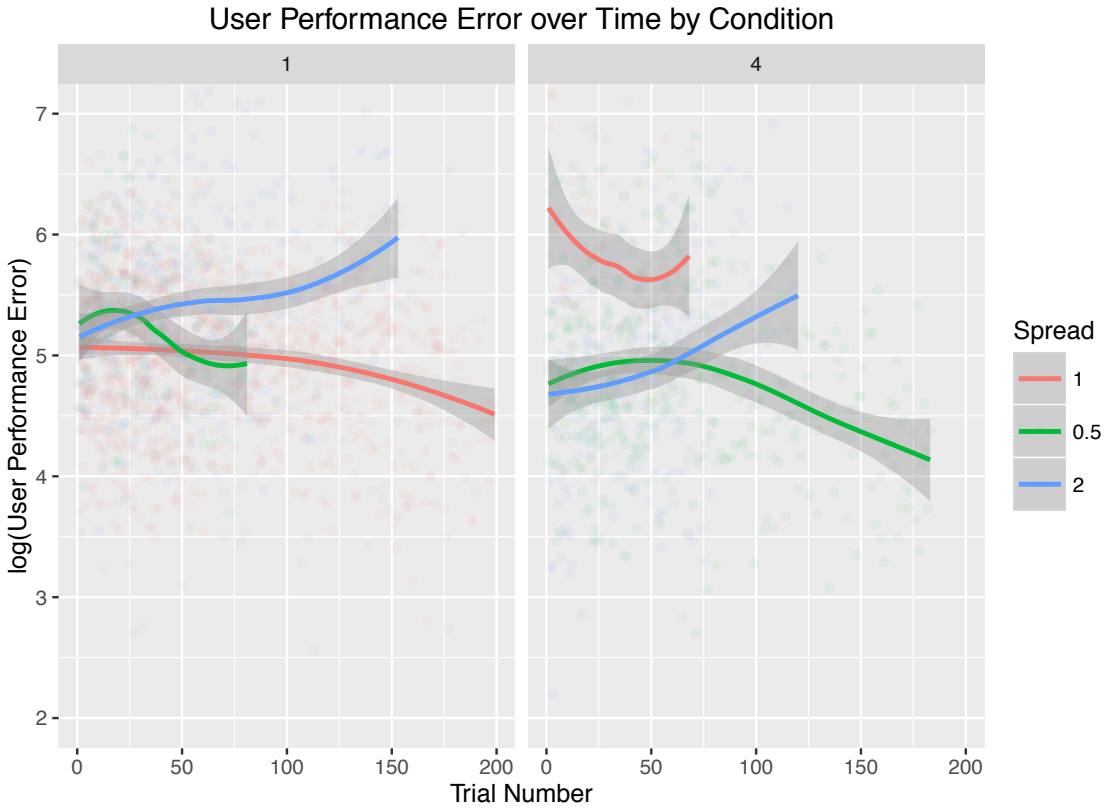


Figure 4.3. User performance error for Experiment Six.
Each panel represents one Condition. The lines represent LOESS curves for each of the spread conditions, and the grey shadows represent the standard error of each.

Since performance is better in both High and Low spread circumstances, it is untenable to suggest that simply lowering Spread yields better performance. Rather, the most reasonable interpretation is that the standard Spread supports the worst performance in among those assigned to the triangle candle body group.

4.8. Experiment Six: Discussion

The sample size is quite low for such a complex experiment design, and so lessons

from these results should be interpreted cautiously. However, this is a good starting point to consider the relationship between coordinate spatial representation (of the sort queried in Experiment Five) and categorical spatial representation (elicited by this Experiment Six). Building on van der Ham *et al.* (2007), categorical spatial representation is the type of spatial cognition that is represented in two or more dimensions and is relative to other objects in the display. In Experiment Six, the categorical representation for the user-generated glyph is “above” or “below” the penultimate icon in the time series. Said differently, the user can draw in the icon that best completes the series relative to icons that were presented previously.

An advantage of this experiment design is that participants must only adjust values in the Y dimension on the graph. In that sense, the vertical coordinate space is the only dimension free to vary when the participant completes the series. In this task, the final glyph is trivially the last item in the series, and so it could not go anywhere but the final slot on the graph. The selected glyph snaps into place on the X axis, and the participants adjust its height and its position along the Y axis. As such, the meaningful dimension — the value of the stock — is the only one that the participant has to make any spatial judgements about.

Perceiving a data point in a data visualization, interpreting it correctly and then using that interpretation to make a decision all involve a complex set of processes. The experiment design in Chapters 2 and 4 builds up from the relatively simpler case of extracting a single dimension to the more complex case of using information from a candlestick plot to make a forecasting judgement. The effect of candle body on user performance error is only observable in the forecasting task, suggesting that the more complex task environment is increasingly sensitive to manipulations in the physical representation of information.

4.9. Chapter Four: Discussion

At the end of the first four experiments, there are tests of how the different representations of candlestick plots impact how quickly people can learn how to use them (Experiment 1), the impact of adding grid lines to help with reading the plots (Experiment 2), how good participants are at completing a pattern of data presented in candlestick form

(Experiment 3) and the impact of an uninformative dimension represented in a complex multivariate data visualization (Experiment 4). Generally, the candle body manipulation impacts performance only in the more complex task environment.

Experiments 5 and 6 were designed to test the importance of relative spacing in how participants understand candlestick plots specifically, but also spatially distributed visual objects generally. The relative spacing of individual icons appeared to do little to impact participants' capacity to extract dimension information from candlestick glyphs. However, in the forecasting task, the relative spacing of glyphs did make a difference for those assigned to the triangle candle body: these participants were better at the forecasting task when the glyphs were relatively close together, in line with initial hypotheses. However, contrary to expectations, the participants in Condition D were also better at forecasting when the glyphs were further apart. The hypothesized trade-off between precision and pattern perception was not supported by the observations in Experiments Five and Six.

A fascinating observation in both groups, however, was that the user performance error increased for participants assigned to the high spread group. This is odd, given that most performance error is level or decreasing over the course of the experiments. While it requires its own set of studies to understand more fully, it may be the case that the erratic-looking presentation of information in the high spread condition invites more rapid fatigue or indifference in participants.

Among all of the complexity invited by Experiments Five and Six, it does appear that using a low Spread and triangular candle bodies would support generally better performance in candlestick plots.

Chapter 5. Chapter Five: Testing Perceptual Economics

The experiments to this point all explore what it is to read a graph: the role of salience, grid lines, distracting information, and distributing icons in space. It's necessary to examine the perception of a graph, especially for a plot like the candlestick which hadn't yet been studied in the context of human perception. A second contribution of this thesis is to move toward a science of economic data visualization by considering perception and decision making together. Integrating perception and action across multiple levels of abstraction is a challenging pursuit for cognitive science. Data visualization is an effective domain to attempt to do, as discussed below. This experiment is meant to serve as the capstone, building upon findings from Experiments One through Six to test whether improvements to candlestick plot design informed by results from the simpler task environments can be combined to make a better graph for users compared to standard candlestick plot.

5.1. Experiment Seven

To examine the influence of perceptual changes on decision making behaviour, we must first have a task in which participants are making a decision. The information extraction experiments (One, Two, and Five) are relatively simple tests of the impact of perception on reading a information contained by candlestick-type glyphs. Building slightly in complexity are Experiments Three, Four and Six, forecasting tasks which require the participant to integrate information from a series of glyphs and generate an estimate of what comes next in the series. The forecasting tasks to this point require extrapolating from a pattern, which has the properties of decision making, but does not actually require a discrete response in the midst of uncertainty like traditional decision making literature does. To bridge the findings of perceptual differences on pattern perception to decision making literature, this experiment is an extension of forecasting tasks in that participants make a decision based on the information they extract from the presented time series data.

5.1.1. A Science of Economic Data Visualization

Movement toward a science of data visualization is gaining momentum, and for good reason. Data visualization is increasingly important as a communication tool, and society is relying upon graphs to assess data and make decisions. Although there have been practices established for how best to represent data (Bertin, 1983) and empirical work to verify the efficacy of different representations (Cleveland & McGill, 1984; Croxton, 1932) there remains a gap. There needs to be a cohesive framework within which to study data visualization. While practitioners of data design have outlines of best practices, or intuitions of what works in data visualizations they have not fully explored the visualization problem space (e.g. establishing best practices for visualizing live, ever-changing datasets; Rensink, 2014), and the deluge of data in recent years is introducing a new class of problems that earlier data visualization designers didn't need to consider. Even if the type and volume of data was not rapidly changing, the effectiveness of existing design practices remains largely unexplored. Establishing a science of data visualization will help to both inform and assess designers' data visualization innovations (Resink, 2014; Ware, 2004).

Basic science does benefit from applied sciences' demand, on occasion (Shneiderman, 2016). Ware notes that "[j]ust as engineering has influenced physicists to become more concerned with areas such as semiconductor technology, […] the development of an applied science of visualization can encourage vision researchers to intensify their efforts in addressing such problems as 3D space and task-oriented perception" (2004, pp. 27). The external impetus for vision science to explore more complex tasks to inform data visualization may exist, but is likely overshadowed by the need for vision science insights in advancing in virtual reality or computer vision. Both virtual reality and computer vision benefit advances in data visualization, of course, but visualization is not the driving force for next steps in vision science. Alternatively, considering data visualization as its own, pure science that encompasses both the generation and interpretation of visualization can be a framework to understand human cognition more broadly.

Considering physics as a model science, a good framework explicitly describes a set of entities the field is interested in, questions that can be asked of those entities and how to answer those questions about entities in the field. Rensink (2014) suggests that a

science of visualization will then need to effectively define visualization, identify questions we can ask of visualizations and how to evaluate those questions. He notes that a good scientific framework for data visualization will consider “the nature of the visualization task, the computational issues involved, and the nature of the human viewer” (Rensink, 2014, pp. 2). That quotation captures a critical step forward: Rensink’s outline of a science of visualization collects both the generation and the perception of a visualization under the “science of visualization” umbrella while acknowledging that different problems (visualization tasks) require different solutions, and that there is an important human element (the nature of the human viewer) to explore.

5.1.2. Information Processing

The concept of extended-vision (Rensink, 2014) holds that the human observer and the data visualization can be treated as one information processing system. In the extended-vision information processor, the first step is transforming data to a graph, the second is transcribing the graph to a visual representation and the third step is transforming the visual representation into the conceptual representation. The information processing approach is a familiar one in cognitive psychology. Decades earlier, Marr (1982) suggested three levels of analysis that are required to understand information processing systems: the computational level, representational level and physical level.

To fit Rensink’s extended-vision thesis into Marr’s information-processing framework it is necessary to split the generation and perception of data visualization back up again. Consider the generation of a visualization as one information processing problem (Simkin & Hastie, 1987), the perception of that visualization as a second information processing problem and add a transcription step between generation and perception. Said differently, each of Rensink’s “transformation” steps is a separate information processor. The transformation steps are nominally information processing steps (transforming input to output) and there are reasons to believe the graphical and visual representations can be dissociated (e.g. if a graph uses red/green but the observer is colourblind) so the graphical and visual representations are not part of the same information processing unit. The collection of these three components (visualization generation translation, visualization perception translation and transcription) can be grouped as a single, larger processor to maintain the framework Rensink builds for a

science of data visualization. This spin on the extended-vision thesis also allows that the output of the generation transformation acts as input of the perception transformation. Practically, another advantage of thinking about the transformation steps as paired information-processing problems means research from computer graphics and human vision can directly inform the generation and perception transformations in the extended-vision thesis. Additionally, research on human-computer interaction can aid in understanding the combined human observer/data visualization unit.

5.1.3. Measurement

While the qualities of data visualization can be couched in general vision terms, there still remains some critical steps to take to have a full scientific framework. Above, the properties set as a target for a science of visualization were a) an understanding of what the science of visualization includes, b) the types of questions that can be asked in the framework and c) a way to evaluate those questions. The extended-vision thesis (Rensink, 2014) encompasses the interest of a science data visualization: it's the system of an observer and a graph that together uncover properties of data. Keeping that general definition satisfies the first property. There are a massive number of specific types of questions that can be asked in a domain. Given the current science of visualization framework, the possibilities are based on the graphs' performance (e.g. memory load, display errors) and observer's experience and performance. Of particular interest is how effective the graph is in helping an observer get an accurate understanding of the data.

Recall that the representational level is one of three levels of analysis Marr posits as necessary to understand an information-processor. Visual representation, in his theory, allows three possibles stages of interpretation after the image is presented: the primal sketch (basic two-dimensional properties like blobs and lines), the 2 1/2 dimensional sketch (including, for example, orientation and depth information) and the 3 dimensional model representation (including, for example, the layout of shapes and representation of shape volume).

Cartographer Jaques Bertin published *Sémiologie Graphique* (1967; English Translation: *The Semiology of Graphics*, 1983) to describe how best to graphically represent multidimensional data. Like the extended-vision thesis, Bertin's contributions

can be aligned with Marr's theory. Rather than speaking of information-processing, though, this is specifically about a representational framework for data visualization. Bertin notes that static visual graphics can represent information through plane variables (a point, line, the area of a mark, the organization of marks) and "retinal" variables (size, value, texture, colour, orientation and shape). Plane variables share properties that describe Marr's primal sketch while Bertin's "retinal" variables are of the type included in Marr's 2 1/2 dimensional sketch. The parallels between the theories of a cartographer trying to establish best practices for creating graphs and a vision scientist trying to unify a framework for his field are compelling (Pinker, 1990; Green, 1998), and point toward an interdisciplinary convergence of ideas. The definition of planar and retinal variables is not Bertin's primary contribution, though. Image theory is.

Image Theory (Bertin, 1983; 1967) has been used as a framework for understanding the perception of multi-dimensional visualizations. His "image" is rapid impression formed after the presentation of a visual stimulus. For example, the image of a scatter plot would be a cloud of points. He describes a good graphic as showing "not only the leaves; it should show the branches as well as the entire tree" (pp. 10, 1983). He also provides a general operationalization of what he suggests is the best standard measurement for graphs: efficiency.

Like a lot of cognitive psychology researchers at the time, Bertin's measures of choice were accuracy and reaction time. His theory of how people read information off of graphs is a serial, multi-stage process like strict information approaches of the time, and so increased reaction times indicated increased processing load. If an observer can answer a question asked of them correctly off of two different graphs, the better graph is the one that requires less time for the observer to make a response. The advantage of efficiency over other measures is that it can be applied to any kind of graph, and any kind of question that has an objectively accurate response. The downside of relying on efficiency as a measure is that it is as opaque as any other reaction time measure and gaining meaningful insights requires clever experiment design.

5.1.4. Decision Making

The science of visualization is defined as the system of an observer and a graph

that together uncover properties of data. Questions that can be asked in the science of visualization are tricky to specify because, up until now, there are a lot of different kinds of graphs, a lot of different kinds of people, and together they comprise a lot of different visualization systems. The questions of the science of visualization are questions about the graphs' performance and the observers' performance separately and together. The measurement that seems applicable to all circumstances is efficiency. This is vague and under-constrained in part because there are so many details and various applications of graphs that it's hard to think about generalities that capture a sufficient number of them to include in an outline for the science of visualization.

This changes if you add decision making. If the goal of a data visualization is to communicate synthesized data points for easy and clear interpretation by an observer, then the data should influence the observer's decision. A graph's job is to communicate the properties of data to inform decisions, and so including a decision outcome after the presentation of a graph offers a more cohesive framework for scientifically approaching data visualization. Studies asking people to report a value directly off of a graph without interpreting it are good for testing the observers' surface level understanding of the data, but they do not capture the influence of different graph features in providing a general impression of the data and don't offer a good understanding of the graph and the observer as a system, which is a goal of a science of visualization.

Including decision making in a science of visualization also acknowledges that an observer will be influenced by prior knowledge, biases, heuristics and their own utility functions in response to whatever is being visualized. Having decision making as an explicit part of the scientific framework means that measurement from traditional decision science can be used for evaluation. Additionally, this offers a connection to real world data visualization problems. Graphs are meant to represent data, and data are meant to support decisions. Using realistic decision-making problems that include data visualizations can assuage some concerns about scaling and generalizability (Kingstone, Smilek, Ristic, Friesen & Eastwood, 2003). Even if the academic interest is more about simple perception of graphs than about higher level decision making, recording a continuous decision outcome in addition to the traditional reaction time and accuracy measures offers a third dimension to basic psychophysical explanations of behaviour.

Economic and decision science is a natural extension of the science of

visualization as outlined by Rensink (2014). Behavioural economics famously stands on the shoulders of early cognitive psychology. Economics, generally, is about the allocation of valued resources. The influence of data visualizations on decision making can be a one-off research program, or it can be the start of a bigger field of the influence of visualizations on decision. Different physical representations of information influence how people pay attention to stimuli (McColeman & Blair, 2015), and the manner in which people pay attention to stimuli influences their decisions (Armel, Beaumel & Rangel, 2008). It is expected, then, that manipulations of data visualizations influence people's decisions. If so, this has critical implications for design practice and ethics.

5.1.5. Decision Science

An early extension of work on decision making with uncertainty, as introduced by Bernoulli (1738; c.f. translated 1954), acknowledges that a dollar for a poor person is more valuable than a dollar for a rich person. Bernoulli moves from objective, expected value to personalized expected *utility.* He understands utility of a dollar in the context of that dollar compared to each person's previous gains. Bernoulli's theory was a significant step forward in understanding the subjectivity (or at very least the contingent nature) of decision making. Additionally, he notes that lower material wealth corresponds to loss aversion in risky decisions relative to the same potential loss for the well-to-do. It's this consideration of the biased, fallible, inconsistent, and emotional human behind the choice that motivates much of decision science — this work included.

Centuries of economic theory stand on the shoulders of Bernoulli's insight. Expected Utility Theory (von Neumann & Morgenstern, 1944) did a respectable job of describing group decision making behaviour, and fit with the then-dominant idea that humans were rational animals. Upon inviting some possibility that humans are not good intuitive statisticians and allowing that humans are not always rational decision makers, it became apparent that Expected Utility Theory was less effective at predicting the behaviour of individuals than of groups (Kahneman & Tversky, 1979; Rabin, 2000). Some people do buy lottery tickets, even though the expected utility is not in their favour. Two scenarios that have the same expected utility can yield different responses based on how the information is framed. The framing manipulation was through text vignettes, so at a higher level than the information presentation manipulation in the current work, but the

sentiment of the manipulation is similar: the way in which information is accessed impacts how people make economic decisions. As an improvement upon Expected Utility Theory, Prospect Theory distinguishes two phases to model decision making under risk: editing (also known as framing, the phase for understanding the problem) and valuation (assessing possible outcomes) (Tversky & Kahneman, 1992).

The editing phase is where data turns into usable information. It encompasses processes that filter input and simplify the options before valuation: coding, combination, segregation, cancellation, simplification, and dominance detection, (Kahneman & Tversky, 1979). Coding, in Prospect Theory, is the perception of outcomes in some context such as gains and losses. The processes included in the Coding step of Prospect Theory mirror much of cognitive psychology's interests, where researchers examine what information in the world means for an observer, and how it comes to take that meaning. Combination, segregation, and cancellation can be described as algebraic reductions of the problem space; simplification is the process of rounding values; and dominance detection means spotting a particularly poor (or dominated) choice and eliminating it from contention.

The two phases of prospect theory are nominally simple, but mask immense complexity. The focus of the current work — the visual representation of decision-relevant information — is captured by Coding in Prospect Theory, which in itself is just a subsection of the first phase. Even if the editing phase were simpler and could be fully understood (which, at the moment is not true) moving from edited information to decision is non trivial.

Decisions usually require some deliberation and yield some categorical outcome (Gold & Shadlen, 2007). They can be as innocuous as "chicken or fish?" for dinner, or as weighty as "should we pull the plug?" for someone on life support. Most people respond differently to these questions; less important decisions appear to require less deliberation and less evidence before moving onto to making a choice. Even if the decision-maker is given probabilities associated with outcomes in each scenario, emotional responses to the situation and the potential outcomes impact the decision-maker's perception of the scenario (Zajonc, 1980; Thaler, 1980) or the strategy they employ to make a decision (Mikels, Maglio, Reed, & Kaplowitz, 2011). The subjective value of potential outcomes effectively predicts what decision is likely to be made (Tversky & Kahneman, 1981).

Decisions are often influenced by bias, emotion, context, or even the modality

through which information is presented. However, we know little about how visual representation of information affects decisions. This blindspot, the relationship between visual representation and decision making, is unfortunate because visual representation of information is a serious factor in decision making. This is especially important if we rely more on data driven decisions in the age of Big Data 2.0 (Provost & Fawcett, 2013), in which we more frequently make data driven decisions informed by visual representations of data. Much of the editing phase of Prospect Theory can be offloaded to a data visualization, freeing up the cognitive resources of the observer to dedicate to the valuation phase and making a good decision (Corso, Hammitt & Graham, 2001).

5.2. Experiment Seven: Method

Rather than adjusting a selected glyph to complete an observed pattern, the participants were asked whether they would like to buy the stock that they're presented with if they have to sell it at the end of the next time step. This is essentially asking, "do you think the stock will close higher tomorrow than it closed today?" The design of the stimuli builds on results from the earlier experiments. The ability of participants to effectively learn how to make judgements in the face of uncertainty is tested in the context of two types of candlestick plots.

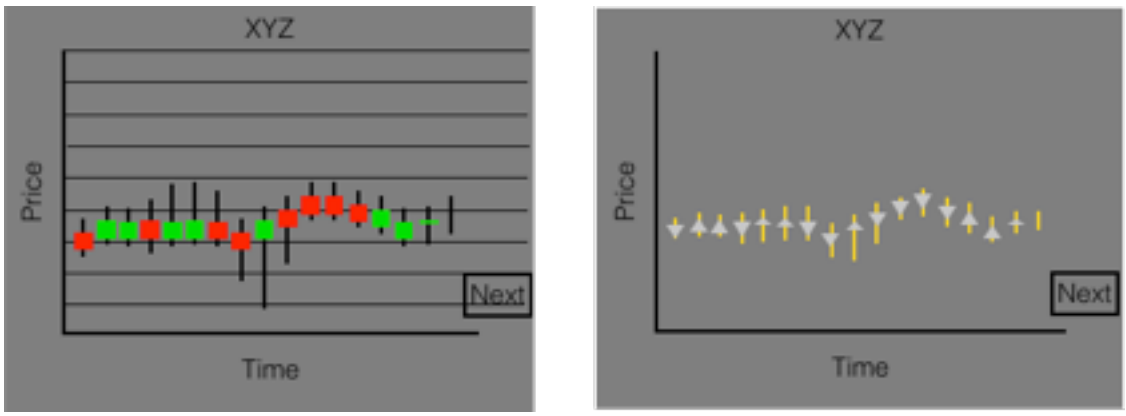


Figure 5.1. A schematic representing the between subjects conditions in Experiment Seven

The qualities associated with the worst user performance in earlier experiment are combined to generate stimuli for the first group. The second group is presented with stimuli associated with better user performance.

The participant will start the experiment with 2500 points (the equivalent of having already scored perfectly for five trials in the forecasting experiment). They know they're being asked whether they'd like to buy stock if they have to sell it at the next time step, but they're not told how to read the candlestick chart. On each trial, the participant sees the fixation cross, the stimulus, and is then prompted to answer "yes" or "no" to the question, "If you buy this stock, you will have to sell it at the end of the next day. Do you want to buy it?" If they answer yes, they'll be asked how much they want to buy, from their 2500 points. The difference between the closing price on the final day from the stimulus and the closing price that they bet on is their score for that trial. Their score is multiplied by 10% of their bet (to keep number manageable). For example, if the participant bets 50 points that the stock closes higher the next day, and it does by 15 points, the participant adds 75 points to their total score for 2575 following the payoff formula

$$\text{payoff} = 0.1(\text{Bet}_t) * (c_t - c_{t-1}) \qquad (1)$$

where *t* is the time point, Bet is the number of points the participants wagers on the closing price at time *t* being higher than the closing price at time *t*-1, and *c* is the closing price.

During the feedback phase, the participant's answer cannot be shown since they are explicitly queried only about one direction. While in Experiments Three, Four and Six, the participant's response was displayed before the correct answer, in Experiment Seven, only the number of points awarded is displayed. Otherwise, the methods are the same as Experiment Three.

5.3. Experiment Seven: Results

Table 5.1. Coefficient estimates from linear mixed effect model predicting log(UPE) Experiment Five considering Candle Condition and Spread

	Coefficient Estimate (β)	Standard Error	*T*	*p*
Intercept	71.34	125	0.57	0.57
TrialID	0.65	2	0.40	0.69
Condition D	-28.21	183	-0.14	0.87
TrialID x Condition D	-0.63	3	-0.25	0.80

Prospect Theory suggests that people will be risk averse in uncertain

environments. The analysis tested this first by looking that the number of points earned over the experiment: the idea being that as participants learnt about the task, the structure of the problem would become more predictable and their task uncertainty would decrease. A linear mixed effects model with Trial and Condition as predictors examined the effect of each on the reward; it was predicted that participants' reward would increase with Trial as they develop more confidence in reading the graph and attempt to maximize their payoff. This prediction was not supported by the linear mixed effects model fit to points earned.

The model shows no effect of condition or trial. A post-hoc test was conducted on the proportion of trials in which no risk was taken. The mean proportion of trials with no earnings was not statistically greater in Condition D (mean proportion of trials with zero risk = 0.130) than in Condition A (mean proportion of trials with zero risk = 0.052) as shown by a t-test (T = -1.296, df = 10.192, p = 0.224).

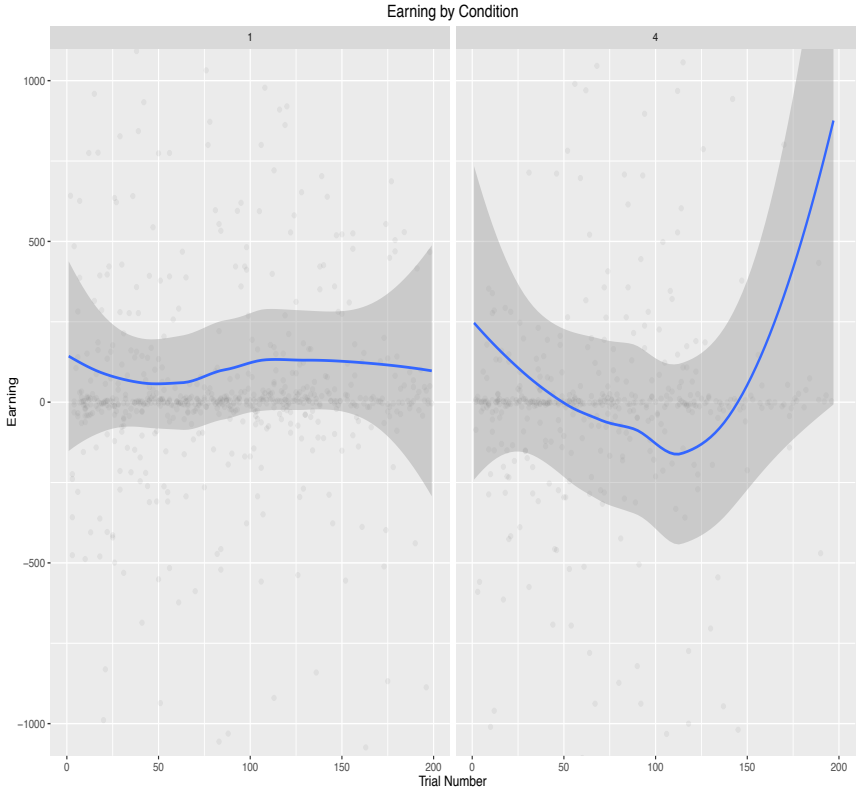


Figure 5.2. User performance between group in Experiment Seven. There was no overall difference between groups

5.4. Experiment Seven: Discussion

When tested in an applied setting, the candle body associated with the best performance in the best circumstance (triangular body, no grid lines, as indicated by Experiment 2) did not support better investing decisions than the candle body associated with the worst forecasting performance (Condition A, Experiment 3) that most closely mirrors actual practice. The physical representation of the candlestick icons did not appear to impact gross performance indicators of investor confidence, either, in that both groups opted not to bet at similar rates, and their earnings were comparable.

It was expected that, given the same number of remaining points, participants assigned to the red/green candle body condition would be more sensitive to the visual representation of uncertainty and would be loss averse in keeping with earlier results from prospect theory (Kahneman & Tversky, 1992). Results based upon prospect theory are well-replicated but cultural differences have been noted (Brumagim & Xianhua, 2005). Failure in this Experiment 7 to observe differences between groups in response to the physical properties of the candlestick plots may be due to cultural differences, but the participants' cultural background was not surveyed and so that explanation cannot be explicitly tested with these data. Perhaps more interestingly, and more in keeping with findings throughout this thesis, is the possibly that visual representations of uncertainty do not impact participants' decision-making as much as uncertainty represented in prose. The method by which properties of the environment are conveyed to participants impacts how those properties impact the participants' understanding and subsequent decisions.

Chapter 6. Conclusion

Using a science of economic visualization can help practitioners to create better data visualizations to support effective decision making. It is an interesting science in its own right, though: we can use lab-generated data of the sort used through these experiments to control the input of the data visualization process and gather knowledge through careful rigour, and we can use real-world decisions and the graphs that informed them in the interest of external validity in developing and honing psychological theories.

In these seven studies, the candle Condition was found not to impact simple information extraction (Experiments One and Two). It was found, however that the presence of grid lines negatively impacted performance, supporting some visualization practitioners in their stance that everything in the graph should be informative (Few, 2004; Kosslyn, 2006) and the rest is “chart junk” (Tufte, 1983). In Experiment Three, when participants were forecasting what came next in a financial time series represented by candlestick plots, there was an impact of Condition on performance: people were more accurate in their judgements when they were assigned to Condition C and reading a graph with the triangular candle body. This difference did not hold in Experiment Four, where the presence of an irrelevant dimension removed the triangle candle body advantage relative to the rectangular candle body in Condition A. Experiments Five and Six tested the influence of the icon size, wherein Experiment Five was information extraction task and Experiment Six was a forecasting task. There was no advantage of the triangle candle body in Experiment Five, but there was in Experiment Six. In Experiment Six, the triangle candle condition outperformed the rectangle candle condition when the icons were smaller or larger than normal.

Putting it all together, Experiment Seven was meant to test whether the general data-driven improvements to candlestick plot design would allow for better performance during actual risk-taking. There was no discernible difference between the two groups, however, identifying another source of complexity in predicting human performance, even when using the same stimuli for different tasks.

6.1. The importance of task environment for testing cognitive phenomena

The physical properties of a data visualization can differently impact the observers' performance on a task, depending on what the task is. The first major lesson from these experiments is that task is crucially important to what variables influence performance. In this thesis, there were three broad categories of task: information extraction (Experiments One, Two and Five), forecasting (Experiments Three, Four and Six) and risky decision-making (Experiment Seven). The influence of the physical properties of the candlestick icons on performance changed in different task environments.

In these seven experiments, all of the dimensions represented in the candlestick plot (with the exception of the noisy dimension in Experiment Four) were important to completing the task, but there were differences between tasks that elicited different performance across the four dimensions. Participants' user performance error was not impacted by the different candle glyph types in the information extraction tasks (Experiments One, Two, and Five), but it was during forecasting (Experiments Three and Six). This variation may be due in part to different task requirements on the cognitive resources of participants while interpreting the candlestick plots. Attention to task-relevant subsets of the visual environment is known to change as the task does.

For example, eye movements can be used as an index of overt attention. When participants were asked to extract different information from a painting, their eye movement patterns changed in response to different questions (Yarbus, 1967). If the task was to estimate the ages of the people in the image, the participant looked to faces; if it was to estimate the wealth of the family in the image, the participant looked to furniture and around the room. When different subsets of the environment are more valuable for supporting specific task goals, they are more likely to be fixated. In this regard, there was a relatively simple correspondence between the task goals (extract requested information) and the oculomotor activity of the participant (search for properties of the painting that match the task set). In that case, information gathering matches closely with information requirements. While the painting was the same for the different questions, the oculomotor movements were not. Since the eyes are the gate by which visual information enters the cognitive system, Yarbus's findings suggest that information gathering changes flexibly based on the task requirements.

In the present information extraction experiments (One, Two and Five), there is also a relatively straightforward relationship between information gathering and information requirements. Because the participants do not know which dimension they will need to report, judicious performers will gather the position of all four in preparation for the response phase of the trial. Treating each dimension as a separate item requires maintenance of four values in working memory (one each for open, close, high and low price). It is reasonable to maintain four items or chunks in working memory given historical (Miller, 1956) and more contemporary estimates (Cowan, 2001) of working memory capacity.

Since each dimension of a single icon can be maintained in working memory, it might be expected that perceptual costs of inefficient icons (Condition A) would be minimal as long as participants had enough time to sufficiently attend to the four dimensions. As long as they are effectively encoded, they can be maintained. Given earlier work that shows attention driven by arrows (Kingstone, Smilek, Ristic, Friesen & Eastwood, 2003), it is likely that there is actually an advantage of Conditions C and/or D for drawing attention the close price, but the self-timed nature of the tasks in this paper render it unobservable. The structure of these experiments is such that even if the first dimension encoded into working memory was the close dimension on every trial, it would not be reflected in user performance error if the other dimensions were encoded too. A promising avenue for future research is to replicate Experiment One with two additional conditions manipulating the stimulus presentation time. For the speeded condition, participants may view the candlestick glyph for only 750 milliseconds, approximately long enough to make three fixations. In such an environment where time pressures are high, participants are made to select as much of the information as they can. It is expected that in this environment, for Conditions C/D, that the close price would be selected and more accurately reported than the remaining dimensions. Additionally, it's expected that the close price would be more accurately reported for Conditions C/D than the close price in Condition A, where the shape of the icon does not reflexively draw attention to the part of the glyph that represents the close price.

Even with the present data, though, better performance is observed in the triangle candle bodies during forecasting but not in information extraction. Both studies are self timed during the stimulus presentation and response phases of the trials, so it's not just a

matter of how much opportunity participants have to extract the information. There must be an additional difference introduced between tasks that showcases the differences between glyph representations in one task type but not the second.

Navalpakkam and Itti developed a model that integrates task differences with physical properties of the environment to predict attention (2005). The model takes a visual scene, prior knowledge and task set as input. Low level features are extracted from the image (Itti, Koch & Nieber, 1998) to output a gist, a layout map, and a salience map. Together with task specification and prior knowledge, the gist, layout and salience map are used to calculate where attentional selection is most likely to occur in topographical space. This is a critical step toward a realistic representation of how cognition operates, but there remains a couple of gaps before such a model could be used to explain the cognition of data visualizations. One remaining question is how the spatial representation for the selectively attended item/dimension corresponds to externally meaningful values. During information extraction tasks, participants are asked to translate a point in two dimensional space into a dollar value. This translation process is another step between the participant and the data during which any existing differences between candle conditions may be washed out.

Another gap between good models of human cognition and their application to slightly more complex tasks such as data visualization is that they cannot determine how attention, driven by the task environment (and other top-down factors) and exogenous salience (and other bottom-up factors), relates to the overt response. While great for eye movements or other metrics of selective attention, models such as the one by Navalpakkam and Itti (2005) are not equipped to predict higher level performance.

Category learning models have better connected attention to stimulus dimensions and performance (e.g. Barnes, McColeman, Stepanova, Blair & Walshe, 2014; Gluck & Bower, 1988; Kruschke, 1992; Kruschke & Johansen, 1999; Love, Medin & Gureckis, 2004), but in these models it is assumed that the task is always to make a category judgement. As such, the models better connect attention, environment and response, but with the very notable advantage of having constrained the response space to a few categories. Most models of category learning also fail to consider the visual properties of the environment that can draw visual attention (excepting Barnes, Blair, Tupper, & Walshe, 2015; Kruschke & Johansen, 1999). Unsurprisingly, models developed in

different domains have different pros and cons: visual attention models fail to consider response (Navalpakkam and Itti, 2005) and categorization models often fail to consider visual factors.

As discussed in Chapter One, some models of data visualization are better poised to integrate perception, attention, and decision making in response to visual stimuli, though perhaps at such a high level of abstraction that it's difficult to use them to predict human performance. In all of visual attention, category learning, and data visualization, though, there is yet to be model that can account for the influence of changing task environments on the perception of the same stimulus set. This is a critical opportunity to integrate levels of analysis and to consider a more nuanced approach to how humans connect to their environment, or how we act as a system with the information we observe (van Wijk, 2005). Data visualization is a very promising avenue for cognitive psychology research.

When visual environments become sufficiently complex, the observer is more likely to see patterns. It may be that the 99 glyphs presented in the forecasting task are far too numerous to consider each glyph individually, and so spatial selective attention to individual icons is fundamentally the wrong way to consider human performance. Indeed, 99 well exceeds estimates of working memory capacity. Rather, the forecasting tasks invite a richer pattern perception, where the whole time series representation is greater than the sum of its glyphs.

6.2. The interconnectedness information visualization components

A second lesson from these studies is the interconnected and complex way in which information visualization is impacted by manipulations to their components parts: simply adding an uninformative dimension (Experiment Four) eliminated the effect of candle body type previously observed in Experiment Three. The connection between visual input, cognitive processing of that input and higher level output is non-linear. Whether the introduction of an unpredictive dimension in that experiment made the forecasting task too difficult or whether it pervasively impacted the perception of the remaining dimensions has yet to be tested, but in either case the effect of introducing

noise washed out any earlier observed differences between candle condition groups. The interconnectedness of information visualization components is important to consider when designing data displays, but more generally, is critical for studying cognition as a whole. Carefully controlling the environment to test properties of visual cognition is critically important for inferring a causal relationship between stimulus and behaviour, but there is a risk that controlling the environment without also exploring the behaviour in adjacent contexts eliminates the meaningfulness of the research conducted.

In Experiment Three, where participants were asked to complete a time series pattern, there was a difference between the Condition A and Condition C. Participants in Condition C were more accurate in their forecasting predictions. However, in Experiment Four when one dimension was replaced with random values, there was no difference between groups. Introducing random noise to one dimension appears to have eliminated the performance advantage that came with the triangle candle bodies. One strong possibility is that the noisy dimension interrupted the participants' pattern perception of the overall Gestalt (Wertheimer, 1923, translated by Ellis 1938) of the time series.

The methodology of this thesis is well positioned to support a modern approach to Gestalt psychology research (Jäkel, Singh, Wichman & Herzog, 2016) in that different data visualizations are designed to help users see patterns and Gestalt researchers are interested in exactly that phenomenon (Brandes, Nick, Rockstrah & Stefen, 2013). Histograms would be effective stimuli to study the Gestalt law of symmetry, and scatter plots would be good stimuli to study the Gestalt law of proximity, for example. Candlestick plots could be effectively leveraged to test the perceptual boundaries of the law of continuity. While it would require a full series of studies to truly uncover the importance of glyph similarity to the overall perception of the candlestick time series, it is presently thought that the introduction of a noisy dimension disrupts the similarity Gestalt percept and so negates any advantage that may have come from the triangle candle body.

6.3. Information integration and rule-based task differences

The third overarching finding from these studies is that information integration and rule-based tasks differently impact the relationship between perception and cognition. Information integration is elicited more by Experiments Three, Four, Six and Seven

wherein a greater pattern is observed through component icons (Ashby & Maddox, 2005). In such task environments, combining features of the task stimuli offers insight into the data above and beyond querying any one of those features one their own. In contrast, Experiments One, Two and Five require participants to perform a simpler task of learning the connection between one component of the candlestick icon and then extracting one particular value from it. This task is more similar to traditional rule-based categorization problems, excepting for the continuous nature of the participant's response in the Experiments herein.

A rule based categorization task typically requires participants to develop an "if X then Y" sort of problem-solving strategy. A defining characteristic of rule-based tasks is that they can be easily described with language (Ashby & Maddox, 2005). For example, "if an animal has a beak and feathers, it is a bird". While the candlestick plot conveys continuous values, the strategy for learning which value is assigned to which stimulus dimension is rule-based. The easily-verbalizable rule for the candlestick glyph (Condition A) is "if it is green, then close is on top; if it is red, then close is on the bottom". The use of such rule is directly queried in the information extraction tasks (Experiments One, Two and Five) where participants are asked to report the value of a single dimension. Their job is to query the values represented by the glyph using the dimension-appropriate rule.

Manipulations of the candlestick glyph did not impact user performance error in Experiments One, Two and Five. Information extraction and utilization of the rule observed at the level of the participant's response were similar between the candle conditions. This is befitting of the logical nature of cue use in rule based tasks. So long as the perceptual qualities of the stimulus do not impede the observer from interpreting the values of the stimulus, those values may be used in a logical, abstract manner independent from their physical manifestation. Neurologically, rule-based decision making corresponds to activity in frontal-stratal circuitry (Ashby, Noble, Filoteo, Waldron, & Ell, 2003; Ashby, Maddox & Bohil, 2002; Seger & Cinotta, 2005; Schönberg, Daw, Joel & O'Doherty, 2007), implicating the basal ganglia and frontal lobe regions in making logical decisions. Similarities between candle conditions in the information extraction experiments is consistent with the logical use of information that is more robust to visual variation.

In contrast, however, the forecasting experiments (Experiments Three, Six) did exhibit differences between the candle conditions. In these experiments, the triangle

candle body outperformed the standard red/green rectangle candle glyph representation. As discussed above, the forecasting task in Experiments Three and Six is based on a richer stimulus set that exceeds the capacity for working memory to individually represent each icon, and the task itself is better performed using a general pattern percept. Combining complicated information in a hard-to-verbalize manner to make a decision is the mainstay of information-integration category learning research (Ashby & Maddox, 1990). Models that specialize in information-integration tasks (Ashby, Paul & Maddox, 2011) emphasize procedural learning as critical to performance. Procedural learning is often implicit and developed through practice more than it's developed through training. Neurologically, information-integration based decision making corresponds to increased activity in the putamen and caudate nucleus (Ashby, Ell & Waldron, 2003; Ashby & Maddox, 2010; Knowlton, Mangels, & Squire, 1996; Nomura & Reber, 2008). Since information is understood more procedurally in information-integration tasks, participants rely less on abstract rules and are more sensitive to how the information is actually presented. As with other types of category learning research, there is a relative lack of evidence about the physical properties of category dimensions and their influence on information-integration categorization types, but using data visualizations to study human cognition offers a great foundation to further manipulate icons and the physical properties of a display to query the relationship between visual stimuli and information integration.

6.4. The role of a science of data visualization in cognitive psychology

Data visualization as a domain nicely combines multiple levels of analysis in traditional cognition research including sensation and perception, attention and learning, concepts and categorization, and information-integration and decision-making. It is necessarily connected to the world beyond the human skull, too, in that it is simplifying external data for a human observer. Data visualizations change with some properties of the environment through externally measured values. In this respect, they serve as a predictable extension of human cognition into the real world. In Chapter Six, connections between Marr's information processing framework (1982) and data visualization were developed, following Rensink's extended-vision thesis (2014) wherein the human and the data visualization are modelled as a single information processor. While conceptualizing data visualization as an information processing problem is an important step to bring it

more formally in line with existing cognitive science (Fisher, Green, & Arias-Hernàndez, 2010), there had remained a problem of how to measure the output of the information processing system. In this thesis, by using a single outcome variable across three distinct types of tasks, there is some consistent measurement across multiple levels of analysis.

Perhaps most critically, though, this thesis offers a way to query information extraction, forecasting and decision making building upon a single type of representation in the candlestick plots. Decision making is ultimately the outcome of interest to data visualization designers. Offering a plot that best supports human observers in making the right data-driven decision is the whole point of developing a visualization in the first place. Decision making, then, must be part of a science of data visualization. Working from the simplest case (a single icon) toward the more complicated application of the candlestick plot to support decision-making is a strategy that could be used for any type of data visualization, many of which can then be used to test the application of theories of human cognition.

References

Aha, D. W., & Goldstone, R. L. (1992). Concept learning and flexible weighting. In Proceedings of the fourteenth annual conference of the Cognitive Science Society (pp. 534-539).

Armel, K. C., Beaumel, A., & Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. Judgment and Decision Making, 3(5), 396–403.

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. Memory & cognition, 30(5), 666-677.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. Annu. Rev. Psychol., 56, 149-178.

Ashby, F. G., & Maddox, W. T. (2010). Human category learning 2.0. Annals of the New York Academy of Sciences, 1224, 147–161. http://doi.org/10.1111/j.1749-6632.2010.05874.x

Ashby, F. G., & Alfonso-Reese, L. A. (1998). A neuropsychological theory of multiple systems in category learning. Psychological review, 105(3), 442.

Ashby, F. G., Ell, S. W., & Waldron, E. M. (2003). Procedural learning in perceptual categorization. Memory & Cognition, 31(7), 1114-1125.

Ashby, F. G., Noble, S., Filoteo, J. V., Waldron, E. M., & Ell, S. W. (2003). Category learning deficits in Parkinson's disease. Neuropsychology, 17(1), 115.

Ashby, F. G., & Paul, E. J. (2011). COVIS. In E. M. Pothos & A. J. Wills (Eds.), Formal approaches in categorization (65-87). Cambridge: Cambridge University Press.

Baluch, F., & Itti, L. (2011). Mechanisms of top-down attention. Trends in Neurosciences, 34(4), 224–210.

Barnes, J.I., Blair, M.R., Tupper, P. & Walshe, R.C. (under review). Tempus: An active, neural and temporal model of category learning and attention.

Barnes, J.I., Blair, M.R., Tupper, P. & Walshe, R.C. (2015). A dynamic neural field model of self- regulated eye movements during category learning. In Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., & Maglio, P. P. (Eds.), Proceedings of the 37th Annual Meeting of the Cognitive Science Society (pp. 148-153). Austin, TX: Cognitive Science Society.

Barnes, J.I., McColeman, C.M., Stepanova, E., Blair, M.R. & Walshe, R.C. (2014). RLAttn: An actor-critic model of eye movements during category learning. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), Proceedings of the 36th Annual Conference of the Cognitive Science Society (pp. 1892-1897). Austin, TX: Cognitive Science Society.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language, 68(3), 255-278.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Mixed-effects modeling with R. Journal of Statistical Software, 67, 1–48.

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. Academy of management review, 3(3), 439-449.

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. Econometrica: Journal of the Econometric Society, 23-36.

Bertin, J. (1983). Semiology of graphics: diagrams, networks, maps. Madison, WC: The University of Wisconsin Press.

Bischof, W. F., & Di Lollo, V. (1990). Perception of directional sampled motion in relation to displacement and spatial frequency: Evidence for a unitary motion system. Vision Research, 30(9), 1341-1362.

Blair, M. R., Watson, M. R., & Meier, K. M. (2009). Errors, efficiency, and the interplay between attention and category learning. Cognition, 112(2), 330–6.

Blair, M. R., Watson, M. R., Walshe, R. C., & Maj, F. (2009). Extremely selective attention: eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. Journal of Experimental Psychology. Learning, Memory, and Cognition, 35(5), 1196–206.

Boguski, M. S., Arnaout, R., & Hill, C. (2009). Customized care 2020: how medical sequencing and network biology will enable personalized medicine. F1000 biology reports, 1, 73.

Bott, L., Hoffman, A. B., & Murphy, G. L. (2007). Blocking in category learning. Journal of Experimental Psychology: General, 136(4), 685.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. Spatial vision, 10, 433-436.

Brandes, U., Nick, B., Rockstroh, B., & Steffen, A. (2013, June). Gestaltlines. In Computer Graphics Forum, 32(3), 171-180. Blackwell Publishing Ltd.

Broadbent, D. E. (1957). A mechanical model for human attention and immediate memory. Psychological review, 64(3), 205.

Brumagim, A. L., & Xianhua, W. (2005). An examination of cross-cultural differences in attitudes towards risk: testing prospect theory in the people's Republic of China. Multinational Business Review, 13(3), 67-86.

Brynjolfsson, E., & McElheran, K. (2016). Data-Driven Decision Making in U.S. Manufacturing. Center for Economic Studies. Retrieved from http://ssrn.com/abstract=2722502

Byrne, R. M., & Johnson-Laird, P. N. (1989). Spatial reasoning. Journal of memory and language, 28(5), 564-575

Cadenas, E. S., Cardenas, D., Delgado, G. S., & Perales, J. (2015). The hidden cost of coaching: Intentional training of shot adequacy discrimination in basketball hampers utilization of informative incidental cues. Perceptual and Motor Skills, 120(1), 139–58.

Callaghan, T. C. (1989). Interference and dominance in texture segregation: Hue, geometric form, and line orientation. Attention, Perception, & Psychophysics, 46(4), 299-311.

Callaghan, T. C. (1990). Interference and dominance in texture segregation. In Visual Search, D. Brogan, Ed. Taylor & Francis: New York, New York. pp. 81–87

Caputo, G., & Guerra, S. (1998). Attentional Selection by Distractor Suppression. Vision Research, 38(5), 669–689.

Carter, L. F. (1947). An experiment on the design of tables and graphs used for presenting numerical data. Journal of Applied Psychology, 31(6), 640.

Cleveland, W. S., & McGill, R. (1984). Graphical Methods Graphical Perception : Theory , Experimentation , and Application to the Development of Graphical Methods. Journal of the American Statistical Association, 79(387), 531–554.

Corso, P., Hammitt, J., & Graham, J. D. (2001). Increased information leakage from text. The Journal of Risk and Uncertainty, 23(2), 165–183.

Couclelis, H., Golledge, R. G., Gale, N., & Tobler, W. (1987). Exploring the anchor-point hypothesis of spatial cognition. Journal of Environmental Psychology, 7(2), 99-122.

Crapo, A. W., Waisel, L. B., Wallace, W. A., & Willemain, T. R. (2000, August). Visualization and the process of modeling: a cognitive-theoretic view. In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 218-226). ACM.

Croxon, F. E. (1932). Graphic Comparisons by Bars, Squares, Circles, and Cubes. Journal of the American Statistical Association, 27(177), 54–60.

Dacke, M., Nilsson, D. E., Scholtz, C. H., Byrne, M., & Warrant, E. J. (2003). Animal behaviour: insect orientation to polarized moonlight. Nature, 424(6944), 33-33.

Dalkey, N., & Helmer, O. (1963). An Experimental Application of the Delphi Method to the Use of Experts. Management Science, 9(3), 458–467.

Ellis, W. (1938). Gestalt Theory. A source book for Gestalt Psychology.

Enns, J. T., & Rensink, R. A. (1990). Sensitivity to three-dimensional orientation in visual search. Psychological Science, 1(5), 323-326.

Few, S. Perceptual Edge. (2005). Effectively Communicating Numbers: Selecting the Best Means and Manner of Display.

Fisher, B., Green, T. M., & Arias-Hernández, R. (2011). Visual analytics as a translational cognitive science. Topics in Cognitive Science, 3(3), 609–625. http://doi.org/10.1111/j.1756-8765.2011.01148.x

Gaspar, J. M., & McDonald, J. J. (2014). Suppression of salient objects prevents distraction in visual search. Journal of neuroscience, 34(16), 5658-5666.

Gaspar, J. M., Christie, G. J., Prime, D. J., Jolicœur, P., & McDonald, J. J. (2016). Inability to suppress salient distractors predicts low visual working memory capacity. Proceedings of the National Academy of Sciences, 113(13), 3693-3698.

Gigerenzer, G., Hertwig, R., Van Den Broek, E., Fasolo, B., & Katsikopoulos, K. V. (2005). “A 30% chance of rain tomorrow”: how does the public understand probabilistic weather forecasts? Risk Analysis an Official Publication of the Society for Risk Analysis, 25(3), 623–629.

Gluck, M. A, & Bower, G. H. (1988). From conditioning to category learning: an adaptive network model. Journal of Experimental Psychology. General, 117(3), 227–47.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. Annual Review of Neuroscience, 30, 535–74.

Gorman, A. D., Abernethy, B., & Farrow, D. (2011). Investigating the anticipatory nature of pattern perception in sport. Memory & Cognition, 39, 894–901.

Gottlieb, J. (2007). From Thought to Action: The Parietal Cortex as a Bridge between Perception, Action, and Cognition. Neuron, 53(1), 9–16.

Marie, T., Ribarsky, W., Green, T. M., & Visualization, C. (2009). Building and applying a human cognition model for visual analytics. Information Visualization, 8(1), 1–13.

Greensmith, T. M. (2016). The Personal Equation of Interaction for Categorization of Composite Glyphs. In IEEE Hawaii International Conference on System Sciences (1456-1465).

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: a meta-analysis. Psychological assessment, 12(1), 19.

Hammer, R., Sloutsky, V., & Grill-Spector, K. (2012). The interplay between feature-saliency and feedback information in visual category learning tasks. Proceedings of the 34th Annual Conference of the Cognitive Science Society, 420–425.

Hammer, R. (2015). Impact of feature saliency on visual category learning. Frontiers in psychology, 6.

Hammer, R., Sloutsky, V., & Grill-Spector, K. (2015). Feature saliency and feedback information interactively impact visual category learning. Frontiers in psychology, 6.

Healey, C. G., Booth, K. S., & Enns, J. T. (1995). Visualizing real-time multivariate data using preattentive processing. ACM Transactions on Modeling and Computer Simulation, 5(3), 190–221.

Healey, C. G., & Enns, J. T. (2011). Attention and Visual Memory. IEEE Transactions on Visualization and Computer Graphics.

Heer, J., & Bostock, M. (2010). Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. Proceedings of the 28th Annual Chi Conference on Human Factors in Computing Systems, 203–212.

Hegarty, M. (2011). The Cognitive Science of Visual-Spatial Displays : Implications for Design, 3, 446–474.

Hoffman, D. D., & Singh, M. (1997). Salience of visual parts. Cognition, 63(1), 29-78.

Humphrey, K., & Underwood, G. (2009). Domain knowledge moderates the influence of visual saliency in scene recognition. British Journal of Psychology, 100, 377–398. http://doi.org/10.1348/000712608X344780

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11), 1254–1259.

Jäkel, F., Singh, M., Wichmann, F. A., & Herzog, M. H. (2016). An overview of quantitative approaches in Gestalt perception. Vision research, 126, 3-8.

Jarvenpaa, S. L. (1990). Graphic displays in decision making - the visual salience effect. Journal of Behavioral Decision Making, 3(4), 247–262. http://doi.org/10.1002/bdm.3960030403

Jetzek, T., Avital, M., & Bjorn-Andersen, N. (2014). Data-driven innovation through open government data. Journal of theoretical and applied electronic commerce research, 9(2), 100-120.

Johnston, J. C., & Pashler, H. (1990). Close binding of identity and location in visual feature perception. Journal of Experimental Psychology: Human Perception and Performance, 16(4), 843.

Johnson-Laird, P. N. (1983). Mental models: Towards a cognitive science of language, inference, and consciousness (No. 6). Harvard University Press.

Julesz, B., & Bergen, J. R. (1983). Human factors and behavioral science: Textons, the fundamental elements in preattentive vision and perception of textures. The Bell system technical journal, 62(6), 1619-1645.

Julesz, B. (1984). Toward an axiomatic theory of preattentive vision. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), Dynamic aspects of neocorticalfunction (pp. 585-612). Neurosciences Research Foundation.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An analysis of Decision under Risk. Econometrica, 47(2), 263–291. http://doi.org/10.1017/CBO9781107415324.004

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. in B.A. Campbell & R.M. Church (eds.), Punishment and aversive behavior (pp. 279-296). New York: Appleton-Century-Crofts.

Kingstone, A., Smilek, D., Ristic, J., Friesen, C. K., & Eastwood, J. D. (2003). Attention, Researchers! It Is Time to Take a Look at the Real World. Current Directions in Psychological Science, 12(5), 176–180.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. Perception, 36(14), 1.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. Science, 273(5280), 1399.

Koide, N., Kubo, T., Nishida, S., Shibata, T., & Ikeda, K. (2015). Art Expertise Reduces Influence of Visual Salience on Fixation in Viewing Abstract-Paintings. Plos One, 10, e0117696.

Kosslyn, S. M., Chabris, C. F., Marsolek, C. J., & Koenig, O. (1992). Categorical versus coordinate spatial relations: computational analyses and computer simulations. Journal of Experimental Psychology: Human Perception and Performance, 18(2), 562.

Kosslyn, S. M. (1987). Seeing and imagining in the cerebral hemispheres: A computational approach. Psychological review, 94(2), 148.

Kosslyn, S. M. (2006). Graph design for the eye and mind. New York, NY: Oxford University Press.

Kruger, J., Blanck, H. M., & Gillespie, C. (2006). Dietary and physical activity behaviors among adults successful at weight loss maintenance. International Journal of Behavioral Nutrition and Physical Activity, 3(17), 1–10. http://doi.org/doi:10.1186/1479-5868-3-17

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. Journal of Experimental Psychology. Learning, Memory, and Cognition, 25(5), 1083–119. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10505339

Kruschke, J. K. (1992). ALCOVE: An examplar-based connectionist model of category learning. Psychological Review, 99(1), 22–44.

Lam, H., Bertini, E., Isenberg, P., Plaisant, C., & Carpendale, S. (2012). Empirical Studies in Information Visualization: Seven Scenarios. Visualization and Computer Graphics, IEEE Transactions on, 18(9), 1520–1536. http://doi.org/10.1109/tvcg.2011.279

Lawrence, M. J., Edmunson, R. H., & O'Connor, M. J. (1985). An examination of the accuracy of judgemental extrapolation of time series. International Journal of Forecasting, 1, 25–35.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

Lewandowsky, S., & Spence, I. (1989). The perception of statistical graphs. Sociological Methods and Research, 18(2–3), 200.

Lohse, J. (1991). A cognitive model for the perception and understanding of graphs. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Reaching through Technology CHI 91, 137–144. http://doi.org/10.1145/108844.108865

Louie, K. (2013). Integrating salience and value in decision making. Proceedings of the National Academy of Sciences, 110(40), 15853-15854.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: a network model of category learning. Psychological Review, 111(2), 309–32. http://doi.org/10.1037/0033-295X.111.2.309

Lurie, N. H., & Mason, C. H. (2007). Visual Representation : Implications for Decision Making. Journal of Marketing, 71(1), 160–177.

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. Behavioural Processes, 66(3), 309-332.

Marr, D. (1982). Vision: A computational approach. San Francisco, CA: W. H. Freeman

McColeman, C. M., & Blair, M. R. (2015). During category learning , top-down and bottom-up processes battle for control of the eyes. Proceedings of the 36th Annual Conference of the Cognitive Science Society, 1559–1564.

McColeman, C. M. (2012). The Influence of Salience and Similarity on Selective Attention. (Unpublished master's thesis). Simon Fraser University, Burnaby, Canada.

McColeman, C. M., Barnes, J. I., Chen, L., Meier, K. M., Walshe, R. C., & Blair, M. R. (2014). Learning-induced changes in attentional allocation during categorization: A sizable catalog of attention change as measured by eye movements. PLoS ONE, 9(1). http://doi.org/10.1371/journal.pone.0083302

Mikels, J. A, Maglio, S. J., Reed, A. E., & Kaplowitz, L. J. (2011). Should I go with my gut? Investigating the benefits of emotion-focused decision making. Emotion (Washington, D.C.), 11(4), 743–753. http://doi.org/10.1037/a0023986

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological review, 63(2), 81.

Mirman, D. (2014, February 11). Three ways to get parameter-specific p-values from lmer. Retrieved from http://lme4.0.r-forge.r-project.org/lMMwR/lrgprt.pdf

Modarres, M., Mirsamadi, M., & Peyman, G. A. (1996). Prevalence of congenital color deficiencies in secondary-school students in Tehran. International ophthalmology, 20(4), 221-222.

Munoz, D. P., Broughton, J. R., Goldring, J. E., & Armstrong, I. T. (1998). Age-related performance of human subjects on saccadic eye movement tasks. Experimental brain research, 121(4), 391-400.

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. Vision research, 45(2), 205-231.

Neisser, U. (1967). Cognitive Psychology. New York: Appleton.

Simon, H. A., & Newell, A. (1972). Human problem solving: The state of the theory in 1970.

Nison, S. (1994). Beyond candlesticks: New Japanese charting techniques revealed (Vol. 56). John Wiley & Sons.

Nomura, E. M., & Reber, P. J. (2012). Combining Computational Modeling and Neuroimaging to Examine Multiple Category Learning Systems in the Brain. Brain Sciences, 2(2), 176–202. http://doi.org/10.3390/brainsci2020176

Nomura, E. M., Maddox, W. T., Filoteo, J. V, Ing, a D., Gitelman, D. R., Parrish, T. B., Mesulam, M. & Reber, P. J. (2007). Neural correlates of rule-based and information-integration visual category learning. Cerebral Cortex (New York, N.Y. : 1991), 17(1), 37–43. http://doi.org/10.1093/cercor/bhj122

Park, V., & Datnow, A. (2009). Co-constructing distributed leadership: District and school connections in data-driven decision-making. School leadership and Management, 29(5), 477-494.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. Vision research, 42(1), 107-123.

Pashler, H. (1988). Cross-dimensional interaction and texture segregation. Attention, Perception, & Psychophysics, 43(4), 307-318.

Patterson, R. E., Blaha, L. M., Grinstein, G. G., Liggett, K. K., Kaveney, D. E., Sheldon, K. C., … Moore, J. A. (2014). A human cognition framework for information visualization. Computers & Graphics, 42, 42–58. http://doi.org/10.1016/j.cag.2014.03.002

Payne, J. W. (1982). Contingent decision behavior. Psychological bulletin, 92(2), 382.

Pinker, S. (1990). Theory of Graph Comprehension. In R. O. Freedle (Ed.), Artificial intelligence and the future of testing (pp. 73–123). Hillsdale, N.J: L. Erlbaum Associates.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. Big Data, 1(1), 51-59.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. The Behavioral and Brain Sciences, 22(3), 341-65-423. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11301517

Rabin, M. (2000). Risk Aversion and Expected-Utility Theory: A Calibration Theorem. Econometrica: Journal of the Econometric Society, 68(5), 1281–1292. http://doi.org/10.1111/1468-0262.00158

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. Cognitive Psychology, 51(1), 1–41. http://doi.org/10.1016/j.cogpsych.2004.11.001

Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert chess players: Evidence from eye movements. Psychological Science, 12(1), 48–55.

Rensink, R. A. (2006). Attention, Consciousness, and Data Display. Proceedings of the American Statistical Association, Statistical Graphics Section.

Rensink, R. A. (2014). On the prospects for a science of visualization. In Handbook of Human Centric Visualization (pp. 147–175). http://doi.org/10.1007/978-1-4614-7485-2

Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: Application of a theory. Inhibition and learning, 301-336.

Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. Psychonomic Bulletin & Review, 9(3), 507–513. http://doi.org/10.3758/BF03196306

Rosenbaum, D. A. (1975). Perception and extrapolation of velocity and acceleration. Journal of Experimental Psychology. Human Perception and Performance, 1(4), 395–403. http://doi.org/10.1037/0096-1523.1.4.395

Ruotolo, F., van Der Ham, I. J., Iachini, T., & Postma, A. (2011). The relationship between allocentric and egocentric frames of reference and categorical and coordinate spatial information processing. The Quarterly Journal of Experimental Psychology, 64(6), 1138-1156.

Sadalla, E. K., Burroughs, W. J., & Staplin, L. J. (1980). Reference Points in Spatial Cognition. Journal of Expeirmental Psychology: Human Learning and Memory, 6(5), 516–528.

Shneiderman, B. (2016). The new ABCs of research: Achieving breakthrough collaborations. Oxford University Press.

Schönberg, T., Daw, N. D., Joel, D., & O'Doherty, J. P. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. Journal of Neuroscience, 27(47), 12860-12867.

Shultz, K. S. & Whitney, D. J. (2005). Measurement in Action. Thousand Oaks, California: Sage Publications

Schütz, A. C., Trommershäuser, J., & Gegenfurtner, K. R. (2012). Dynamic integration of information about salience and value for saccadic eye movements. Proceedings of the National Academy of Sciences, 109(19), 7547-7552.

Seger, C. A., & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. Journal of Neuroscience, 25(11), 2941-2951.

Shepard, R., Hovland, C., & Jenkins, H. (1961). Learning and memorization of classifications. Psychological Monographs, 75(13).

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. Psychological review, 81(3), 214.

Simkin, D., & Hastie, R. (1987). An Information-Processing Analysis of Graph Perception. Source Journal of the American Statistical Association, 82(398), 454–465. http://doi.org/10.1080/01621459.1987.10478448

Sargent, T. J., & Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. New methods in business cycle research, 1, 145-168.

Sims, C. A. (1980). Macroeconomics and reality. Econometrica: Journal of the Econometric Society, 1-48.

Spence, I., & Lewandowsky, S. (1991). Displaying proportions and percentages. Applied Cognitive Psychology, 5(1), 61-77.

Spence, I. (1990). Visual Psychophysics of Simple Graphical Elements. Journal of Experimental Psychology: Human Perception and Performance, 16(4), 683–692. http://doi.org/10.1037//0096-1523.16.4.683

Stone, M., & Bartram, L. (2010). Whisper , Don't Scream: Characterizing Subtle Grids. IEEE Transactions on Visualization and Computer Graphics, 17(10), 1444–1458. http://doi.org/10.1109/TVCG.2010.237

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: reinterpreting salience. Journal of Vision, 11, 5. http://doi.org/10.1167/11.5.5

Thaler, R. H. (1980). Toward a positive theory of consumer choice. Journal of Economic Behavior & Organization, 1, 39–60. http://doi.org/10.1016/0167-2681(80)90051-7

Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. Attention, Perception, & Psychophysics, 50(2), 184-193.

Theeuwes, J. (1992). Perceptual selectivity for color and form. Attention, Perception, & Psychophysics, 51(6), 599-606.

Theeuwes, J. (2004). Top-down search strategies cannot override attentional capture. Psychonomic bulletin & review, 11(1), 65-70.

Theeuwes, J. (2010). Top–down and bottom–up control of visual selection. Acta psychologica, 135(2), 77-99.

Thompson J.J., Blair M.R., Chen L., & Henrey A.J. (2013) Video Game Telemetry as a Critical Tool in the Study of Complex Skill Learning. PLoS ONE 8(9): e75129

Tipples, J. (2002). Eye gaze is not unique: Automatic orienting in response to uninformative arrows. Psychonomic Bulletin & Review, 9(2), 314–318. http://doi.org/10.3758/BF03196287

Towal, R. B., Mormann, M., & Koch, C. (2013). Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. Proceedings of the National Academy of Sciences, 110(40), E3858-E3867.

Tranel, D., & Kemmerer, D. (2004). Neuroanatomical correlates of locative prepositions. Cognitive Neuropsychology, 21(7), 719-749.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. Cognitive psychology, 12(1), 97-136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. Psychological review, 95(1), 15.

Tufte, E. R. (1983). The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, 211(4481), 453–458.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and uncertainty, 5(4), 297-323.

van der Ham, I. J., & Borst, G. (2011). Individual differences in spatial relation processing: effects of strategy, ability, and gender. Brain and cognition, 76(1), 184-190.

van der Ham, I. J., van Wezel, R. J., Oleksiak, A., & Postma, A. (2007). The time course of hemispheric differences in categorical and coordinate spatial processing. Neuropsychologia, 45(11), 2492-2498.

Port, R. F., & Van Gelder, T. (Eds.). (1995). Mind as motion: Explorations in the dynamics of cognition. MIT press.

van Rullen, R., & Thorpe, S. J. (2001). The Time Course of Visual Processing: From Early Perception to Decision-Making. Journal of Cognitive Neuroscience, 13(4), 454–461. http://doi.org/10.1162/08989290152001880

van Wijk, J. J. (2005, October). The value of visualization. In Visualization, 2005. VIS 05. IEEE (pp. 79-86). IEEE.

van Zoest, W., & Donk, M. (2007). The effects of salience on saccadic target selection. Visual Cognition, 12(2), 353–375. http://doi.org/10.1080/13506280444000229

van Zoest, W., Donk, M., & Theeuwes, J. (2004). The Role of Stimulus-Driven and Goal-Driven Control in Saccadic Visual Selection. Journal of Experimental Psychology: Human Perception and Psychophysics, 30(4), 746–759. http://doi.org/10.1037/0096-1523.30.4.746

van Zoest, W., Hunt, A. R., & Kingstone, A. (2010). Representations in visual cognition: It's about time. Current Directions in Psychological Science, 19(2), 116-120.

Vessey, I., & Galletta, D. (1991). Cognitive fit: An empirical study of information acquisition. Information systems research, 2(1), 63-84.

Vessey, I. (1994). The effect of information presentation on decision making : A cost-benefit analysis. Information and Management, 27, 103–119.

Von Neumann, J., & Morgenstern, O. (1945). Theory of games and economic behavior. Princeton, NJ: Princeton University Press.

Waisel, A. W. C. L. B., Wallace, W. A., & Willemain, T. F. (2000). Visualization and The Process of Modeling: A Cognitive-theoretic View. In KDD: Proceedings (No. 6, p. 218). AAAI Press.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: evidence for multiple category learning systems. Psychonomic Bulletin & Review, 8(1), 168–76. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11340863

Walker, R., Walker, D. G., Husain, M., & Kennard, C. (2000). Control of voluntary and reflexive saccades. Experimental Brain Research, 130(4).

Ware, C. (2004). Information visualization: perception for design. Waltham, MA: Elsevier.

Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond Bar and Line Graphs : Time for a New Data Presentation Paradigm, 1–10. http://doi.org/10.1371/journal.pbio.1002128

Wertheimer, M. (1923). Laws of organization in perceptual forms. A source book of Gestalt Psychology.

Wever, E. G., & Zener, K. E. (1928). The method of absolute judgment in psychophysics. Psychological Review, 35, 466–493. http://doi.org/10.1037/h0075311

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. Acta psychologica, 41(1), 67-85.

Wolfe, J. M. (2001). Guided Search 4.0: A guided search model that does not require memory for rejected distractors. Journal of Vision, 1(3), 349-349.

Woltman, H., Feldstain, A., Mackay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. Tutorials in Quantitative Methods for Psychology, 8(1), 52–69.

Xu, Y., & Franconeri, S. L. (2015). Capacity for visual features in mental rotation. Psychological science, 26(8), 1241-1251.

Xu-Wilson, M., Zee, D. S., & Shadmehr, R. (2009). The intrinsic value of visual information affects saccade velocities. Experimental Brain Research, 196(4), 475-481.

Yarbus, A. (1967). Eye Movements and Vision. New York: Plenum Press.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. American Psychologist, 35(2), 151–175. http://doi.org/10.1634/theoncologist.9-90005-10

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. Memory & cognition, 34(2), 387-398.

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. NeuroImage, 32(4), 1826–1836. http://doi.org/10.1016/j.neuroimage.2006.04.199