INTERFERENCE IN THE PERCEPTION OF CORRELATION FOR TWO-POPULATION SCATTERPLOTS

by

Madison Ann Elliott

M.A., Towson University, 2013

B.A., Roanoke College, 2011

A thesis submitted in partial fulfillment of the requirements for the degree of
Master of Arts
in
The Faculty of Graduate and Postdoctoral Studies

(Psychology)

The University of British Columbia

(Vancouver)

August 2016

© Madison Ann Elliott, 2016

Abstract

Rensink & Baldridge (2010) first operationalized a methodology to study perceptual performance for single population scatterplots. Rensink (2014) emphasized that this methodology could be extended to understand the perception of more complex displays. In this paper, the original methodology is extended to examine the perception of two-population scatterplots, containing target and distractor populations. In three experiments, we investigated the nature of performance for a discrimination task used to measure viewer precision (defined by parameter k in our analysis) and accuracy (defined by parameter b in our analysis). The results show that perception for two-population scatterplots is non-trivially different from the perception of single population scatterplots. Namely, there is a significant degree of interference for selecting and discriminating the target populations in each display. This interference occurs due when the target and distractor populations are featurally distinct from one another, violating assumptions from Feature Integration Theory and Guided Search. Discrimination performance also degrades as a function of the density of the distracting population. Findings from this work not only help solidify a methodology to study the mechanisms underlying ensemble selection and inhibition in the presence of multiple populations, but also guide design choices for the creation of more complex scatterplot displays.

Preface

All research described in this report was conducted through the Department of Psychology at the University of British Columbia under the supervision of Dr. Ronald A. Rensink. I was responsible for programming the stimuli and experimental procedures in the VCL Framework, a Java software library that I helped develop with other members of the lab. I was also responsible for conducting or supervising data collection, for conducting all data analysis, and for writing this manuscript.

Data from Chapter 1 were presented as a poster at the Vision Sciences Society's 2015 Annual Meeting.

Data from Chapter 2 were presented as a poster the Vision Sciences Society's 2016 Annual Meeting.

Data from Chapters 1, 2, and 3 were presented as a talk at the 2016 International Conference on Thinking.

This research ("The Perception of Correlation in Scatterplots") was approved by the UBC Behavioural Research Ethics Boards (certificate number H10-01207).

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to sincerely thank my supervisor, Dr. Ronald Rensink, for his unwavering support and brilliant advice during my MA degree. Dr. Rensink encouraged me to explore my data creatively, to think critically and innovatively about research in general, and to work with other members of the lab to achieve a novel, cohesive project.

I would also like to thank my committee members. Dr. Darko Odic, for his friendship, and for providing constant inspiration and fresh perspectives throughout the phases of this project. Also Dr. Jiaying Zhao, for her helpful advice and research inquiries.

I would like to thank the participants who performed each of the experiments in this paper.

A massive thanks to the members of the Visual Cognition Lab. First and foremost, thank you to the members of the Coding Team, for devoting every other Saturday to the tireless development of our experimental software application framework. I would like to offer my deepest appreciation for the work that Yana Pertels, Daniel Chen, Amon Ge, Hubert Ngu, Taj Darra, Mario Cimet, and Paulo Apolinar put into this project. Thanks to my Research Assistants: Ida Mattson, Hilary Bate, and Andy Wu, who helped with data collection, cleaning, and analysis. Finally, thank you to Lisa Shiozaki, who provided endless insight into the practice of data visualization, and with whom most of this project was discussed and theorized.

1 INTRODUCTION

1.1 Motivation

Rensink and Baldridge (2010) developed a rigorous methodology to study the perception of correlation in single population scatterplot displays. Their work drew from long-established experimental designs and procedures from vision science. The results yielded two measures: one for accuracy (b), and another for precision (k). The authors' goals were not only to design better, more effective scatterplot displays, but also to have a means of evaluating observer low level visual performance on their tasks. This pattern of thinking was extended in Rensink (2014), where it was suggested that the foundational work should be extended to more features and correlation displays. Indeed, it appeared that viewer performance was invariant to many design manipulations (Rensink, 2016). Beyond the design evaluation implications, work extending this methodology can offer new insight into how populations of data are selected or inhibited in attention, during a relevant behavioral task.

In these 3 experiments, we examine perceptual behavior for two-population scatterplots. The goal of this work was to extend and operationalize the original methods for this type of stimuli, as well as to investigate the perceptual processes and mechanisms underlying observer performance. In order to understand what implications exist in this realm, it is important to outline the pertinent research topics from vision science that are potentially involved with viewing scatterplots.

1.2 Connections with Visual Perception

1.2.1 Ensemble perception

Humans are capable of extracting a diverse range of information from groups of objects in visual scenes. This includes identifying, or segmenting targets from a background (Wolfe et al., 2002), understanding spatial relationships of objects in a scene (Franconeri et al., 2012; Zhao et al., 2013; Alvarez & Oliva, 2008), estimating numerosity (Dehaene et al., 2008; Ross & Burr, 2010), and averaging, or getting summary information from a set (Ariely, 2001; Haberman & Whitney, 2007). In order to form a perceptual understanding of complicated information, the visual system compresses environmental structure to aid accurate and efficient representations of observed stimuli. Rensink (2016) suggested that scatterplots might be perceived via this type of coding.

Previously, it was thought that objects needed to be serially attended to in order to make sense of a scene (Egeth & Yantis, 1997). However, research on the rapid perception of gist showed that in fact, explicit selection is not needed (Oliva & Torralba, 2006). Early research on ensembles, or populations defined by scene statistics, was motivated by findings showing that the visual system does an excellent job of capitalizing on regularities within scenes. Humans can rapidly glean a gist understanding, or simple categorization at a basic level (Larson & Loschky, 2009) of novel, and even blurry images (Potter, 1975, 1976; Potter et al., 2004; Schyns & Oliva, 1994). In these studies, participants could identify the semantic category of the entire scene, as well as attribute information for various objects in the scene as quickly as 100 milliseconds (ms) after they were presented. This work was seminal to the reconceptualization of how object selection in attention contributes to the holistic perception, or in other words, simple categorical representation of visual information in a scene.

More recently, ensemble research has focused on understanding what constitutes a scene statistic, and how these statistics are used in efficient encoding and representation of information. Based on the well-accepted information theory developed by Shannon (1948), it has been proposed that the visual system represents stimuli by calculating efficient codes based on the statistics of an image or scene ensemble. This coding is used to reduce both redundancy and noise without losing too much visual information (Simoncelli & Olshausen, 2001). It has been suggested that such codes can allow for sufficiently accurate representations, even when processing is capacity-limited (Alvarez & Oliva, 2009). Initially, researchers observed that groups of low-level visual features could serve as scene statistics for ensemble encoding. Features like size (Ariely, 2001; Chong & Treisman, 2003) and orientation (Parkes et al., 2001) were among the first to be identified. Ariely (2001) found that when participants viewed displays containing different numbers of spots in varying sizes, they could easily infer the mean size of each group of spots. Interestingly, despite showing remarkable accuracy for the average size in a set, viewers encoded very little information about the individual spots themselves (i.e., participants were unable to individuate members of a set based on size). Even when viewers were explicitly instructed to pay attention to individual spots in the display, results indicated that they encoded the scene summary. This work paved the way for recent ensemble research by uncovering two main points: 1) that visual encoding of size ensembles is a rapid, automatic process for sets of objects, and 2) that this summary representation is not reliant on the explicit selection or encoding of any information about individual objects' size (aside from the overall range in some cases). In summary, the representation of a set of objects does not appear to be a straightforward configuration of its individual items or objects.

Chong and Treisman (2003) showed that ensembles could be encoded, irrespective of the distribution viewed. In their results, average size summary thresholds were the same for comparisons between samples drawn from the same distribution and samples drawn from different distributions. In addition to this, the authors found that sample means could be very accurately computed for different spatial locations (i.e., in the right versus left visual fields). In three discrimination task experiments, Chong and Treisman (2005) took these findings a step further. Participants viewed various displays and had to decide which sample of circles in the displays had the higher mean size. These samples varied by color, numerosity and density, and in the third experiment a cueing paradigm was used to examine how automatically ensembles could be formed and encoded by viewers. This work aimed to determine whether or not participants could perform averaging summaries over displays with various structural constraints such as variations in density, and distractor samples. The results confirmed previous findings that viewers do not rely on information from individual objects to form summary conclusions about the scene, and also confirmed that the summary statistic used was the mean and not the median or mode size in the display. Mean size judgments were also found to be invariant to changes in density or numerosity in the samples. The most surprising finding from this paper was that discrimination for mean size of sets of different colors was still highly accurate, and regardless of whether the cue preceded or followed the circle displays, performance was the same. The authors compared thresholds for mean size judgments in this experiment with results from the third experiment in Chong and Treisman (2003), and found that these color discriminations yielded results just as accurate as the spatial location discriminations from Chong and Treisman (2003). This result was especially surprising with regards to feature integration models from Treisman and Gelade (1980), Treisman and Gormican (1988), and past research on binding from Treisman (1998), in the sense that participants did not appear to need to select and bind color or size to individual items to perform the necessary averaging computations. This suggested that the formation of ensembles in perception is a highly automatic and reliable process. Similarly, Parkes et al. (2001) showed that mean orientation is correctly identified for Gabor patches presented in the periphery, even when observers failed to individuate a central target patch. Again, this failure was not due to interference from object crowding near the target, or particular attentional deployment in the scene. Haberman and Whitney (2007, 2009) extended the conceptualization of scene statistics to higher-level stimuli in their studies, which showed that observers could precisely represent the mean emotion expressed over of a group of faces. All of the studies mentioned ruled out the possibility of serial processing and individuation, asserting that each of these summary behaviors is due to an explicit averaging process over the contents of the scene.

1.2.2 Crowding

Another perceptual phenomenon to consider in the context of ensemble statistics is visual crowding. This is defined as impairment in the recognition of objects or features presented away from the fovea due to the presence of neighboring objects. Crowding occurs when distracting objects or features closely flank a target object or feature. Older work on crowding suggested that it is normally damaging to holistic scene perception, because crowding is detrimental to the encoding of individual object representations. This is the case because crowding is disruptive to serial attentional selection (Kahneman et al., 1983; Stockdale, 1978). However, more recent findings suggest that object crowding may actually contribute to a visual "chunking" process that better enables ensemble representations (Haberman & Whitney, 2012). Research from Intriligator and Cavanagh (2001) suggested that crowded displays caused viewers to engage in summary statistical strategies to process information. Attentional selection of a specific target and/or specific distractor objects does not occur in this process. Nevertheless, these findings only addressed crowding by similar objects.

As the study of crowding progressed, many of its nuances and definitions became better defined. For instance, a review by Whitney and Levi (2011) outlined several important advances in understanding how crowding both limits and increases spatial attentional selection. One important aspect of defining crowding is the degree of physical space between stimuli. Bouma's "Law" (1970), which states that "at a given point and direction in the visual field, critical spacing, measured from the center of a target object to the center of a similar flanking object, is the same for all objects" (Rosen et al., 2014), helped clarify the combination of parameters necessary in critical spacing for determining targets and distractors in this context. This spacing varies depending on the similarity between target and flankers, the complexity of the scene, type of stimuli displayed, and their distance from the fovea. As mentioned above, in displays of crowded objects of similar types, feature information is pooled or averaged (Parkes et al., 2011), however if the nature of the flakers is dissimilar from the target, a "pop-out" effect will occur for the target and distractor information is lost in attention. In this case, attentional selection of the target cannot be inhibited. This occurs for a wide range of dissimilarities, including shape, size, orientation, polarity, spatial frequency, depth, color, motion, and order (Whitney & Levi, 2011).

What remains to be understood in this area of research is the effect of a small crowding occurrence (or multiple crowding occurrences) in a larger display. If crowding occurs in only a small region of the display, how exactly does it affect the overall averaging process? This is clearly dependent on whether information is lost due to the crowding (i.e., there is high dissimilarity between target and flankers; you could imagine this happening if a target-population dot is flanked by distractor-population dots in a two-population scatterplot). However, even if information is successfully pooled in a crowded region, does that pooling occur at the same time as the overall ensemble formation? Is this process hierarchical? Does it occur automatically, pre-attentively, or does it require more a more active cognitive process?

Despite the well-established set of behaviors for ensemble encoding, a substantial number of questions remain surrounding its underlying perceptual mechanisms (for a review, see Haberman & Whitney, 2012). For instance, it is unknown whether statistical representations are a serial or parallel process in attention. Additionally, it is unclear whether spatial or featural information takes precedence during the encoding of ensembles. In order to speculate further on these questions, it is important to outline attentional behaviors across other perceptual contexts. One of the most prominent examples of visual information summary in attention occurs during a phenomenon known as visual search.

1.2.3 Feature selection in visual search

Visual search is the act of identifying a target object among a set of distracting, or irrelevant stimuli. There has been an established connection between ensemble codes and visual search. In particular, controversies surrounding the nature and mechanisms underlying pop-out effects (Eckstein, 1998; Itti & Koch, 2000; Wolfe, 2003) may be addressed by understanding how the visual system capitalizes on its averaging abilities to detect dissimilarity. To further understand these discrepancies between perceptual successes and failures in attentional selection, it is necessary to examine the literature on illusory conjunctions and models of guided search.

Treisman and Schmidt (1982) hypothesized that diversions in attention, or too-high attentional loads would cause features to be incorrectly bound together in viewers' perceptual representations. In other words, viewers would confuse object features in the scene when selection was distracted or inhibited. These mistaken reports occur for brief stimulus presentations (~200ms), even when the items are featurally distinct (Prinzmetal et al., 1995). Cohen and Ivry (1989) posited that attentional processing requires at least 100-200ms to accurately define featural information and spatial positions for items spaced 1^{o} or less from each other in foveal vision, otherwise incorrect bindings are likely to occur. Items in the periphery are subject to binding errors at even broader spatial separations and for slower presentation times. This visual phenomenon is known as "illusory conjunction". The experimental stimuli were small (~3.5in x ~1in), hand-drawn letters or shapes, and numeric digits on cards. Depending on the experiment, the letters and shapes could vary in size, "solidity" (or fill), color, and spatial location. The digits were always black, but the letters or shapes could be any of five different colors. These cards were presented to the viewers in a tachistoscope, and participant responses were given verbally. All tasks were variations on a digit-recognition task (participants had to report the digit on the card) that was either followed by a working memory task, or occurred at the same time as an object recognition/visual search task (participants had to report whether a specific object was present in the scene). The major findings from this paper were 1) illusory conjunction, or a mis- combining of feature information occurred under high attentional load (although it is unclear whether the information was encoded improperly, or recombined improperly after encoding), 2) conjunction errors were not always conveyed via participants guessing due to a lack of information. Some conjunction errors were conveyed via confident reports of objects that were not actually in the display (e.g., reporting a pink X, when in fact there was only a pink O and a yellow X in the display), 3) illusory conjunctions did not occur due to problems with verbal encoding (or due to verbal encoding at all), 4) illusory conjunctions cannot be explained by failures in working memory recall, 5) physical distance/spatial location did not affect illusory conjunction, 6) features appeared to be coded independently, and illusory conjunctions were not related to true combinations of features within objects on the display. Put more simply, features were "exchanged" freely between objects when misbindings and illusory conjunction reports occurred, 7) the frequency of illusory conjunctions was not dependent on the frequency of errors related to individual features (i.e. mistaking one color for the other, or an X for an N). Although, illusory conjunctions occurred more frequently for color and shape, they were also present for solidity and size. In fact, the number of reported illusory conjunctions could increase as feature error reports decreased. This final result suggested that attention is actually used to recombine separately/independently acquired feature information. Much of this work conflicts with models that tout global or holistic visual scene processing. In summary, Treisman and Schmidt (1982) demonstrated that the whole could not be considered greater the sum of its parts; object features and spatial locations are often muddled and confused in attention.

Theeuwes (1992) showed that full top-down selectivity for a unique, salient target was not possible in visual search tasks when a salient distracting item, even with different features was present. For instance, in a search task for a green circle among green squares, when a red square appears as a distractor, observers' search times were significantly slower. This occurred even when the distractor was identified and rehearsed prior to the task. Later, work from Wolfe (1994) re-examined a number of visual search and feature integration findings. Through a series of simulation studies with target and distractor objects, a new attentional selection, or "guided search" model was developed to explain novel search behavior differences due to additional nuances in the displays. Under the guided search model, feature searches appeared to be rapid and "parallel", instead of serial, when displays contained high target/distractor discrepancies. However, the less featural difference between target and distractor, the less efficient the visual search performance. Additionally, if the targets were categorically unique enough, some feature similarities with distractors would not decrease search efficiency. Finally, efficiency of conjunctive search was better for more salient stimuli. It should be noted that the guided search model was not fully able to account for findings by Theeuwes (1992), leaving notably unresolved ambiguity in the feature selection literature.

These findings helped lay the behavioral groundwork for ensemble research, as new search performance variations begged an explanation for how target and distractor populations are encoded and grouped. One of the main advantages from Wolfe's (1994) guided search model was the consideration of a greater number of factors resulting in parallel versus serial processing of scenes. Still, this consideration yielded somewhat different conclusions from early feature integration models (Treisman & Gelade, 1980; Treisman & Schmidt, 1982) regarding the use of feature selection versus feature inhibition in attention. Under the guided search model, it appears that selection is indeed possible under high attentional load, so long as object features are salient or different enough.

Visual search models offer further explanations for selecting and inhibiting visual stimuli in a scene. The relationship between ensemble work and search is clear, since both processes require a mental summary of visual information. It seems logical to argue that scene statistics help people understand populations of distractors and targets, via featural and spatial information. Although behavioral findings have undoubtedly conveyed humans' limited attentional resources in search procedures (Pashler, 2016), summary mechanisms that aid distractor inhibition (or slow target identification) remain largely undefined and poorly understood. In this thesis, we build upon these conclusions and propose a new way to study task-driven ensemble perception.

1.3 Connections with Information Visualization

1.3.1 Information visualization

Before we describe the methods for our work, it is important to detail the other side of our research motivation. In addition to addressing traditional topics in vision science, our work contributes equally to the increasingly important, but poorly understood phenomenon of visualizing information and data. Humans process more of their environment through the visual system than by any other sensory means combined (Ware, 2012). Our visual cognition is capable of rapid, flexible pattern discovery, which is complimented by adaptive, automatic decision-making. Given this exquisite perceptual toolset, it should come as no surprise that we have developed complex, high-level strategies for synthesizing our environment in visual terms. Evidence of information visualization practices date back to the earliest records of human civilization ("How we did data visualization before computers", 2016). A very early example comes from ancient Quechuan societies in South America, who translated their base-ten number values into a system of categorized rope lengths and knots, called "talking knots" (a bit of a misnomer- too bad they didn't call them "seeing knots"). By physically representing abstract numerical information, these people could visually encode them, which facilitated communication and refined computational procedures.

Mankind has come a long way since talking knots, from the advent of geographic maps to describe and navigate our vast and un-seeable world, all the way to modern visualization software tools and packages like the Javascript D3 library, R Studio®, and Tableau®. Current technology allows us to translate massive loads of digital information stored in incomprehensible tables of data, into digestible graphic summaries, which now facilitate interactive, exploratory visual displays. Over the past decade, scientists have recognized the tremendous potential that visual interfaces provide us: we can utilize successful visualizations to characterize, discover, abstract, and even generate task scenarios and form hypothesis from our data (Munzner, 2009). The advantages of studying and improving information visualization are obvious, especially with the expansive prospects offered by big data collection (Keim et al., 2013). While contemporary tech companies like Google and Facebook are pushing the limits of computing and artificial intelligence with their user information, even classical sciences like physics are benefitting from previously impossible large-scale simulations of astronomic data. Today's society has become, without a doubt, data-driven; and information visualization is a vital component of this paradigm shift. But this realization begs an important question: how can we comprehensively understand visualizations in order to optimize them?

1.3.2 Towards a collaborative science of data visualization and evaluation

Rensink (2014) proposed the need for a scientific methodology, or framework, to characterize and evaluate visualizations. In his proposal, he points out that designers have long-explored the many possible dimensions of data display, and historically, such designers have had good intuitions about what makes certain visualizations "work". However, simply basing design decisions around gut instincts or aesthetic preferences of an expert group is not enough. That isn't to say these experts are wrong in their evaluations, but instead it is proposed that borrowing and refining methods from perceptual psychology and vision science can explain the reasons behind "good" visualizations' success, empirically. Just as crucially, these scientific methods can be used to understand multifaceted visualizations, to explore increasingly large and complex datasets.

As discussed previously, vision science experiments do not generally seek to explain complex and poorly defined perceptual experiences. Instead, vision science tends to focus on a well-defined instance of perception, such as selecting a target among distractors (i.e. visual search), and designing controlled experiments to toggle and evaluate relevant variable components of this experience (i.e. number of distractors or feature salience of the target; Wolfe et al., 2002). Rensink (2014) suggests that the same kind of paradigm can be applied to the realm of information visualization, and advocates the characterization and study of well-defined aspects of performance (i.e. viewer accuracy of observed Pearson's r correlation value) and components associated with data displays (i.e. the size of the axes or color of the dots in a scatterplot). Thoughtful development of this kind of paradigm could result in a research framework that is generalizable to a massive visualization and vision science audience.

Recently, the prospect of using vision science techniques to understand information visualization was strongly echoed by Albers-Szafir et al. (2016), in their survey of ensemble encoding in various types of data displays. Although the authors' proposal is much more focused on a two-way conversation between visualization experts and vision scientists: emphasizing that while vision science can help the evaluation and design of better displays, these displays can in turn motivate new directions in perceptual psychology research. For instance, ensemble encoding has previously been studied to understand how people can do things like estimate the number of books in a shelf, and how items (like books) are spatially encoded in vision. By the same principle, several researchers (see Albers-Szafir et al. 2016; Cleveland at al., 1982; Cleveland & McGill, 1986) have proposed that vision researchers should look to visualizations, which also frequently use meaningful, spatially distributed items in a display (i.e., dots in a scatterplot) to investigate how ensemble encoding works. In a similar vein, work from Harrison et al. (2014) showed that traditional vision science methodologies could be extended to empirically evaluate a number of visualizations. This bi-directionality of application feeding research, and vice versa, can be leveraged quite broadly to help fulfill some of the original aspirations from Rensink (2014): to clearly define perceptual experiences and isolate their fundamental variables of interest. Both of these papers offer promising approaches towards not only improving the nature of design in visualization, but also promoting a deeper understanding of perceptual mechanisms in the human visual attention. Using graphs as controlled visual stimuli yields experimental data with two clear outcomes: 1) an applied, prescriptive measure of design, and 2) evidence regarding how visual features in visualizations are perceived and understood.

1.3.3 Scatterplots as visual stimuli

Scatterplots have proven to be exemplary controlled visual stimuli. Past research on the perception of correlation largely focused on direct, numerical estimation tasks (for reviews, see Boynton, 2000; Doherty et al., 2007). In these experiments, participants were typically asked to simply report the magnitude of the correlation they perceived in the display. Much like direct estimation performance in approximate number research, these studies showed that viewers could rapidly extract correlation information, although viewers did typically underestimate physical correlation *r* (Bobko & Karren, 1979; Cleveland et al., 1982; Strahan & Hansen, 1978). Interestingly, viewer expertise did not seem to influence basic perceptual estimation performance, as results were largely independent of participants' familiarity with statistics (Meyer & Shinar, 1992; Meyer et al., 1997). Despite drawing important conclusions about how the estimation of correlation behaves, these studies had some limitations. The most striking issue with past work was that few studies have directly investigated the precision of scatterplot viewing behavior (for a review, see Rensink, 2016); instead, most research has focused solely on accuracy.

Rensink and Baldridge (2010) addressed this limitation in their two-part methodology. Their experiments used a controlled set of stimulus parameters: black dot clouds with 100 data points, each containing gaussian distributions with equal variance in both dimensions. They demonstrated robust and clean participant performance on two concrete scatterplot tasks: discrimination and direct estimation of Pearson's r correlation. These methods were simply extensions of classic vision science approaches to assessing how well stimulus properties can be discriminated or directly perceived and estimated. In discrimination tasks, a just noticeable difference (JND) can be calculated to show when viewers are able to discriminate two side-by-side stimuli (e.g. squares of differing brightness) 75% of the time. For many low-level visual properties, plotting JNDs shows a simple linear behavior described by Weber's Law: where if p denotes a physical property (again, e.g. brightness), JND(p) = dp = kp. Note that values of for Weber fraction (k) are generally between 0.02-0.08 (Coren et al., 2004). This method and JND calculation was easily applied to scatterplots and the discrimination of correlation values, generating results that measure viewer precision. Remarkably, JNDs for Pearson's *r* correlation in scatterplots are a precise representation of Weber's Law. Rensink & Baldridge (2010) showed that discrimination performance can be described simply: JNDs are proportional to the distance from $r = 1$. These results indicate that despite seeming complex, scatterplots are actually a very successful visualization of a simple property: correlation. Additionally, the perception of correlation behaves like the perception of other low-level visual quantities; further suggesting that correlation is in fact, likely to be a scene statistic or ensemble property (Rensink, 2016).

By using this behavioral foundation, we can take a "bottom-up" approach to studying the perception of correlation. Scatterplots are often stripped down to single population graphs, containing something like 100 black dots each. But once perceptual behavior is established, the effect of feature manipulations (e.g., changing the color or number of dots, or adding more than one data population) can be studied. This building-block approach is humbling, in so far as it highlights how challenging it will be to study complex, multi-dimensional visualizations. However, just as natural scene representations continue to be a major challenge for vision scientists to study, by accruing evidence about their components, we build a better understanding of the whole. The same logic applies to studying visualization displays: the more we understand about simple displays, the more we can learn about multifaceted visualizations.

1.4 Approach

In this thesis, we develop a new methodology to examine visual extraction of multiple correlation ensembles in a discrimination task. This work extends the approach of Rensink and Baldridge (2010) and Rensink (2016), to examine perceptual behavior for two populations of data. To our knowledge, this work is the first investigation of the visual system's ability to select and inhibit ensembles simultaneously. Experiment 1 examines perceptual interference in discrimination performance for "simple" two-population scatterplots, using color (populations of red dots and black dots) as the featural marker of each correlation population. In order to glean a deeper understanding for spatial relevance in correlation discrimination, Experiment 2 investigates the effect of density on discrimination performance. Finally, Experiment 3 investigates the effect of a using novel population feature: orientation, in order to determine whether color is unique in the case of perceptual interference during correlation discrimination.

2 EXPERIMENT 1. Discrimination For Two-Population Scatterplots: Color

2.1 Introduction

The purpose of this experiment was to investigate performance for scatterplots containing two data populations, each of a different color. It is common practice to dichotomize a variable on a scatterplot, requiring viewers to process two sets of information on the same set of axes. Rensink (2014) suggested that once a foundational understanding of the perceptual processes behind viewing simple scatterplots was established, researchers could begin investigating various feature manipulations and design changes, to deliberately work towards a controlled study of more complex displays. We implemented a design similar to that of the discrimination task from Rensink and Baldridge (2010), with the addition of a second population of "irrelevant", or "distracting" red dots on each scatterplot. This is the first extension of work by Rensink and Baldridge (2010) to address how multiple correlation populations are simultaneously represented and discriminated.

Importantly, the goal of this work was not only to measure perceptual performance, but also to identify behavioral markers for the underlying perceptual mechanisms. As mentioned in the introduction, Rensink (VSS 2015, 2016) suggested that scatterplots might be perceived via ensemble coding. If so, this study can investigate how multiple ensembles are processed in parallel during a discrimination task. Past ensemble research focused solely on single ensembles (see review by Haberman & Whitney, 2012), but never looked at the case of one ensemble affecting another during a perceptual decision. Visual search research has investigated issues more relevant to this thesis, such as multiple targets (Chun & Potter, 1995), multi-focal attention (Cavanaugh

& Alvarez, 2005), selection for groups (Theeuwes, 1992), and search for sets of similar items (Cave & Wolfe, 1990). However, this is the first study to look explicitly at attentional deployment and selection for multiple, concurring scene statistics in a display.

So what did we expect to happen? As a starting point, we looked to conflicts between Feature Integration Theory (Treisman, 1980) and Guided Search (Wolfe et al., 1989). Using its most straightforward interpretation, Feature Integration Theory would posit that the distracting correlation population should be selected and inhibited successfully during the discrimination task, so that it would not interfere. A color difference should be salient enough to rapidly and pre-attentively identify a target among distractors. However, it is possible that the parsing of ensembles based on spatial information could conflict with selecting information based on color. Im and Chong (2009) showed observers are able to select ensembles based on color; however, they mention that in order to complete a mean size judgment and discrimination, it was necessary for participants to use both color and spatial information. This issue has not been explicitly addressed in either Feature Integration Theory or Guided Search models of attention. Both spatial and featural selectivity processes are hypothesized to occur extremely early in attention, and there is no clear proposition of which would take precedence in a task like this. Thus, it is not clear whether these processes would conflict to slow and worsen performance, or be used together, similar to guided search predictions, to produce fast and accurate performance. But based on the potential for this conflicted dual-process, we expected results to be non-trivial and unique compared to single population viewing; with the likelihood for some kind of perceptual interference to occur when selecting and discriminating the target correlations.

2.2 Methods

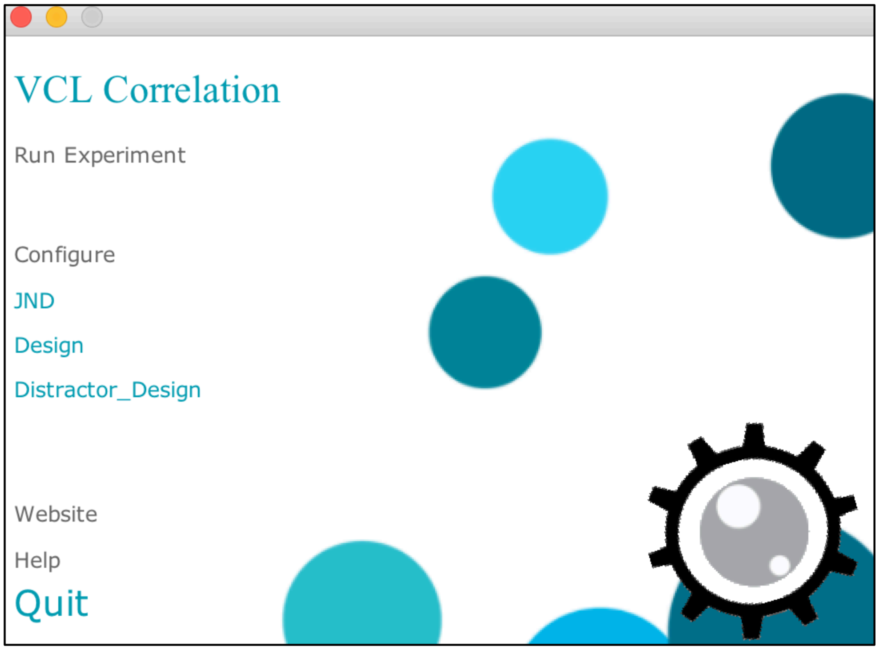
2.2.1 Participants

Participants were recruited through the Reservax® online appointment sign up system, and were paid $10 for partaking in the study, which lasted 60 minutes or less. A total of 15 participants between the ages of 18-35 were recruited, with the average age being 22 years old. 12 of the participants had experience with scatterplots, while 3 reported no familiarity with either scatterplots or the concept of correlation. Participants were screened via self-report for an understanding of instructions in English, and were required to have normal or corrected-to-normal vision. No participants were excluded from analysis.

2.2.2 Apparatus

Participants were each seated approximately 57 cm from a 17" Dell CRT monitor, which displayed the experimental task. Experiments were developed and run using the Visual Cognition Lab (VCL) Correlation Framework (see https://github.com/UBC-VCL/VCLCorrelation). The VCL Framework is experimental vision science software, developed in the Visual Cognition Lab at The University of British Columbia. The VCL Framework is made up of three primary separable components: the graphic user interface (GUI), data input/output (IO), and experiment framework. These components were used together to create and execute the experiment. Experimenters simply configured the task parameters on the GUI (see Figure 2.1) before running the application.

Figure 2.1 VCL Framework Application GUI

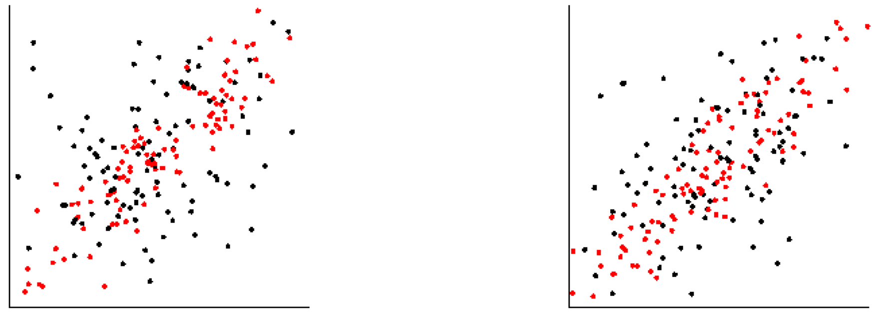


Displaying the configuration for Experiment 1.

2.2.3 Stimuli and experimental conditions

The stimuli used in this experiment are shown in Figure 2.2. The parameters and attributes of these displays were designed and developed to match the scatterplot displays from Rensink and Baldridge (2010). Stimuli were side-by-side, two-population scatterplots, each of 5° vertical extent x 5° horizontal extent. Both plots contained 100 normally distributed black target dots, as well as 100 normally distributed red distractor dots. The means of both target and distractor populations were set to 0.5 of their extent, and the standard deviation of both target and distractor populations were 0.2 of this extent. Each scatterplot correlation contained less than .0001 error in correlation upon display generation. In other words, for any target correlation t, the scatterplot correlation $r = t \pm 0.001$.

Figure 2.2 Side-By-Side Two-Population Scatterplots



Displaying a target correlation population with Pearson's r = .3, and a distractor correlation population with Pearson's r = .9.

In order to generate both target and distractor correlation populations, the experimental framework drew pseudo-random numbers from a Gaussian distribution, which were generated and stored in a list before being drawn on the screen. First, an *x*-coordinate was chosen, scaled, and translated to the axes on the screen. Second, a *y*-coordinate was created and transformed to create a correlated pair (x, y'). The equation used for this process is shown in Figure 2.3. If the program generated a point greater than 2 standard deviations from the mean of the existing points, that point was discarded and replaced with a new point that fell within the 2 standard deviation threshold. This constraint was implemented to prevent points from being drawn outside of the axes. Once the lists of correlation point coordinates were generated for target and distractor populations, the actual points were drawn in the display. A distractor point was always the first to be drawn, followed by a target point, and so on. This was to ensure an even chance of occlusion and overlap by both populations, to eliminate the possibility of a depth illusion or pop out effect for either population. (Note: the stimuli generation was perceptually instantaneous, occurring within a single screen refresh). Points were created using Java Graphics.2D library ellipses, with diameter set to 3 pixels. Axes were 1 pixel wide and always black in color. Finally, the VCL Framework also implemented a scaling algorithm to ensure uniform size of display across any monitor 15" or larger.

Figure 2.3 Equation to Generate Correlated Pairs of *x* and *y* Coordinates for Scatterplot Points

$$y' = \frac{\lambda x + (1-\lambda)y}{\sqrt{\lambda^2 + (1-\lambda)^2}}, \text{ where } \lambda = \frac{r^2 - \sqrt{r^2 - r^4}}{2r^2 - 1}$$

This experiment used a fully within-subjects design to test target correlation discriminations across three conditions of distractor population values. Target correlations of Pearson r = .3, .6, and .9 were tested in a 3 x 4 condition design with distractor correlations of Pearson r = .3, .6, .9, and a control condition with no distractors present. Each target correlation appeared once per distractor correlation, which was repeated in a bidirectional Latin Square balanced design, resulting in a total of 24 sub-conditions per experiment. See Table 2.1 for an example of stimuli presentation in this design.

Table 2.1 Example Stimulus Presentation and Ordering for Experiment 1

Sub-Condition	Target Correlation Pearson *r* Value	Distractor Correlation Pearson *r* Value
1	.3	.3
2	.6	.3
3	.9	.3
4	.3	.6
5	.6	.6
6	.9	.6
7	.3	.9
8	.6	.9
9	.9	.9
10	.3	N/A
11	.6	N/A
12	.9	N/A
13	.9	N/A
14	.6	N/A
15	.3	N/A
16	.9	.9
17	.6	.9
18	.3	.9
19	.9	.6
20	.6	.6
21	.3	.6
22	.9	.3
23	.6	.3
24	.3	.3

Scatterplots appeared on every trial after a 300 ms delay following each response key press, and there was no time limit for the duration of stimulus presentation; the scatterplots remained on the screen until participants made their response.

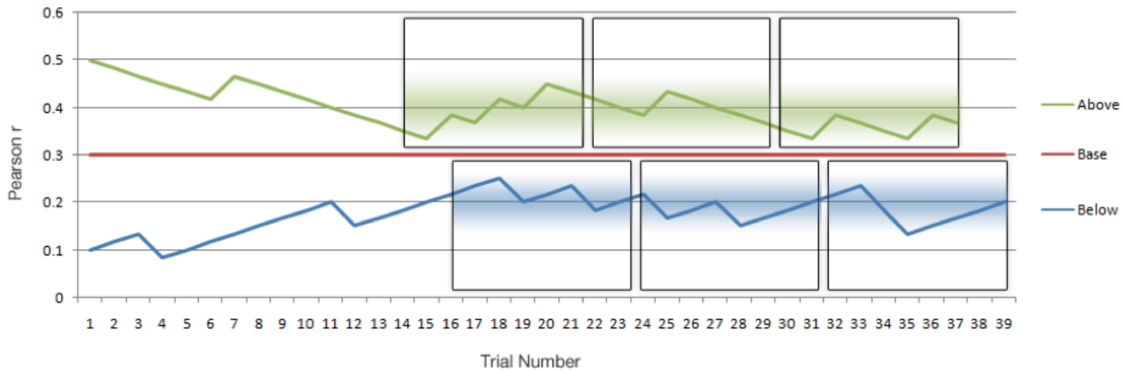
2.2.4 Procedure

Participants were shown side-by-side scatterplots, each containing a target and distractor population. One of the scatterplots always showed a higher correlation value (the test plot, relative to the base plot). The participant's task was the judge whether the right or left scatterplot was showing a higher Pearson's *r* correlation of black target dots, while ignoring the populations of red distractor dots in the display. The experimenter explicitly told participants to "Ignore all red dots you see in the display, and only base your discriminations on the black dots". If the participant believed that the graph on the left was showing a higher correlation of target dots, they were instructed to press the "Z" key. If the participant believed that the graph on the right was showing a higher correlation of target dots, they were instructed to press the "M" key. After each trial, participants received text feedback in the center of the screen, indicating whether they were "correct" or "incorrect" in their response. Following this feedback, participants had to press the space bar to continue to the next trial. Participants were allowed as much time as they wanted to complete each discrimination trial. The experimenter emphasized that accuracy was the most important part of the task, and that there was no pressure to respond in a given amount of time. Additionally, the experimenter informed participants that if they were completely unsure about a discrimination judgment, it was encouraged to simply make their best guess. Participants completed 18 practice trials with feedback, in order to become familiar with the experiment. The experimenter then elicited self-reported feedback from participants to ensure that they understood how to complete the task.

This discrimination task was used to measure observer precision, as described by the sensitivity of observers to differences in target correlation values, while ignoring distractor correlation populations. Closely following the original procedure from Rensink and Baldridge (2010), we used a common staircase algorithm to adjust the target test plot correlation values after every participant response (on each trial). Test plots (containing the higher, adjustable correlation) and base plots were randomly drawn to the left or right of each other on every trial, to ensure that no perceptual learning could occur for which side the higher correlation appeared. Note that in our design, the test plots always converged towards the base plots from above, meaning they always contained the higher correlation value. Initial differences between target correlation values were always 0.1; when participants made a correct response, this difference was decreased by 0.01, thus increasing the difficulty of the task. When participants made an incorrect response, the difference between target correlation values was increased by 0.03, making the next trial easier. Scatterplot coordinates were generated and drawn independently, on each and every trial. This was to ensure that observers based their responses on general perceptual properties of correlation, and eliminated the possibility of observers being able to encode spatial locations of points for each correlation value. Participants continued to make discrimination judgments until a JND, defined as a 75% steady state accuracy, could be calculated. The VCL Framework used a convergence algorithm to determine JNDs. Observer performance was measured over a moving window of 24 trials, divided into 3 sub-windows of 8 trials each (the original image describing this calculation, from

Rensink & Baldridge, 2010, is shown in Figure 2.4). After 24 trials, the average variance within each of the sub-windows was compared to the variance of the averaged of the sub-windows. This test was repeated until the variance was equal to 0.25, at which point the algorithm was considered to converge on a JND. If the algorithm did not converge by trial number 52, the average of the sub windows from the last 24 trials was simply used to calculate the JND. As depicted in Table 2.1, the base correlations of the target populations tested were r = .3, .6, and .9. No manipulation was applied to the generation of distractor correlation values; they always contained the same Pearson's r value on both the test and base scatterplot when present in the display.

Figure 2.4 Schematic of Threshold Algorithm



Schematic obtained from Rensink and Baldridge (2010). The distance from base correlation is adjusted until the variance of the averages of the sub-windows is 0.25 of the average variance within the sub-windows. In the experiments here, only variants above the base value were tested.

2.2.5 Analysis

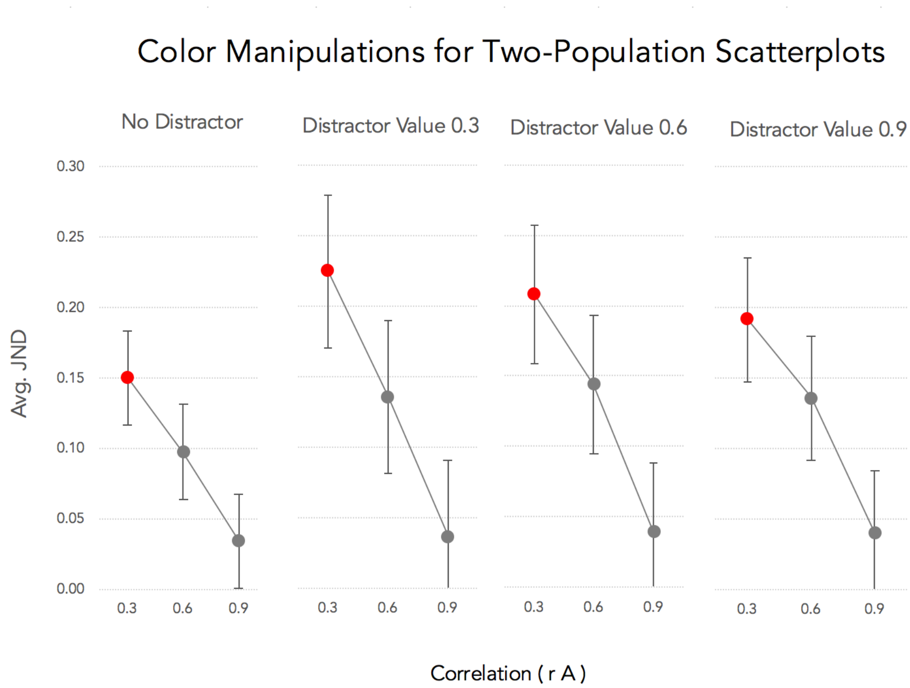
JNDs were calculated and adjusted using the same methods from Rensink & Baldridge (2010), for each target correlation value, over each of the three distractor-present conditions, as well as the no-distractor control condition. A least squares line was fit to the JND performance for each condition, across each participant, in order to obtain slopes and *y*-intercepts. From this, we determined a variability parameter, (k), and a bias offset parameter, defined as the reciprocal of the intersection of the JND line with the *x*-axis, (b), were also calculated for each condition overall, according to the original analysis proposed by Rensink and Baldridge (2015). These values serve the same purpose in the current analysis, as descriptors of accuracy (b), and precision (k), for performance in the perception of correlation.

2.3 Results

JNDs were analyzed with a two-way within-subjects ANOVA using factors target correlation ($r = .3$, $r = .6$, $r = .9$) and distractor correlation condition (none, $r = .3$, $r = .6$, $r = .9$). Tests of between-subjects effects revealed a significant main effect of distractor base, $F(3, 227) = 8.43$, $p = .00$; a significant main effect of target base correlation, $F(2, 227) = 334.59$, $p = .00$; and a significant interaction between the two, $F(6, 227) = 02.43$, $p = .03$. A post hoc Tukey test showed significant differences ($p < .05$) between each of the three distractor base conditions and the no distractor present condition at target correlation $r = .3$ and $r = .6$, though none of the three distractor base conditions were significantly different from one another at any target correlation value. A one-way ANOVA showed no significant difference in b values, which were transformed pre-analysis using a probit function, as suggested in Rensink (2016), $F(3, 76) = .794$, $p =$

.501. However there was a significant difference in k values across distractor value conditions, as determined by a one-way ANOVA, $F(3, 76) = 4.622$, $p = .005$. Post hoc results from this analysis revealed the same significant differences as the Tukey test from the two-way ANOVA; only distractor versus no distractor conditions were significantly different at target r = .03. Results are visualized in Figure 2.5 below.

Figure 2.5 Mean JNDs At Target Correlations *r* = .3, *r* = .6, and *r* = .9 for Experiment 1



Correlations are displayed across four distractor conditions (*r* = .3, *r* = .6, *r* = .9, and no distractor present) as obtained in Experiment 1. Error bars reflect standard error. Mean JNDs for each condition are plotted along the *y*-axis and denoted by color, with the *x*-axis representing the target correlation value.

2.4 Discussion

Much like the results from Rensink and Baldridge (2010), JNDs produced linear behavior across the target correlation values, and R^2 values were high (m = .988). Performance in the no-distractor condition served as a successful replication of previous work, and showed a similar bias parameter (b = .92) and slope (k = -.19) to the foundational discrimination results. Although there was variability across participant results, no subject had to be excluded, based on a 2.5 standard deviation criteria from the mean JNDs. The overall consistency of the data yielded a promising conclusion that the foundational discrimination methodology can be extended and applied to two-population scatterplots.

Loosely following our predictions, the data show an interesting pattern of perceptual inference when a distractor correlation population is present. It is evident that Feature Integration Theory assumptions about selection for color in attention do not hold in the context of these results: there was a significant degradation of discrimination ability in the presence of differently colored distractors. Furthermore, post-hoc analyses from the two-way ANOVA on JNDs, and the analysis of k values, showed this perceptual interference was most pronounced at target correlation *r* = .3. There was no evidence of interference at target correlation *r* = .9, due to expected ceiling effects for observer JNDs. It should be noted that the non-significant difference in b values indicates that there was no change in accuracy for the task; only precision showed interference, suggesting that the process had become more noisy.

These findings suggest that the decline in perceptual performance may be a function of spatial layout in the display. Although no strong conclusions can be drawn as to the nature of this viewing behavior, perhaps there were spatial-conjunctive failures at target correlation $r = .3$ and $r = .6$, due to the spread and proximity of both target and distractor dots. Treisman and Schmidt (1982) postulated that attentional overload could lead to mis-bindings and illusory conjunctions in the context of search, an idea that could account in part for the results of this experiment. This notion raises an important question to be addressed in future studies: will the number of dots in the distracting display affect viewers' discrimination ability?

3 EXPERIMENT 2. Effect of Density on Two-Population Scatterplot Discrimination

3.1 Introduction

In Experiment 1, we showed interference in the perception of two-population scatterplots. Discrimination performance significantly declined, in terms of precision, when distractor populations were present in the display. To further investigate the mechanisms underlying this behavior, we needed to test whether performance was degraded due to spatial mis-bindings. The interference effect from Experiment 1 could have occurred because of inefficient and inaccurate serial processing, as suggested by Treisman and Gormican (1988), but the issue of spatial proximity between the points must be addressed first. Results from this work could also help resolve conflicts in the crowding literature, i.e. the debate as to whether featurally dissimilar flankers help (promote a pop out effect) or hurt (cause a detrimental pooling effect) in attention for the target, depending on their spatial proximities (see Parkes et al., 2011; Whitney & Levi, 2011).

In Experiment 2, we used a design similar to Experiment 1, except we varied the density of the distractor population by increasing or decreasing the number of dots in each sub-condition. We hypothesized that when fewer distractor points were present, discrimination performance would be more precise, showing lower JNDs. Following this line of thinking, we hypothesized that the worst discrimination performance would occur when the number of distractor dots was greater than then number of target dots.

3.2 Methods

3.2.1 Participants

Participants were again recruited through the Reservax® online appointment sign up system, and were this time paid $5 for partaking in the study, which lasted 30 minutes or less. A total of 24 participants between the ages of 18-35 were recruited, with the average age being 23.5 years old. All participants reported previous experience with scatterplots and the concept of correlation. Participants were again screened via self-report for an understanding of instructions in English, and were required to have normal or corrected-to-normal vision. Data from two participants had to be excluded from the analysis due to program failures during testing, which resulted in incomplete responses.

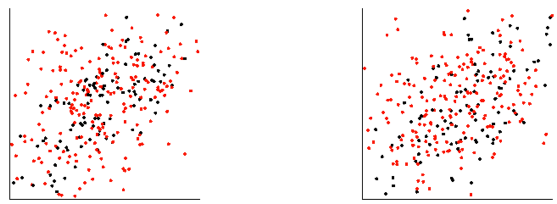
3.2.2 Apparatus

Participants were each seated approximately 57 cm from a 27" iMac monitor, which displayed the experimental task. Despite the hardware changes from Experiment 1, nothing about the look or feel of the display was different from Experiment 1, since the VCL Framework implements a scaling algorithm to ensure stimuli consistency across all monitors. Experiments were developed and executed using the Visual Cognition Lab (VCL) Correlation Framework (see https://github.com/UBC-VCL/VCLCorrelation). Experiment 2 was configured and run using the same GUI shown in Experiment 1.

3.2.3 Stimuli and experimental conditions

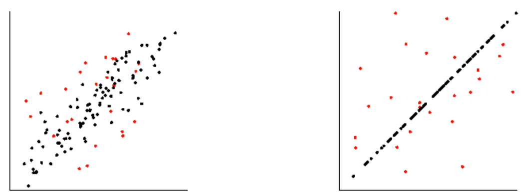
The stimuli used in Experiment 2 were identical to those used on Experiment 1, with the exception of the density manipulations and a fixed distractor correlation value at $r = .3$. Each of the 5 conditions had a different number of distractor population points: 0, 25, 50, 100, and 200 points. For examples of stimuli, see Figures 3.1 and 3.2. As before, target correlation populations always contained 100 points. Target dots were again represented by black dots, and distractors by red dots. The same VCL Framework methodology was used to generate and draw the points on the display.

Figure 3.1 Density Manipulations with 200 Distractors



Example stimuli for conditions with 200 distractor dots present in the display. Correlation was always $r = 0.3$ for distractor dots.

Figure 3.2 Density Manipulations with 25 Distractors



Example stimuli for conditions with 25 distractor dots present in the display. Correlation was always $r = 0.3$ for distractor dots.

This experiment used a within-subjects: target correlations of Pearson $r = .3$, .6, and .9 were tested in a 3 x 5 condition design with number of distractor points 25, 50, 100, and 200, and an additional control condition with no distractors present. Crucially, the distractor correlation was always set to $r = .3$. Each target correlation appeared once per distractor population, which was balanced across observers using a Latin Square balanced design, resulting in a total of 15 sub-conditions per experiment. See Table 3.1 for an example of stimuli presentation in this design.

Table 3.1 Example Stimulus Presentation and Ordering for Experiment 2

Sub-Condition	Target Correlation Pearson *r* Value	Distractor Dot Density
1	.3	0
2	.6	0
3	.9	0
4	.3	25
5	.6	25
6	.9	25
7	.3	50
8	.6	50
9	.9	50
10	.3	100
11	.6	100
12	.9	100
13	.3	200
14	.6	200
15	.9	200

Just as in Experiment 1, scatterplots appeared on every trial after a 300 ms delay following each response key press; there was no time limit for the duration of stimulus presentation; the scatterplots remained on the screen until participants made their response.

3.2.4 Procedure

The procedure was much the same as Experiment 1. Participants were told to discriminate correlation values based on the black target dots in the display, while always ignoring the red dots. Observers were not informed that the distractor population would stay at the same correlation value throughout the experiment, and no participant reported any unusual observations about the distractor population following their task completion. The same staircase algorithm was applied to the task, and the same convergence criteria were used to calculate JNDs. "Z" and "M" keys were used to indicate right or left scatterplots, and this time participants were exposed to 15 practice trials, in order to become familiar with the experiment. The experimenter again elicited self-reported feedback from participants, to ensure that they understood how to complete the task.

3.2.5 Analysis

JNDs were calculated and adjusted using the same methods from Rensink & Baldridge (2010), for each target correlation value, over each of the four distractor-present conditions, as well as the no-distractor control condition. Variability parameter (k), describing precision, and bias offset parameter (b), describing accuracy, were also calculated for each condition.

3.3 Results

JNDs were analyzed with a two-way within-subjects ANOVA using factors target correlation ($r = .3$, $r = .6$, $r = .9$) and distractor correlation condition (none, 25 dots, 50 dots, 100 dots, 200 dots). Tests of between-subjects effects revealed a significant main effect of base correlation, $F(2, 315) = 224.26$, $p = .00$; a significant main effect of number of distractor points, $F(4, 315) = 7.22$, $p = .00$; and a non-significant interaction between the two, $F(8, 315) = 1.82$, $p = .07$. A post hoc Tukey test showed significant differences ($p < .05$) between performance at target $r = .3$ for 0 dots and 50, 100, and 200 dots; 25 dots and 100 and 200 dots; 50 dots and 200 dots.

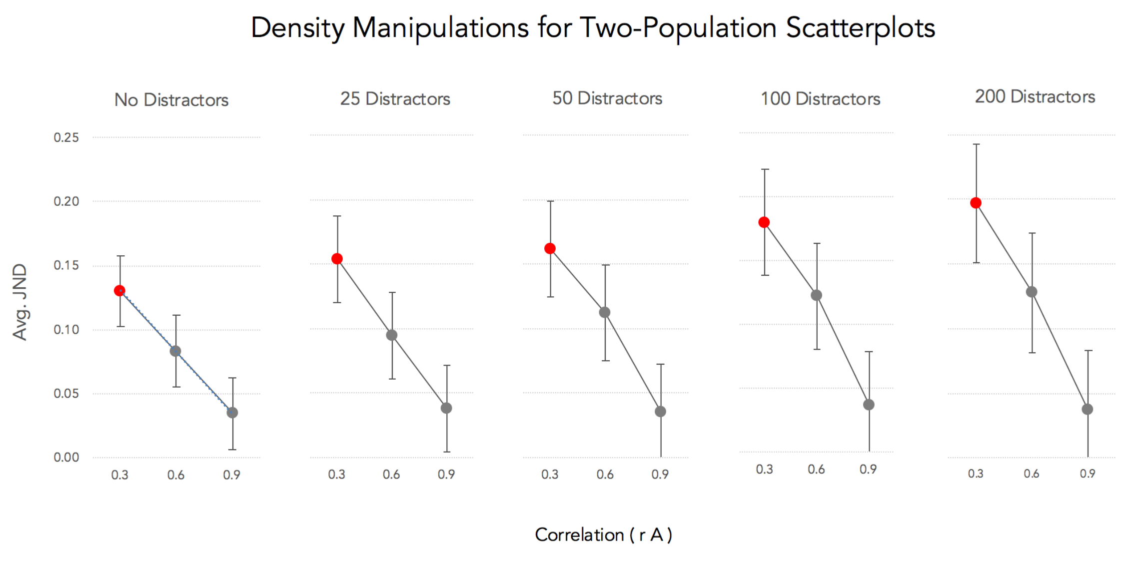
A one-way ANOVA showed no significant difference in b values, which were transformed pre-analysis using a probit function, as suggested in Rensink (2016), $F(4, 105) = 1.120$, $p = .351$. However there was a significant difference in k values across density conditions, as determined by a one-way ANOVA, $F(4, 105) = 4.157$, $p = .004$. Post hoc results from this analysis revealed the same significant differences as the Tukey test from the two-way ANOVA. Post hoc results from the two-way ANOVA are shown in Table 3.2. Results are visualized in Figures 3.3 and 3.4 below.

Table 3.2 Tukey Test Results at Target Correlation *r* = .3 for Experiment 2

Number of Distractor Points	0 Dots	25 Dots	50 Dots	100 Dots	200 Dots
0 Dots	--	--	*	*	*
25 Dots	--	--	--	*	*
50 Dots	*	--	--	--	*
100 Dots	*	*	--	--	--
200 Dots	*	*	*	--	--

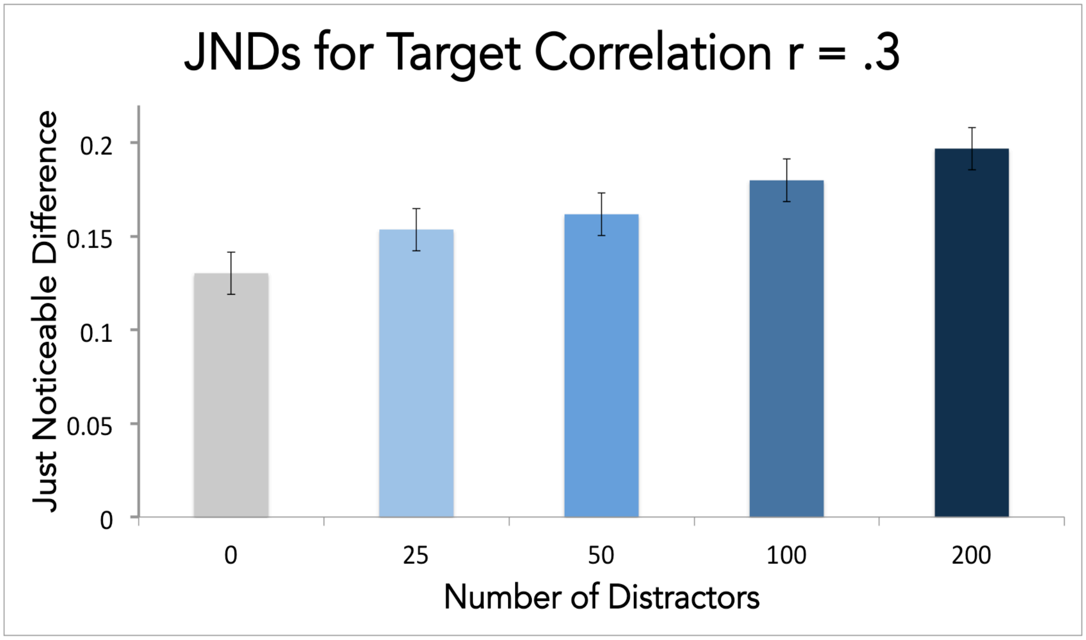
*$p < .05$.

Figure 3.3 Mean JNDs at Target Correlations $r = .3$, $r = .6$, and $r = .9$ for Experiment 2



Correlations are displayed across five distractor conditions (25 dots, 50 dots, 100 dots, 200 dots, and no distractor present) as obtained in Experiment 2**.** Error bars reflect standard error. Mean JNDs for each condition are plotted along the *y*-axis and denoted by color, with the *x*-axis representing the target correlation value.

Figure 3.4 Mean JNDs at Target Correlation *r* = .3



Correlations are displayed across five distractor dot conditions (25 dots, 50 dots, 100 dots, 200 dots, and no distractor present) as obtained in Experiment 2. Error bars reflect standard error. Mean JNDs for each condition are plotted along the *y*-axis, with the *x*-axis representing the number of distractor dots in the display.

3.4 Discussion

Experiment 2 showed further results similar to Rensink and Baldridge (2010). JNDs produced linear behavior across the target correlation values, and R^2 values were high once more (m = .99). Performance in the no-distractor condition from this experiment served as a successful replication of previous work, and showed a very similar bias parameter (b = .86) and slope (k = -.16) to the foundational discrimination results. There were again no participant exclusions, based on a 2.5 standard deviation criteria from the mean JNDs. All of these findings suggested that the methodology could be successfully extended to investigate density manipulations of distractor populations for the perception of correlation in scatterplots.

The main results from the two-way ANOVA, and the analysis of k's and b's, very closely followed our original predictions. The data showed a clear pattern of perceptual interference for target correlations, which occurs as a function of the number of distractor points in the display. The more red dots in the distractor population, the worst observers were at discriminating target correlation values. Consistent with the findings from Experiment 1, the worst performance occurred at target correlation $r = .3$. Additionally, the non-significant difference in b values suggests that this performance was not due to a change in accuracy (bias), and was instead attributable to an increase in noise, i.e., true interference.

These findings provided a replication of the feature selection failures for color that were identified in Experiment 1. Results from this experiment provided a more granular view of how viewing behavior is affected by the density and spatial proximity of the distractors. Although the presence of any distractors in the display was enough to cause significant perceptual interference, reducing the number to just 50 points or fewer yielded similar performance to interference for the 100 dot distractor populations of correlation value $r = .9$ from Experiment 1. This work falls mostly in line with assumptions from Treisman and Schmidt (1982), suggesting that an attentional overload resulting in mis-bindings is likely causing the degradation in performance as distractor dot populations increase in size. However, the 200-distractor dot results pose problems for Guided Search frameworks (see Wolfe et al., 1989), as viewers are unable to capitalize on two striking differences between the populations: color and density. This is somewhat surprising in the context of work by Im and Chong (2009), which showed that color and location information could be optimally combined by observers to aid ensemble discrimination tasks. It should be noted however, that these authors' tasks involved single ensembles side by side, rather than two ensembles overlapping in the same display, like the studies in this thesis. Given these featural distinctions, it would have been reasonable to expect an advantageous pop-out effect to occur for the targets.

Results from Experiments 1 and 2 provide significant evidence that parsing and judging ensembles in attention is perceptually unique from both visual search, as well as the observation of a single ensemble. Feature selection failures cause unexpected interference when discriminating target correlation populations- but is this just the case for color and density? In order to be certain that interference is not a phenomenon unique to color distinction, future studies should investigate what happens when additional low-level visual features represent both distractor and target correlation populations.

4 EXPERIMENT 3. Discrimination for Two-Population Scatterplots: Orientation

4.1 Introduction

Based on the results from Experiments 1 and 2, two-population viewing behavior observed so far could be accounted for by models of crowding and spatial mis-binding for targets and flanking distractor dots (Treisman & Schmidt, 1989; Whitney & Levi, 2011). However, it remains unclear whether or not color is a special feature. In Experiment 3, we addressed this question by using different orientations to distinguish target versus distractor correlation populations. Items with differing orientations have been shown to promote both texture segmentation (Landy & Bergen, 1991) and pop-out in visual search (Nothdurft, 1991; Wolfe, 1994); the effect is magnified further when opposing cardinal orientations (90^{o} vertical versus 180^{o} horizontal lines) are used. Thus, an interference result would confirm that color is not a unique case in the perceptual of two-population scatterplots, and that the visual system indeed treats correlation ensembles differently from other kinds of stimuli.

In Experiment 3, we used a design similar to Experiment 1, except we changed the dots in target and distractor populations to vertically and horizontally oriented lines. We hypothesized that changing the featural representation of data points would not significantly improve discrimination performance for the target correlations. We expected the same pattern of behavior as in Experiment 1, and interference was expected to increase as a function of decreasing distractor correlation values.

4.2. Methods

4.2.1 Participants

Participants were again recruited through the Reservax® online appointment sign up system, and were paid $5 for partaking in the study, which lasted 30 minutes or less on average. A total of 31 participants between the ages of 18-35 were recruited for this experiment, with the average age being 22.5 years old. 29 of the participants reported previous experience with scatterplots and the concept of correlation. Participants were again screened via self-report for an understanding of instructions in English, and were required to have normal or corrected-to-normal vision. Data from one participant had to be excluded from the analysis, again due to program failures during testing, which resulted in incomplete responses.

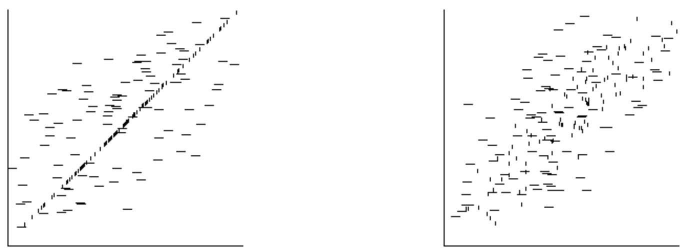
4.2.2 Apparatus

Participants were each seated approximately 57 cm from a 27" iMac monitor, which displayed the experimental task. Experiments were developed and executed using the Visual Cognition Lab (VCL) Correlation Framework (see https://github.com/UBC-VCL/VCLCorrelation). Experiment 3 was configured and run using the same GUI shown in Experiment 1.

4.2.3 Stimuli and experimental conditions

The stimuli used in Experiment 3 were similar to those used on Experiments 1 and 2, but this time each population was represented by line orientation. Vertically oriented lines (8 pixels high x 1 pixel wide) represented target correlation populations, and horizontally oriented lines (1 pixel tall x 8 pixels wide) represented distractor correlation populations. Just as in Experiment 2, there were five conditions, but this time four of them had a different distractor Pearson r correlation value: $r = .3$, $r = .6$, $r = .9$, $r = .999$, and one condition did not contain distractor populations. For an example of the stimuli, see Figure 4.1. Both the target correlation population and distractor correlation populations always contained 100 lines (points). The same exact methodology was used to generate and draw the oriented lines (points) on the display.

Figure 4.1 Side-By-Side Orientation Scatterplots



Displaying a target correlation population with Pearson's *r* = .9, and a distractor correlation population with Pearson's *r* = .6.

This experiment again used a fully within-subjects design to test target correlation discrimination across three conditions of distractor values. Target correlations of Pearson *r* = .3, .6, and .9 were tested in a 3 x 5 condition design with distractor correlations of *r* = .3, .6, .9, .999, and a control condition with no distractors present. Each target correlation appeared once per distractor correlation, which was balanced across observers using a Latin Square balanced design, resulting in a total of 15 sub-conditions per experiment. See Table 4.1 for an example of stimuli presentation in this design.

Table 4.1 Example Stimulus Presentation and Ordering for Experiment 3

Sub-Condition	Target Correlation Pearson *r* Value	Distractor Correlation Pearson *r* Value
1	.3	N/A
2	.6	N/A
3	.9	N/A
4	.3	.3
5	.6	.3
6	.9	.3
7	.3	.6
8	.6	.6
9	.9	.6
10	.3	.9
11	.6	.9
12	.9	.9
13	.3	.999
14	.6	.999
15	.9	.999

4.2.4 Procedure

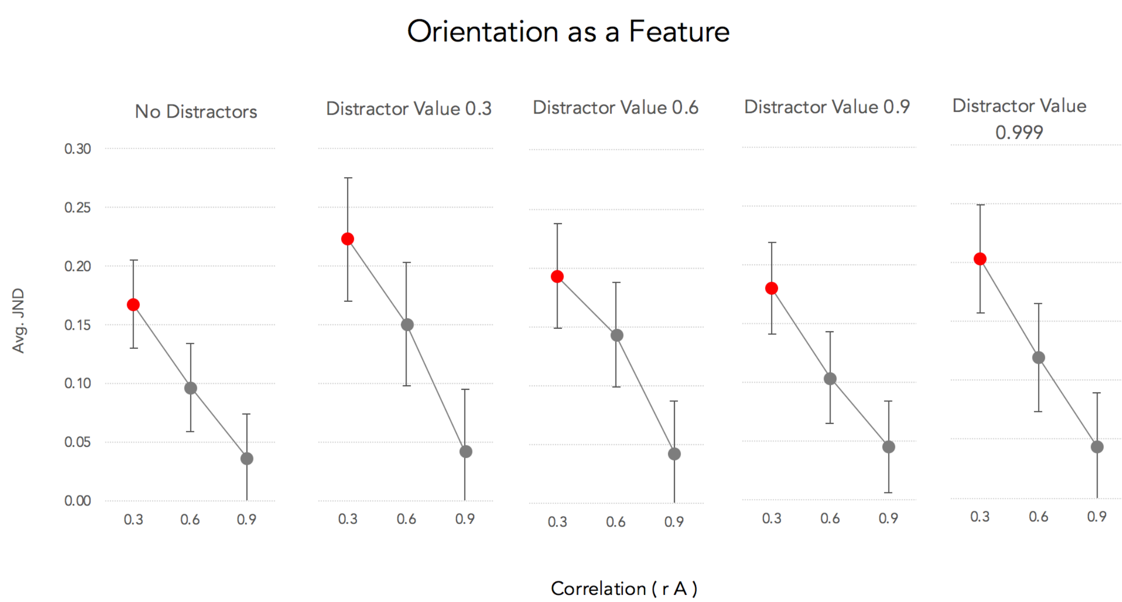
4.2.5 Analysis

JNDs were calculated and adjusted using the same methods from Rensink & Baldridge (2010), for each target correlation value, over each of the four distractor-present conditions, as well as the no-distractor control condition. Variability parameters (k), describing precision, and bias offset parameters (b), describing accuracy, were also calculated for each condition.

4.3 Results

JNDs were analyzed in the same way as Experiments 1 and 2; with a two-way within-subjects ANOVA using factors target correlation ($r = .3$, $r = .6$, $r = .9$) and distractor correlation condition (none, $r = .3$, $r = .6$, $r = .9$, $r = .999$). Tests of between-subjects effects revealed a significant main effect of distractor base, $F(4, 435) = 4.82$, $p = .01$; a significant main effect of target base correlation, $F(2, 435) = 206.68$, $p = .00$; and a non-significant interaction between the two, $F(8, 435) = 1.316$, $p = .23$. A post hoc Tukey test showed significant differences ($p < .05$) between the distractor $r = .3$ condition, and the no distractor present condition for target correlation $r = .3$, as well as a significant difference between distractor $r = .3$ condition, and distractor $r = .9$ condition, for target correlation $r = .3$. There were no other significant differences in distractor value manipulations in Experiment 3. A one-way ANOVA showed no significant difference in probit transformed b values, $F(4, 145) = .180$, $p = .948$. There was also no significant difference in k values across orientation conditions, as determined by a one-way ANOVA, $F(4, 145) = 1.379$, $p = .244$. Results are shown in Figure 4.3 below.

Figure 4.2 Mean JNDs at Target Correlations $r = .3$, $r = .6$, and $r = .9$ for Experiment 3



Correlations are displayed across five distractor dot conditions ($r = .3$, $r = .6$, $r = .9$, $r = .999$, and no distractor present) as obtained in Experiment 3. Error bars reflect standard error. Mean JNDs for each condition are plotted along the *y*-axis and denoted by color, with the *x*-axis representing the target correlation value.

4.4 Discussion

Results in Experiment 3 showed the most dramatic deviation from the original findings in Rensink and Baldridge (2010). JNDs produced linear behavior across target correlation values for four of the distractor correlation values, however, there was a clear non-linear variation in performance when distractor Pearson's $r = .6$. The R^2 value for this condition was noticeably low ($R^2 = .96$) compared to the other 4 conditions (m = .99). No-distractor performance in Experiment 3 did serve as a successful replication of previous work (b = .94), (k = -.22), and were similar to foundational discrimination results, although b was slightly higher than expected. Neither b or k analyses yielded significant results, suggesting that accuracy and precision were not affected greatly. The perceptual interference is only apparent at the broader-level JND analysis. Curiously, there were no participant exclusions, based on a 2.5 standard deviation criteria from the mean JNDs. Despite some nonconformity to the original pattern of findings, these results were consistent enough to support this methodology successfully extending to our orientation manipulations. It is reasonable to assume that participants understood and could perform this task.

Overall, this work shows that features in this task behave very differently than in the processing of multiple ensemble populations, where target populations must be extracted and distracting populations must be inhibited. Contrary to previous work detailing advantageous pop-out effects in attention and perceptual accuracy for discriminating cardinal orientations (Landy & Bergen, 1991; Wolfe, 1994), these findings indicate a clear failure in both processes. Moreover, spatial arguments (Parkes et al., 2011; Treisman & Schmidt, 1989; Whitney & Levi, 2011) begin to unravel in the context of these results, as the non-linearity of the distractor $r = .6$ and unexpectedly superior discrimination performance when distractor $r = .9$ cannot be parsimoniously accounted for by crowded flankers or mis-bindings.

The issue of illusory conjunctions, or even novel object shapes does come into play in Experiment 3, given the nature of the orientation stimuli. Rather than a simple occlusion occurring, when horizontal and vertical lines overlap, they create a new cross-shape in the scene. It may be that these instances are being parsed as another set of stimuli, which may account for the interference behavior. However, work by Wolfe and DiMase (2003) showed that search for the presence and absence of orientation intersections was slow and inefficient, and therefore concluded that these “cross-shapes” should not be considered salient features, nor do they promote efficient search. Nonetheless, it may be the case that these new features operate differently in the context of a scene statistic ensemble population, versus a simple group of objects or features. It could be argued that in some cases, these overlaps could be advantageous for extracting the target correlation values: for instance, if they occur in the periphery of the display and isolate or “knock-out” outliers. In other cases, if too much overlap occurs in the central area of the scatterplot, crucial visual information about the target population may be lost. This hypothesis could account in part for relatively poor performance in the distractor $r = .999$ condition. Future studies could investigate the impact of overlapping orientation stimuli, by recording their occurrence and spatial location in each trial, and correlating that information with JND performance.

5 CONCLUSIONS

5.1 General Discussion

In Experiments 1 and 2, we demonstrated that perceptual interference in the perception of correlation in two-population scatterplots occurs both as a function of increased density and decreased Pearson's *r* in target and distractor populations. Although the mechanisms underlying the extraction (and inhibition) of correlation ensembles remain elusive, it appeared that observers were unable to use color differences to optimize task performance. Additionally, participants were unable to use differences in population numerosity to their advantage. The least optimal performance was observed when a greater number of differently colored distractors were present in the scatterplot display. Experiment 3 showed slightly conflicted results for orientation as a discriminating feature between target and distractor populations. While there was clear evidence for perceptual interference in the presence of distractors, this interference behavior was noisy and variable compared to Experiments 1 and 2. Additionally, it did not follow the same pattern of greater numbers of distractor points leading to worsened discrimination performance for targets.

5.2 Implications for Visual Perception

Work from Rensink (2014) showed that, much like other visual quantities, participants are able to rapidly infer and compare correlations in single population scatterplots. In fact, the correlation representations in these studies were largely complete by 100 ms, and discrimination task performance closely approximated Weber's law, which is commonly associated with discrimination over many types of visual quantities. These results are highly consistent with the idea that some form of holistic processing, or ensemble encoding is involved in the perception of correlation in scatterplots. In the studies here, no time constraint is put on participants, but participant response times appear relatively consistent with previous perception of correlation experiments. Guided search and subsequent hybrid serial/parallel feature selection search models (Wolfe et al., 2011) suggest that attentional selection should be possible for the target dots in correlation displays, given their featural difference (color) from the distractor dots. However, the data suggested that a loss of target population information is occurring in the presence of distractor populations. For the most part, these results were more consistent with attentional overloading, which slows efficiency and accuracy in search tasks, as noted in early feature integration theory and conjunction literature (Treisman & Gelade, 1980; Treisman & Schmidt, 1982).

Surprisingly, literature on guided search and newer feature selection models never fully addressed the nature of the attentional overloading from early feature integration models. As mentioned in the Introduction, neither guided search, nor Feature Integration Theory can account for salient, known distractors slowing search times for uniquely salient targets in Theeuwes (1992). Instead, vision researchers seemed to have focused on discovering and operationalizing conditions where selection is or is not possible. Furthermore, recent work on scene gist, ensembles, and crowding often fails to mention the possibility of attentional overloading; instead focusing on the perceptual successes of global processing. Here, we demonstrated an argument for attentional overload in ensemble extraction, which could be further perceptually complicated by the presence of novel populations, created by overlapping orientation points in the display.

There is still no universally accepted model detailing how exactly humans select and encode individual features and objects. Nor is there a clear answer to when information is lost during localized spatial and density encoding issues, feature binding/illusory conjunction, and crowding. Although the field of vision science is currently saturated with behavioral findings under these effects, the true nature of the results remains elusive.

5.3 Implications for Information Visualization

We were motivated by two primary goals in our work on distracting correlation ensembles. In the realm of vision science, our interest was in identifying new patterns of behavior that could guide insight into the perceptual mechanisms involved with extracting and inhibiting multiple ensembles in attention. And in the realm of data visualization, we were able to ascertain several clear recommendations for the creation of scatterplot displays: 1) single-population displays will always be perceived more accurately (as described by b) and precisely (as described by k) by viewers, and are particularly advantageous when observers are expected to compare data; indeed, they seem robust and invariant to many feature manipulations in the display 2) in a scenario where two populations are used, such as visualizing a dichotomized variable, precision and accuracy will suffer as a function of disparity in N data points between the two populations, 3) while we cannot yet offer an optimal recommendation for which features to use when differentiating data populations, using color appears to produce more predictable and systematic interference; this is important if designers want to be able to predict viewer performance deficits.

5.4 Future Directions

There are many important questions remaining, following results from this thesis. The first is whether viewers are losing spatial information about the target populations as a result of perceiving distractor populations simultaneously. Additionally, the distracting points could be interfering with viewers' ability to discern the shape of the probability envelope of the target population, as described in Rensink (2016), which would lead to very noisy correlation estimates. This could be investigated using a timing manipulation, where one population is briefly displayed first, and then the other. This would give participants a chance to encode information about either the target or distractor populations, and potentially use it to select or inhibit that information. Successful encoding would yield better discrimination performance. Additionally, it will be important to look at the differences in colors (or more generally, features) between target and distractor populations, to ensure that black and red dots are not a unique perceptual case.

Finally, looking forward a bit further, this two-scatterplot methodology should eventually be extended to investigate more complex visualizations, as well as other features to define the populations. Displays like clustered network diagrams, multidimensional scaling planes, and non-linear scatterplot displays are all candidates for the relative near future. Once perceptual behavior for two populations is well understood, we can sequentially and deliberately extend these studies to increasingly complex graphs and data sets. As technologies like Tableau® continue to push the limits of visualization capabilities and design options, there will be an ever-increasing collection of data visualizations to investigate and study.

5.5 Conclusion

In section 1, I outlined several gaps and contradictions in vision scientists' understanding of feature selection and ensemble processing. Additionally, I motivated this work with the applied nature of our findings, which are useful in the realm of data visualization. Experiment 1 solidified that our discrimination task methodology could be successfully applied to two-population scatterplots. Findings from this study demonstrated the occurrence of perceptual interference when a second, distracting population is present. Experiment 2 showed that this interference worsens as density of the distractors increases. Finally, Experiment 3 showed that the interference effect was not limited to a single feature: color.

This study showed that representing the correlation data populations with differently oriented stimuli yields a more complex flavor of perceptual interference in our discrimination task. Overall, this work identified novel perceptual behaviors and presented brand new research potential for the investigation of multiple ensemble discrimination. These results suggest that ensemble selection is a unique process in attention, and does not conform to models of Guided search, Feature Integration Theory, or single ensemble processing.

References

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: extracting statistical summary representations over time. *Psychological Science, 21*, 560–567.

Alvarez, G. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*, 122-131.

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*(4), 392-398.

Alvarez, G.A., & Oliva, A. (2009) Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proc. Nat. Acad. Sci. U.S.A., 106*, 7345–7350.

Ariely, D. (2001). Seeing sets: representation by statistical properties. *Psychological Science,* 12, 157–162.

Bobko, P., & Karren, R. (1979). The perception of Pearson product moment correlations from bivariate scatterplots. *Personnel Psychology*, *32*(2), 313-325.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226,* 177-178.

Brady, T., & Alvarez, G. (2011). Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384–392.

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, *9*(7), 349-354.

Chong, C., & Treisman, A. (2005). Statistical processing: computing the average size in perceptual groups. *Vision Research, 45*, 891-900.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 109.

Cleveland, H. (1982). Information As a Resource. *Futurist*, *16*(6), 34-39.

Cleveland, W. S., Diaconis, P., & McGill, R. (1982). *Variables on scatterplots look more highly correlated when the scales are increased* (No. PJH-12). HARVARD UNIV CAMBRIDGE MASS DEPT OF STATISTICS.

Cleveland, W. S., & McGill, R. (1986). An experiment in graphical perception. *International Journal of Man-Machine Studies*, *25*(5), 491-500.

Cohen, A., & Ivry, R. (1989). Illusory conjunctions inside and outside the focus of attention. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(4), 650.

Coren, S, Ward, L.M., & Enns, J.T. (1999). *Sensation and Perception* (5th edition). New York: Harcourt Brace.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science*, *320*(5880), 1217-1220.

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, *48*(1), 269-297.

Elliott, M., Rensink, R., & Ngu, H. (2015). Visual Cognition Lab Correlation Framework [Computer Software]. Retrieved 18th July 2016 from https://github.com/UBC-VCL/VCLMindWandering.

Fougnie, D. & Marois, R. (2006). Distinct capacity limits for attention and working memory: Evidence from attentive tracking and visual working memory paradigms. *Psychological Science, 6,* 526-534.

Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718.

Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. *From Perception to Consciousness: Searching with Anne Treisman*, 339-349.

Marín Díaz, J. P. (2016). *How we did data visualizations before computers.* Retrieved 16th June 2016 from http://www.datasketch.co/en/p/how-we-did-data-visualizations-before-computers.

Harrison, L., Yang, F., Franconeri, S., & Chang, R. (2014). Ranking visualizations of correlation using weber's law. *IEEE transactions on visualization and computer graphics*, *20*(12), 1943-1952.

Humphreys, G. W., Watson, D. G., & Jolicœur, P. (2002). Fractionating the preview benefit in search: Dual-task decomposition of visual marking by timing and modality. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(3), 640.

Im, H. Y., & Chong, S. C. (2009). Computation of mean size is based on perceived size. *Attention, Perception, & Psychophysics*, *71*(2), 375-384.

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43*(3), 171- 216.

Kahneman, D., Treisman, A., & Burkell, J. (1983). The cost of visual filtering. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(4), 510.

Keim, D., Qu, H., & Ma, K. L. (2013). Big-data visualization. *IEEE Computer Graphics and Applications*, *33*(4), 20-21.

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin*, *112*, 24–38.

Landy, M. S., & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Research*, *31*(4), 679-691.

Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision*, *9*(10), 6.

Luck, S., & Vogel, E. (1997) The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279-281.

Meyer, J., & Shinar, D. (1992). Estimating correlations from scatterplots. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *34*(3), 335-349.

Munzner, T. (2009). A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, *15*(6), 921-928.

Nothdurft, H. C. (1991). Texture segmentation and pop-out from orientation contrast. *Vision Research*, *31*(6), 1073-1078.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, *34*, 72-107.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155*, 23-36.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*(7), 739-744.

Pashler, H. (Ed.). (2016). *Attention*. Abingdon, UK: Psychology Press.

Peterson, C.R., & Beach, L.R. (1967) Man as an intuitive statistician. *Psychological Bulletin*, *68*, 29–46.

Potter, M. C. (1975). Meaning in visual sciences. *Science*, *187*, 965-966.

Potter, M. C. (1976). Short term conceptual memory for pictures. *Journal of Experimental Psychology, 2*, 509-522.

Potter, M. C., Staub, A., & O'Connor, D. H. (2002). The time course of competition for attention: Attention is initially labile. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(5), 1149.

Prinzmetal, W., Henderson, D., & Ivry, R. (1995). Loosening the constraints on illusory conjunctions: assessing the roles of exposure duration and attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(6), 1362.

Rensink, R. A., & Baldridge, G. (2010, June). The perception of correlation in scatterplots. In *Computer Graphics Forum* (Vol. 29, No. 3, pp. 1203-1210). Hoboken, NJ: Blackwell Publishing Ltd.

Rensink, R. A. (2014). On the prospects for a science of visualization. In *Handbook of human centric visualization* (pp. 147-175). New York: Springer.

Rensink (2016). The Nature of Correlation Perception in Scatterplots. *Manuscript submitted for publication.*

Rosen, S., Chakravarthi, R., & Pelli, D. G. (2014). The Bouma law of crowding, revised: Critical spacing is equal across parts, not objects. *Journal of Vision*, *14*(6), 10-10.

Ross, J., & Burr, D. C. (2010). Vision senses number directly. *Journal of Vision*, *10*(2), 10.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379-423, 623-656.

Shapiro, K. L., Raymond, J. E., & Arnell, K. M. (1997). The attentional blink. *Trends in Cognitive Sciences*, *1*(8), 291-296.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological Science*, *5*(4), 195-200.

Simoncelli, E. P. & Olshausen, B. A., (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24,* 1193-1216.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059-1074.

Simons D.J., and Rensink, R.A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, *9*, 16-20.

Stockdale, J. E. (1978). Crowding: Determinants and effects. *Advances in Experimental Social Psychology*, *11*, 197-247.

Strahan, R. F., & Hansen, C. J. (1978). Underestimating correlation from scatterplots. *Applied Psychological Measurement*, *2*(4), 543-550.

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

Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*(1373), 1295-1306.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology*, *40*(2), 201-237.

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

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*(1), 107-141.

Ware, C. (2012). *Information visualization: perception for design*. Amsterdam: Elsevier.

Watson, D. G., & Humphreys, G. W. (1997). Visual marking: prioritizing selection for new objects by top-down attentional inhibition of old objects. *Psychological Review*, *104*(1), 90.

Whitney, D., & Levi, D. M. (2011). Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*(4), 160–168.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202-238.

Wolfe, J. M., & DiMase, J. S. (2003). Do intersections serve as basic features in visual search? *Perception*, *32*(6), 645-656.

Wolfe, J. M., Oliva, A., Horowitz, T. S., Butcher, S. J., & Bompas, A. (2002). Segmentation of objects from backgrounds in visual search tasks. *Vision Research*, *42*(28), 2985-3004.

Wolfe, J. M., Võ, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*(2), 77-84.