Color semantics for visual communication

Karen B. Schloss, Melissa A. Schoenlein, Kushin Mukherjee

Abstract Visual communication through information visualizations (e.g., graphs, charts, maps, diagrams, and signage) is central to how people share knowledge. In information visualizations, visual features like color are used to encode concepts represented in the visualization ("encoded mappings"). However, people have expectations about how colors map to concepts ("inferred mappings"), which influence the ability to interpret encoded mappings. Inferred mappings have an effect even when legends explicitly specify the encoded mappings and when encoded concepts lack specific, strongly associated colors. In this chapter we will discuss factors that contribute to inferred mappings for visualizations of categorical information and visualizations of continuous data. We will then discuss how these different kinds of factors can be united into a single framework of assignment inference. Understanding how people infer meaning from colors will help design information visualizations that facilitate effective and efficient visual communication.

1 Introduction

When observers look at information visualizations such as weather maps, political polling charts, and airport terminal signage, the input they receive is just an array of light projected onto the retinas of their eyes. Yet, from this input, observers ultimately

Karen B. Schloss
University of Wisconsin–Madison, Department of Psychology and Wisconsin Institute for Discovery, 330 N Orchard St, Madison, WI 53715, e-mail: kschloss@wisc.edu,

Melissa A. Schoenlein
University of Wisconsin–Madison, Department of Psychology and Wisconsin Institute for Discovery, 330 N Orchard St, Madison, WI 53715, e-mail: schoenlein@wisc.edu

Kushin Mukherjee
University of Wisconsin–Madison, Department of Psychology and Wisconsin Institute for Discovery, 330 N Orchard St, Madison, WI 53715, e-mail: kmukherjee2@wisc.edu

glean knowledge about the world. They find out if it is likely to rain during their afternoon walk, which political candidate is expected to win an election, or which direction to dash to reach their gate before their flight departs.

Extensive perceptual and cognitive processing is needed to go from light stimulating the retina to knowledge about the world. When interpreting information visualizations, this processing includes, but is not limited to, (1) detecting and discriminating visual features (e.g., color, shape, size, texture) [7, 16, 2], (2) mapping visual features onto the concepts they represent and [16, 28, 39, 44] (3) using (1) and (2) to derive implications about information represented in the visualization [45, 37, 9, 51]. In this chapter, we will focus on (2) by asking: how do people infer meaning from visual features?

At first, it may seem like the answer is straightforward: people can simply examine legends, labels, or accompanying text to determine the meanings of visual features. However, the answer is not so simple. People have expectations about the meanings of visual features, and visualization designs that violate those expectations are harder to interpret. Let's consider two examples.

The first example is a study by Lin et al. [18] on the meanings of colors for visualizations of categorical information. In their study, Lin et al. presented participants with colored bar charts in which each color represented a different category (e.g., kinds of fruits) (Figure 1A). In one condition, the colors of the bars were selected by an algorithm that maximized the fit between concepts and colors. In another condition, the colors were default colors used by Tableau visualization software (ignoring the concepts represented in the visualization). The charts included a legend to indicate the category corresponding to each bar color. Participants were asked to answer questions about the data in the chart, and their response time (RT) was recorded. RT is a measure of interpretability, such that faster RTs for correct responses suggest greater interpretability. RTs were faster when the colors were optimized to match people's expectations, compared to the default Tableau colors, even though there was a clear legend indicating the meaning of the colors in both conditions.

The second example is a study by Schloss et al. [37] on the meanings of color for visualizations of continuous data. In their study, Schloss et al., presented participants with colormap data visualizations, in which gradations of colors represented gradations of quantity (Figure 1B). Participants were told that the colormaps represented alien animal sightings on the planet Sparl, and their task was to indicate whether there were more sightings early or late in the day. The colormap visualizations included a legend that specified the mappings between lightness (dark to light) and quantity (greater to fewer sightings). Overall, participants were faster at correctly interpreting the colormap when the legend specified darker colors mapped to more. This is because observers have a dark-is-more bias leading to the expectation that darker colors map to larger quantities (see Section 3.1.2).

In both of these examples, legends indicate the meanings of colors. But, when the encoding indicated in the legend violates people's expectations, visualizations are harder to interpret. Thus, understanding visual communication requires understanding people's expectation about the meaning, or *semantics*, of visual features.

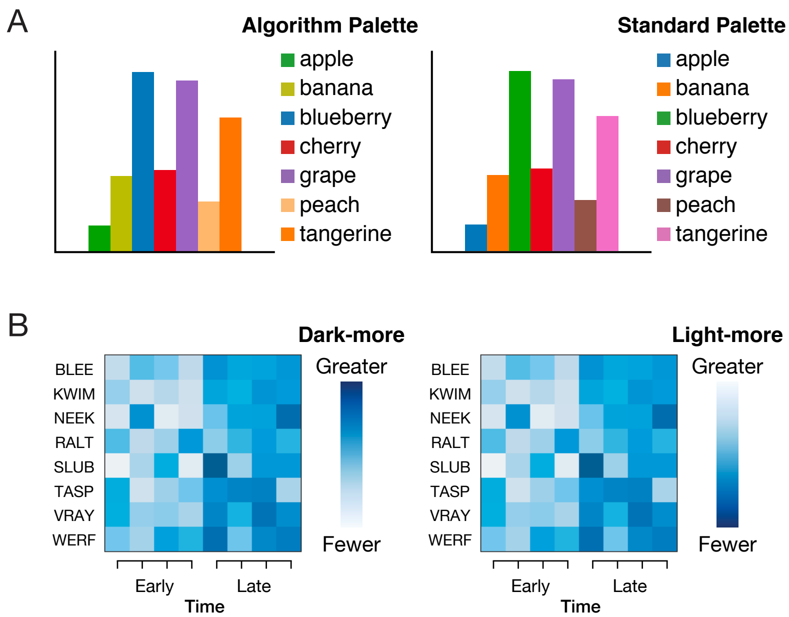

Fig. 1 (A) Bar charts representing fictitious data about fruit sales, with colors selected by an algorithm to maximize fit between concepts and colors (left) or colors determined by a standard Tableau palette order (right) (figure based on stimuli in Lin et al. [18]). (B) Colormap data visualization representing alien animal sightings at different times of day, with a legend specifying dark-more mapping (left) or light-more mapping (right) (figure based on stimuli in Schloss et al. [37]).

1.1 Visual semantics from multiple perspectives

When discussing visual semantics for visual communication there are two perspectives to consider: the perspective of the designer and the perspective of the observer. If these perspectives are aligned, observers are more likely to interpret the message that the designer intended to convey through the visualization [25, 51, 12, 50, 39, 18].

Perspective of the designer. When we use the term 'designer', we do so liberally to refer to anyone who creates a visualization. This could be a professional designer, but it could also be an undergraduate student creating a chart from data in their research methods course, or even a middle school student creating a diagram of the protocol for their science fair project [39]. In cases where people create visualizations for the purpose of exploring and finding patterns in their own data [10], the designer and the observer are the same person.

When a designer creates an information visualization, they use visual features to represent concepts. This mapping between concepts and visual features is called the ***encoded mapping***. For example, if the designer constructs a weather map in which darker colors signify larger amounts of rainfall, the map would have a "dark-more" encoded mapping. Designers may deliberately define the encoded mapping using their own knowledge, using recommendations from other experts, or using recommender system algorithms [18, 43, 39, 22, 19]. Alternatively, designers may rely on software defaults, which automatically assign colors to concepts in a pre-defined order, regardless of the concepts. In such cases, the encoded mapping is created though the designer's actions, but the designer may not explicitly consider the encoded mapping during visualization design.

Perspective of the observer. When we use the term 'observer,' we do so to refer to anyone who looks at visualizations with the goal of gleaning knowledge from what they see. Observers include the general public looking at public health data in the news, travelers looking at maps to find their way, students looking at diagrams to learn about mathematical or scientific processes in the classroom, and academics who look at charts to learn about the latest discoveries in their fields.

Observers' expectations about how visual features should map onto concepts are called ***inferred mappings*** [39]. As established earlier, it is harder for observers to interpret visualizations when the encoded mapping does not match their inferred mappings, even in the presence of a clear legend [18, 37, 47]. Moreover, when the encoded mapping matches their inferred mappings, observers can more easily interpret the meanings of visual features, even in the absence of a legend ([8, 21, 39, 38, 22]. By understanding the nature of observers' inferred mappings, it is possible to design visualizations that match those expectations and thus facilitate visual communication.

1.2 Chapter overview

In this chapter, we will use color as a lens to discuss factors that influence expectations about the meaning of visual features in information visualizations. We will discuss color semantics (i.e., the meaning of colors) in the context of two general kinds of information visualizations: visualizations of categorical information (Section 2) and continuous data (Section 3).

Historically, studies on inferred mappings discussed separate factors relevant for visualizations of categorical vs. continuous information. However, recent work suggests that they can be understood under a single framework [40], as we will discuss at the end of this chapter.

Defining the scope of artifacts that are considered to be "information visualizations" ("visualizations" for short) is a difficult endeavor (see Fox [9] and Chapter 9 of the present book). Stemming from issues raised in Fox [9], we use "information visualizations" broadly, in reference to external graphical representations (and corresponding verbal labels, if present) created to support visual communication. Here, the term "graphical" pertains to non-verbal markings in which visual features (e.g., color, shape, size, and texture) are used by a designer to communicate their intended message [2]. Although this definition of information visualizations includes visualizations of data (e.g., charts), it extends to any encoding system in which designers use non-verbal visual features to communicate their intended message [51, 53]. For example, an encoding system for recycling bins, in which a designer uses different colors to represent different kinds of trash/recyclables, is considered an information visualization. Using this broad definition enables researchers to identify generalizable psychological principles of how people infer meaning from visual features, which transcend specific design formats.

We aim for this chapter to serve two key purposes. First, it will help readers develop an understanding of psychological factors relevant to visual communication. Second, it will provide designers with knowledge that they can apply to help make visual communication effective and efficient. However, color semantics for visual communication is an active field of research. This chapter presents a snapshot of the field as it is today, but we anticipate that the ideas discussed here will evolve with new discoveries about how people infer meaning from visual features.

2 Color semantics for categorical information

In visualizations of categorical information, discrete colors are used to represent distinct concepts. For example, Figure 2A (top) shows a chart in which distinct colors represent different sectors that emit greenhouse gases and Figure 2A (bottom) shows a chart in which distinct colors represent different kinds of management for drinking water facilities. Visualizations of categorical information can be understood in contrast with visualizations of continuous data. Instead of representing discrete categories, visualizations of continuous data represent gradations of quantity, such as farm size across the world and number of African elephants across Africa in Figure 2B. In this section we will focus on visualizations of categorical information, and we will return to visualizations of continuous data in Section 3.

One way to consider color semantics for categorical information is to focus only on the strength of the association between a color and the concept it encodes. Say, the concepts are watermelon and mango, and the chart is about fruit preferences. Mango is strongly associated with shades of orange and watermelon is strongly associated with shades of red. So, if presented with the bar chart in Figure 3A, observers would easily infer that orange encodes mango and red encodes watermelon.

But, what if concepts do not have specific, strongly associated colors, such as the more abstract concepts in Figure 3B? And, what about cases when multiple concepts have similarly associated colors, such as the recycling related concepts in Figure 3C? If one thinks about inferred mappings only in terms of associations between a single concept and single color, they may conclude that color cannot meaningfully encode concepts under such conditions. However, recent work suggests that color semantics is not so limited [39, 22]. To understand why, we must first draw a distinction between color-concept associations and inferred mappings.

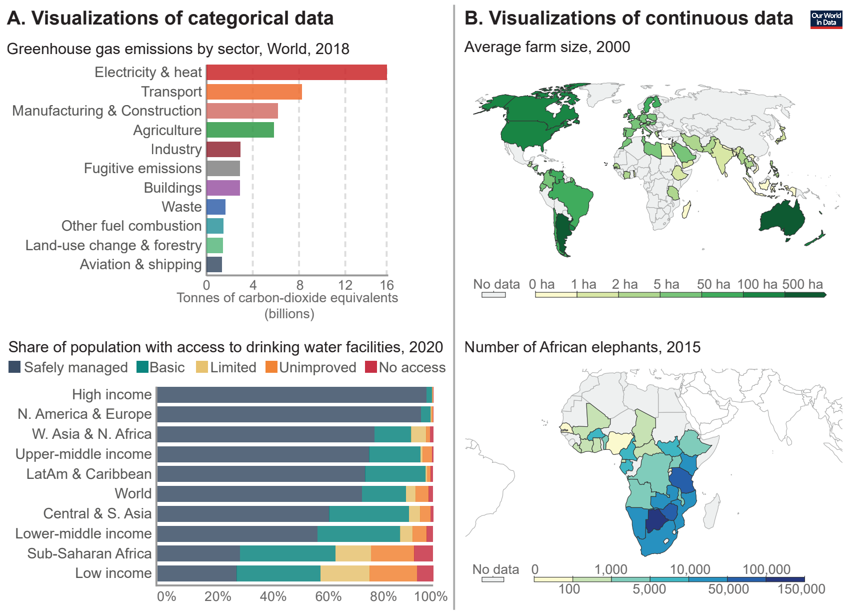

Fig. 2 Examples of visualizations in which color encodes (A) categorical data and (B) continuous data. Figures have been adapted from [33, 32, 30, 31]

2.1 Color-concept associations vs. inferred mappings

It may be tempting to think that people's expectations about the meanings of colors in information visualizations simply depend on the association between an individual color (e.g., yellow) and an individual concept (e.g., banana) represented in the visualization. However, their expectations, or inferred mappings, are far more interesting and complex, as we explain below.

2.1.1 Color-concept associations

Color-concept associations are the degree to which individual colors are associated with individual concepts. Evidence suggests that people form color-concept associations through experiences in the world [41], at least for concepts with directly observable colors. As for more abstract concepts, some have proposed color-concept associations are formed by extension from related concrete objects that do have directly observable colors [36, 48].

For any concept, one can quantify the degree to which that concept is associated with every possible color that a human can perceive. In practice, when researchers measure color-concept associations they sample colors over perceptual color space to make the measurements more tractable [19, 18, 39, 26, 29, 38, 22]. This sample

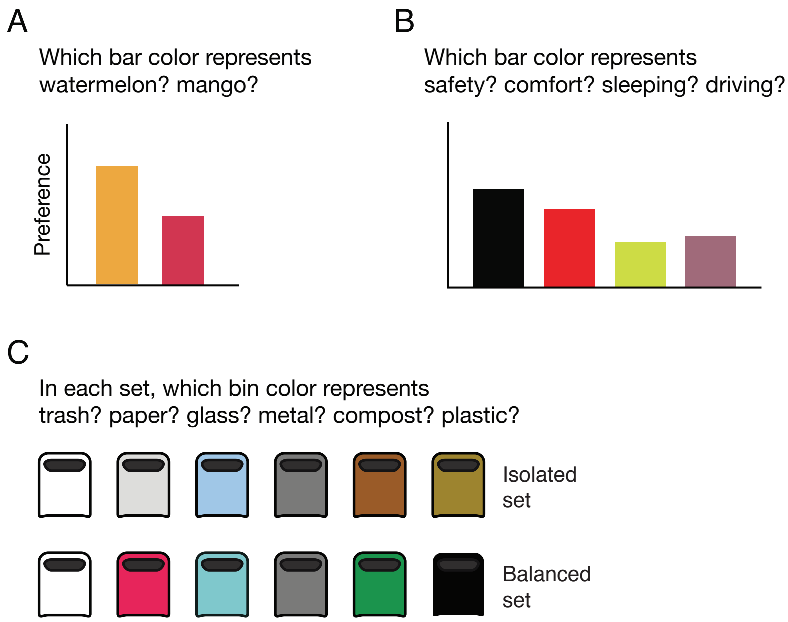

Fig. 3 Examples of visualizations in which colors are used to encode categories, which have been used as stimuli in experiments on inferred mappings. (A) Bar chart representing data about watermelon and mango, which are concepts with strong, specific associations [37]. (B) Bar chart representing data about safety, comfort, sleeping, and driving, which are more abstract concepts with less specific associations [22]. (C) Bins for discarding trash/recyclables, where multiple concepts have similar associations (see Figure 6)[39].

of colors is called the **color library** [22]. Figure 4 shows color-concept associations for the concepts banana, celery, sleeping, and driving [22]. The color library is the "UW-71" colors, which includes 58 colors uniformly sampled over CIELAB color space [29], plus an additional set of light colors required to include saturated yellows (see [22] for details).

Color-concept associations can be measured in multiple ways, including asking people to make judgments of association strength [26, 39, 38, 22, 41] and implementing algorithms that estimate associations from image or language databases [19, 18, 29, 43]. The mean associations in Figure 4 were obtained by presenting participants with a concept at the top of the screen and a color patch below. They rated the association strength between each color and concept on a scale from "not at all" (0) to "very much" (1). Ratings near the middle of the scale (.5) indicate a color was neither strongly associated nor strongly *not* associated with the given concept. For example, in the case of banana in Figure 4, yellows are strongly associated, most blues are strongly *not* associated and greens are in the middle around .5.

Concepts vary in the degree to which they have strong, specific associated colors within a given color library [18, 24], called **specificity** [22].Here we focus on cases when specificity is based on the mean associations across a group of participants, but

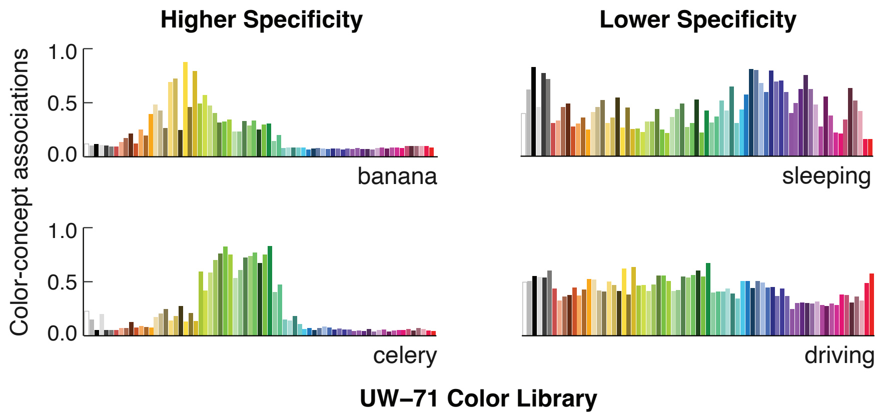

Fig. 4 Mean color-concept association ratings for the concepts banana and celery (higher specificity) and sleeping and driving (lower specificity) from [22]. Data were collected by asking participants to rate how much they associated each color with each concept on a scale from "not at all" to "very much." Thus, the middle of the scale (.5) indicated neutral. The color library was the UW-71 colors, sampled over CIELAB space. Here, the colors are sorted (left to right) according to hue angle and chroma, with the achromatic colors placed leftmost.

specificity could also be defined based on a single person's associations. Concepts have high specificity if they are strongly associated with some colors, and weakly associated with the remaining colors in the color library. For example, Figure 4 shows that celery has high specificity because it is strongly associated with greens, and is weakly associated with the remaining colors. As such, concepts with high specificity have 'peaky' distributions of associations over the color library. In contrast, concepts have lower specificity if they have more uniform distributions over the color library. In a fully uniform distribution, all colors would be equally associated with the concept (i.e., equal bar heights in Figure 4). As shown in Figure 4, the concepts sleeping and driving have lower specificity than banana and celery because their distributions are closer to uniform. Specificity can be quantified using entropy [22, 24], a mathematical measure of the 'peakiness' vs. uniformity of a distribution.

Color-concept associations are essential to interpretations of color meanings in visualizations, but they are only part of the story. This brings us to inferred mappings.

2.1.2 Inferred mappings

Inferred mappings are people's *expectations* about the meanings of each color in an encoding system that maps colors to concepts. Cases arise in which people infer that concepts map to weakly associated colors, even when there are more strongly associated options. To illustrate this point, we will walk through an example from Schloss et al. [39] in which participants inferred the meanings of colors on trash/recycling bins (see Figure 5).

During the experiment, participants were presented with pairs of unlabeled colored bins and were asked which bin was for disposing a target item named at the top of the screen. In some trials the target item was trash (Figure 5A), and in other trials the target item was paper. For each target, participants saw all pairwise combinations of four colored bins (left/right balanced), including white (strongest associate with paper), dark yellow (strongest associate with trash), and red and purple (both weakly associated with trash and paper). The association strengths had been obtained from color-concept association ratings from a different set of participants [39], and are shown in Figure 6. The association strengths for the example trial shown in Figure 5A are represented as a bipartite graph in Figure 5B. In a bipartite graph, edges connect each item in one set (such as colors) to all the items in another set (such as concepts). In this bipartite graph, the circles represent the concepts trash (T) and paper (P), the squares represent the colors purple (Pu) and white (W), and the edge connecting each concept to each color represents the color-concept association strength (thicker indicates stronger associations).

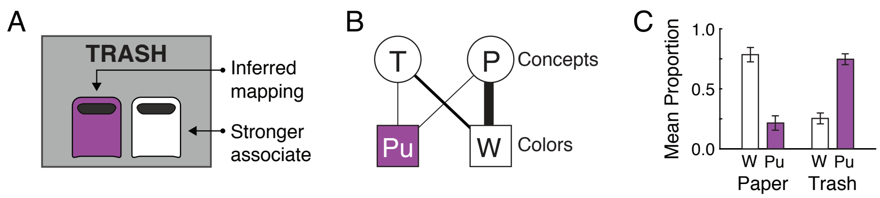

Fig. 5 Illustration of the distinction between color-concept associations and inferred mapping from [39]. (A) Trial in which participants inferred which color mapped to the target concept trash (arrows and labels to the right are for illustration only and were not part of the trial). (B) Bipartite graph showing color-concept association strengths for concepts trash (T) and paper (P) with colors purple (Pu) and white (W). Thicker edges connecting each concept with each color indicate stronger associations. (C) Mean proportion of times participants chose each color when the target was paper or trash (error bars represent standard errors of the means). Participants inferred purple mapped to trash, even though white was more strongly associated with trash.

Schloss et al. [39] considered two possible ways observers might approach this task. The first approach, **local assignment**, is simply to choose the color that is most strongly associated with the target. Local assignment would lead to inferring that white is for trash in Figure 5. The second approach, **global assignment**, is to choose the color that optimizes assignments between all colors and concepts in the encoding system. To determine the optimal assignments in Figure 5, we can compare the total goodness, or "merit" of one possible assignment (e.g., T–Pu/P–W) to the alternative assignment (e.g., T–W/P–Pu) and determine which assignment has greater merit. For now, assume merit is simply color-concept association strength, but we will return to other definitions of merit below. The assignment that pairs trash with purple and paper with white has greater total merit. Thus, the global assignment approach would lead to inferring purple is for trash.

Consistent with global assignment, participants reliably inferred that the purple bin was for trash (Figure 5C), even though trash was more strongly associated with white. This example illustrates the distinction between inferred mappings and associations. Associations serve as the input to global assignment, but further processing leads to people's inferences about the meanings of colors in visualizations. This process is called assignment inference.

2.2 Assignment inference

Assignment inference is the process by which people infer mappings between visual features and concepts in an encoding system [39]. The process was given this name because it is analogous to an *assignment problem* in the field of optimization. Assignment problems are models for assigning items in one set to items in another set in a manner that optimizes **merit**, or the"goodness," of the assignment. [17, 23, 5]. Goodness is defined with respect to a given goal. For example, if the goal is to assign swimmers to strokes in a relay race to minimize time to complete the race, merit is the time it takes each swimmer to complete each stroke. Solving an assignment problem involves finding the *best* pairings such that the overall merit across all pairs is as good as possible (i.e., the total time is as short as possible). The question is, what determines merit in assignment inference?

In our discussion of the trash/paper recycling experiment illustrated in Figure 5, we alluded to the notion of merit in assignment inference as association strength between each color and concept. We explained that global assignment maximizes association strength over possible assignments, even if that means assigning concepts to weakly associated colors when there are more strongly associated options. However, association strength is only one possible way to specify merit, and it is not necessarily the type of merit that humans use in assignment inference.

To study merit in assignment inference, Schloss et al. [39] assumed the role of the designer and created two different color sets (a.k.a. palettes) for trash/recycling bins (Figure 3C). To create the palettes, they used two methods of defining merit, and solved an assignment problem to determine the optimal color set within each definition. The logic of their experiment was that observers would be better at interpreting palettes created using a definition of merit that more closely matched merit in assignment inference. Thus, identifying which palette was easier to interpret would provide insight into the type of merit in assignment inference.

The two color palettes were designed for six types of trash/recyclables (paper, plastic, glass, metal, compost, and trash), using two definitions of merit: isolated merit and balanced merit. Both types of merit were specified as follows, using the color-concept association data shown in Figure 6.

Isolated Merit. Isolated merit for a given color-concept pair is simply the association strength between that color and that concept. It is called "isolated" merit because it is determined by the association between each color and concept in isolation, without accounting for other colors or concepts in the encoding system. When an assignment problem determines the optimal pairings under isolated merit, it se-

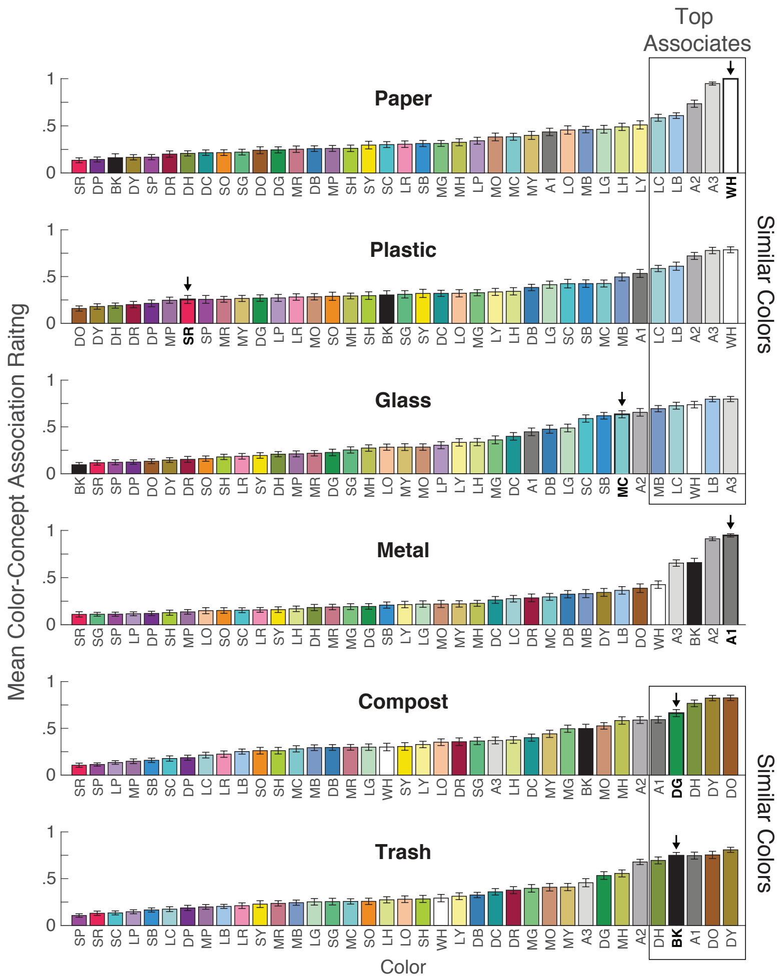

Fig. 6 Mean color-concept association ratings for the Berkeley Color Project 37 (BCP-37) colors and the concepts paper, plastic, glass, metal, compost, and trash (data from [39]). Colors are sorted along the x-axis from weak to strong association. Error bars indicate standard errors of the means. The top associated colors are shared among paper, plastic, and glass, and shared among compost and trash. Arrows point to the optimal colors for each concept using balanced merit.

lects color-concept pairs such that the total association strength across all pairings is as large as possible.

The color palette generated using isolated merit is shown in Figure 7A. Note that paper, plastic, and glass share similar top associated colors, and compost and trash share similar top associated colors (Figure 6). As a result, the colors assigned to those concepts were strongly associated with more than one concept in the encoding system. For example, plastic was associated with its assigned color, light gray (A3), but also was strongly associated with white (WH), the color assigned to paper, and light blue (LB), the color assigned to glass. These observations highlight a potential problem with simply maximizing association strength: it may introduce confusablity when multiple colors are associated with multiple concepts in the encoding system.

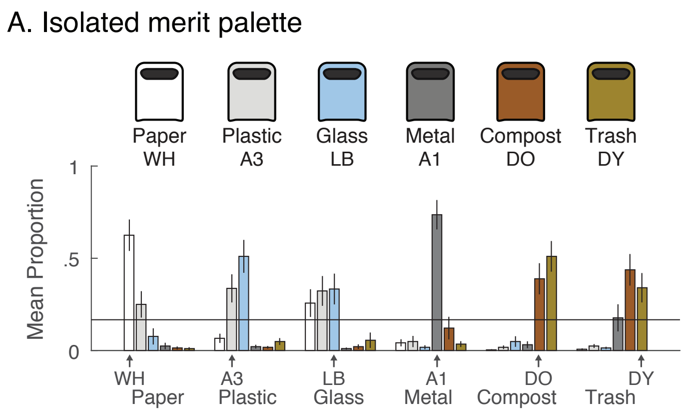

Fig. 7 Color palettes and corresponding plots showing the mean proportion of times participants chose each color for each concept when palettes were generated using (A) isolated merit or (B) balanced merit. Arrows point up to the correct color, specified by the optimal solution to the assignment problem using each definition of merit. Error bars represent standard errors of the means. Data are from [39].

Balanced merit. Balanced merit for a given color-concept pair is computed as the association strength for that color-concept pair, minus the association strength for the color with the next most strongly associated concept. This definition of merit is called "balanced merit" because it balances prioritizing color-concept association strength while avoiding confusability that can arise when a color is strongly associated with multiple concepts in the encoding system. When an assignment problem determines the optimal pairings under balanced merit, it makes the association difference across all color-concept pairs as large as possible.

This method of defining merit can lead to assigning a concept to a weakly associated color, which can occur if the color is more associated with that concept than with all other concepts in the encoding system. For example, the color palette generated using balanced merit is shown in Figure 7B. In this palette, plastic was assigned to red (SR), even though plastic is weakly associated with red, because red is more associated with plastic than with the other concepts (Figure 6).

We note that isolated merit and balanced merit result in the same assignments when there are only two colors and two concepts in the encoding system. However, they can diverge when the number of colors and concepts is larger than two, as in the present experiment.

During the experiment, participants were presented with bins from each palette (between-subjects) along with the list of six concept labels. They were asked to drag the label to the appropriate bin color. Accuracy was specified as the optimal assignment between colors and concepts according to the assignment problem within each source of merit. Figure 7 shows the mean proportion of times participants chose each color for each concept for the isolated merit palette (Figure 7A) and the balanced merit palette (Figure 7B). The optimal color for each concept is indicated by an arrow pointing up to the corresponding bar.

Participants were significantly more accurate for the balanced merit palette than the isolated merit palette, even though some of the associations were weaker in the balanced merit palette. For the isolated merit palette, they showed confusion, especially among white, light gray, and light blue for glass, and among dark orange and dark yellow for compost and trash. For the balanced merit palette, participants consistently identified the correct assignments.

These results suggest that merit in assignment inference is closer to balanced merit than isolated merit. That is, during assignment inference, observers account for the difference in associations, and not just maximal associations when inferring mappings between colors and concepts. These results imply that if a designer aims to create color palettes that are easy for people to interpret, it is better to prioritize association difference rather than association strength.

2.3 Semantic Discriminability

Examining the data in Figure 7, it can be seen that participants were more likely to infer "unique mappings" between colors and concepts in Figure 7B than in Figure 7A. That is, for each concept, there was one color that was chosen more often than all the other colors in Figure 7B, but multiple colors were chosen similarly often in Figure 7A. This ability to infer unique mappings is called **semantic discriminability** [38, 22]. This idea can be understood by analogy with perceptual discriminability. Perceptual discriminability concerns the degree to which observers can see the difference between different colors, whereas semantic discriminability concerns the degree to which observers can discern the difference in meaning between different colors. For a set of colors to be semantically discriminable they must first be sufficiently perceptually discriminable. That is, if colors appear the same, they cannot communicate different meanings [38].

One may presume that semantic discriminability is the same thing as interpretability, but there is an important distinction. Semantic discriminability concerns an observer's inferred mapping, regardless of the encoded mapping specified by the designer. In contrast, interpretability concerns how well observers can discern the encoded mapping specified by the designer. To understand this distinction, consider the bar chart in Figure 8A (left). The chart represents data about the concepts watermelon and mango using two different bar colors, red and orange. Given these two colors and concepts, one would readily infer the mapping that watermelon goes with red and mango goes with orange, not the opposite mapping. As such, these two colors have high semantic disriminability in the context of the concepts watermelon and mango. Now, a designer may choose to encode watermelon using red and mango using orange (matching the observer's inferred mapping), or they may encode watermelon with orange and mango with red (opposite of the observer's inferred mapping). In both cases, semantic discriminability of the colors is the same, but interpretability will be greater for the pairing that matches the observer's inferred mapping (watermelon-red/mango-orange).

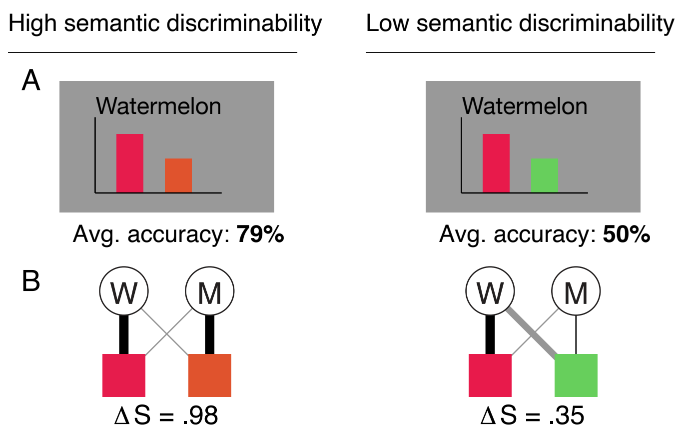

Fig. 8 Color palettes with high vs. low semantic discriminability. (A) Example trials from [38] in which participants inferred which color corresponded to target concept indicated at the top of the screen. Here, the target was watermelon, on other trials the target was mango. The average accuracy is indicated below each example trial. (B) Bipartite graphs showing merit between watermelon (W) and mango (M) and each color, corresponding to the trials above in (A). Black edges correspond to the optimal assignment. Semantic distance (ΔS) for each color pair is indicated below the corresponding bipartite graph.

Schloss et al. [38] developed a method to quantify semantic discriminability using a metric called **semantic distance** (ΔS). Semantic distance is a measure of how likely observers are to infer one assignment over other potential assignment(s) in an encoding system. Figure 8 illustrates examples of color pairs with large and small semantic distance, in the context of the concepts watermelon (W) and mango (M). The bipartite graphs in Figure 8B represent the association strength between each of the two concepts with each of the two colors, corresponding to the visualizations directly above (Figure 8A). The colors in Figure 8 (left) have large semantic distance ($\Delta S = .98$) because the W-red/M-orange assignment has far greater merit than the W-orange/M-red assignment. Even if these associations vary due to noise, the difference in merit between the two assignments is sufficiently large such that W-red/M-orange will remain the optimal assignment (assuming a magnitude of variability that is typical of this kind of association data). The colors in Figure 8 (right) have small semantic distance because the W-red/M-green assignment has only slightly greater total merit than the alternative assignment. If the associations varied somewhat due to noise, the outcome could reverse—the W-green/M-red assignment could have greater merit. For formal details on how semantic distance is computed, see [38].

Having defined semantic distance, the next question is whether semantic distance predicts observers' ability to interpret the meanings of colors in information visualizations. To address this question, Schloss et al. [38] asked participants to interpret the meanings of colors in bar charts with two colored bars, like those in Figure 8A. Each trial had a chart, along with a target concept named above and participants indicated which bar (left/right) they thought corresponded to the target concept. Participants judged many color pairs, which varied in semantic distance and in perceptual distance (i.e., the difference in appearance of the two colors). Responses were scored as "accurate" if they matched the encoded mapping, which was determined as the optimal assignment using balanced/isolated merit (both are the same when there are two colors and two concepts). The charts did not have a legend, so participants did not know which response was correct during the task.

Overall, participants were able to infer optimal mappings, but accuracy increased with increased semantic distance. This effect of semantic distance was independent of effects of perceptual distance. When perceptual and semantic distance conflicted (e.g., high semantic distance, low perceptual distance), semantic discriminability better predicted accuracy. These results suggest that semantic distance does indeed predict observers' ability to interpret the meanings of colors in information visualizations.

2.4 Assignment inference for abstract concepts?

We have established that observers can use assignment inference to interpret optimal mappings between colors and concrete concepts with directly observable colors (e.g., watermelon and mango) [38]. But, is this ability limited to concrete concepts, or might it extend to abstract concepts without directly observable colors (e.g., driving and sleeping)?

Earlier work proposed that some concepts may be "non-colorable," suggesting that such concepts cannot be meaningfully encoded using color [18, 43]. "Colorability" was defined with respect to individual pairs of colors and concepts. Concrete concepts, like banana, celery, grape, and eggplant were considered colorable because they had strong, specifically associated colors (i.e., high specificity), whereas abstract concepts, like sleeping, driving, safety, and comfort, were considered non-colorable because they lacked strong, specific associated colors (i.e., low specificity).

However, this notion of colorability concerns individual concepts alone, and we know from studies on assignment inference that context plays an important role. That is, when inferring mappings between colors and concepts, observers account for all concepts and colors in the scope of an encoding system, not each concept alone (global assignment, see Section 2.1.2). And, their ability to perform assignment inference depends on semantic discriminability of the colors, which concerns the relative associations between all colors and concepts in an encoding system, not just each concept alone. These previous findings imply that observers should be able to use assignment inference to interpret optimal mappings for abstract concepts, insofar as the colors used to encode those abstract concepts are semantically discriminable.

Mukherjee et al. [22] tested this hypothesis in an experiment in which participants interpreted the meanings of colors in visualizations representing data about abstract or concrete concepts[1]. During the experiment, participants were presented with bar charts along with a set of four concept labels, as shown in Figure 9A. Their task was to drag the labels from the top of the chart and place them under the colored bar that they thought corresponded to each concept. Figure 9A shows two example trials, one in which the concepts were all abstract, and the other in which the concepts were all concrete (in other trials abstract and concrete concepts were sometimes mixed).

Each concept appeared in four different concept sets. For example, the concept sleeping appeared with driving, safety, and comfort (set 1), with driving, grape, and banana (set 2), with driving, peach, and cherry (set 3), and with driving, efficiency, and speed (set 4) (Figure 10). For each concept set, the colors of the bars were determined based on the optimal assignment using balanced merit, which selected the four best colors from the UW-71 color library to assign to each of the four concepts.

Overall, participants were able to interpret the optimal mapping between colors and concepts. For example, Figure 9B shows the responses for the stimuli from Figure 9A, plotting the proportion of times participants chose each color for each concept. The arrows below the x-axis point up at the correct color for each concept. Participants consistently chose the correct color for both the abstract and concrete concept sets.

However, the ability to interpret the correct color for a given concept varied depending on semantic discriminability. This relationship is shown in Figure 10. The plots show the proportion of times that participants chose the correct color for the target concepts sleeping (left) and banana (right). Each plot has four points, corresponding to each of the four concept sets in which the target concept appeared. The x-axes represents the semantic discriminability between the correct color and

[1] The abstract concepts had relatively low specificity (close to uniform color-concept association distributions) and the concrete concepts had high specificity (peaky color-concept association distributions), but that correspondence is not always the case (e.g., anger is an abstract concept but has high specificity)

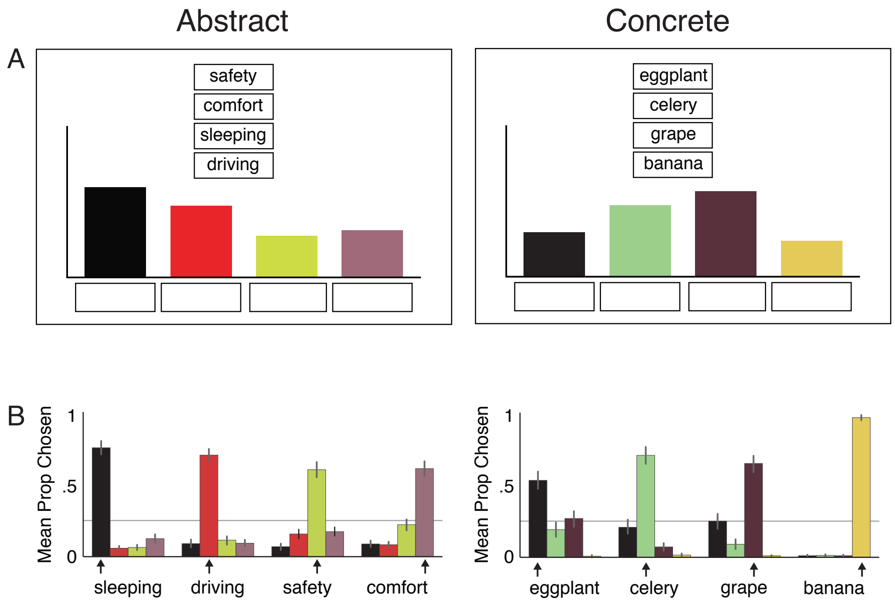

Fig. 9 Examples of (A) experiment stimuli and (B) corresponding data for abstract concepts (left) and concrete concepts (right) from Mukherjee et al. [22]. During the task, participants dragged each concept name to the colored bar that they thought corresponded to each concept. The mean proportion of times participants chose each color for each concept is shown in (B) with arrows pointing up to the correct color for each concept. Error bars represent standard errors of the means and the horizontal gray line represents chance.

the other colors in the palette[2]. The positive slope of the best-fit lines through the points illustrate that accuracy increased with increased semantic discriminability. For example, in set 1, participants were highly accurate at assigning yellow to banana because the other concepts in the set (eggplant, celery, and grape) did not compete with banana for yellow. Yet, in set 4 they were less accurate at assigning yellow to banana because corn competed with banana for yellow. This competition led to yellow being less semantically discriminable from the other colors in set 4 compared to set 1. Figure 10 shows the data for only two out of the 16 concepts tested in the experiment, but the pattern is representative of the full datasets (see [22]).

The results of this experiment suggest that people can use assignment inference to infer optimal mappings for both concrete and abstract concepts. Yet, the ability to do so depends on the semantic discriminability of the colors, which is determined based on all of the colors and concepts in an encoding system. In short, context matters.

[2] Here, when we are discussing semantic discriminability, we are referring to a metric called "semantic contrast". Unlike semantic distance, which quantifies the semantic discriminability of a color palette as a whole, semantic contrast quantifies the distance between a single color and all other colors in the palette. Computational details of these two metrics can be found at [22].

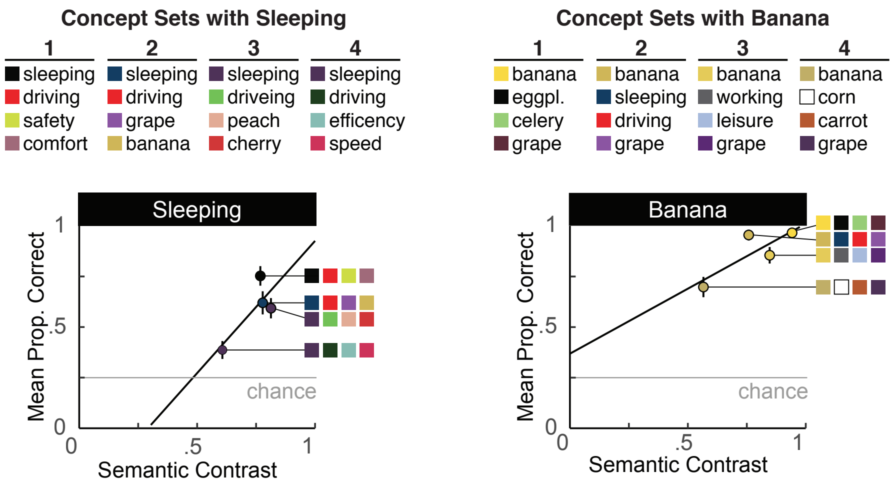

Fig. 10 Top: The four concept sets and color palettes for the concepts sleeping (left) and banana (right) in Mukherjee et al. [22]. Bottom: The proportion of correct responses for the target concepts sleeping (left) and banana (left) as a function of semantic discriminability of the colors in the color palettes. Each point corresponds to each of the four concept sets in which the target concepts appeared. Error bars represent standard errors of the means, and the black lines represent the best-fit lines through the data points.

2.5 Semantic Discriminability Theory

Thus far, we have presented evidence that semantic discriminability is important for interpreting the meanings of colors in visualizations. The next question is, what determines whether it is possible to produce semantically discriminable colors for a set of concepts? To address this question, Mukherjee et al. [22] proposed a theory called **semantic discriminability theory**. The theory states that the ability to produce semantically discriminable colors for a set of concepts depends on the difference between the color-concept association distributions for those concepts.

This theory is illustrated in Figure 11, which shows three pairs of concept sets, one set with very different associations (peach and celery), one with moderately different associations (driving and comfort) and one with very similar associations (eggplant and grape). Below each set of color-concept association distributions is a histogram showing the semantic distance between all pairs of colors in the UW-71 color library for that concept set. Peach and celery, which have a large distribution difference, have many color pairs with high semantic discriminability, and the maximum semantic distance (max ΔS) was a perfect semantic distance of 1. This maximum semantic distance is called the "capacity" for semantic discriminability. Examining the other two concept pairs, driving and comfort (medium distribution difference) have medium capacity for semantic discriminability, and eggplant and grape (small distribution difference) have low capacity for semantic discriminability. Note that eggplant and grape have far higher specificity (peakier distributions) than driving and sleeping, but the capacity for semantic discriminability is lower for the pair eggplant and grape because eggplant and grape have too similar association distributions to produce semantically discriminable colors.

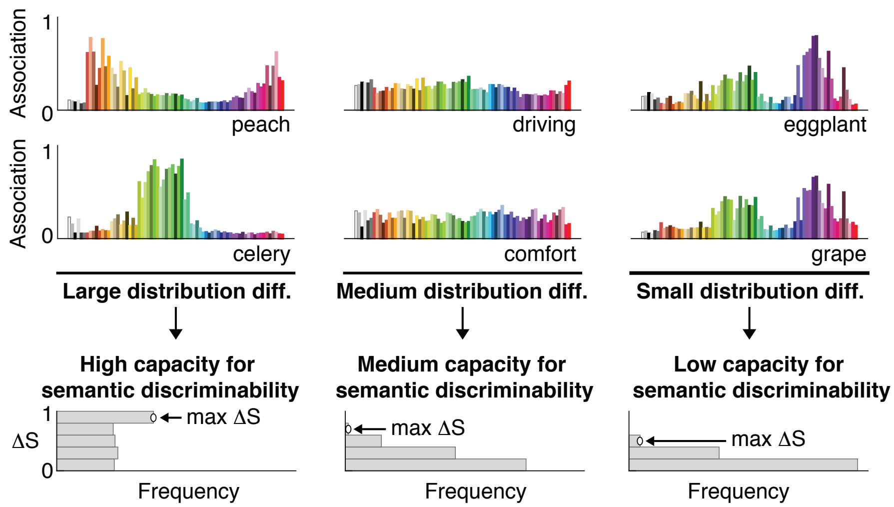

Fig. 11 Color-concept association distributions for concept sets with large, medium, and small distribution differences, which result in high, medium, and low capacities for semantic discriminability, respectively. The top two rows show color-concept association distributions for the 71 colors in the UW-71 color library. The bottom row shows the frequency of color pairs at varying degrees of semantic distance (ΔS). An arrow points to the maximum semantic distance (max ΔS) for each concept set. Figure adapted from [22].

The relation between capacity for semantic discriminability and distribution difference shown in Figure 11 highlights only three concept pairs, but Mukherjee et al. [22] conducted a systematic study of this relationship for all pairwise combinations of 20 concepts (190 concept pairs in total). The concepts included fruits (peach, cherry, grape, banana, apple) vegetables (corn, carrot, eggplant, celery, mushroom), activities (working, leisure, sleeping, driving, eating), and properties (efficiency, speed, safety, comfort, and reliability). In this full dataset, capacity was strongly correlated with distribution difference ($r = .93$). Capacity was also correlated with mean specificity of the individual concepts ($r = .82$), but significantly less so than with distribution difference. When effects of distribution difference and specificity were evaluated in a single model, only distribution difference was a significant predictor of capacity (see [22] for computational details). Aspects of these results for sets of two colors and concepts also extended to sets of four colors and four concepts. These results support semantic discriminability theory, emphasizing the importance of considering the difference between color-concept association distributions, independent of the specificity of each concept's distribution alone.

Semantic discriminability theory was originally formulated and studied with respect to color. However, Mukherjee et al. [22] suggested it as a general theory with potential to extend beyond color to other visual features (e.g., size, shape, texture) and perceptual features in other modalities (e.g., sound, odor, touch).

2.6 Summary and open questions for visualizations of categorical information

We began Section 2 by explaining that the notion of inferred mappings is distinct from color-concept associations. Using assignment inference, observers infer globally optimal assignments between colors and concepts, even if that means assigning a color to a weakly associated concept. We then provided evidence that the ability to perform assignment inference to interpret optimal assignments depends on the semantic discriminability of the colors. Observers can successfully perform assignment inference to interpret optimal assignments for abstract and concrete concepts, as long as the colors representing those concepts are semantically discriminable. Finally, we discussed semantic discriminability, a theory on the constraints for producing semantically discriminable colors for a given set of concepts. Supporting the theory, capacity for semantic discriminability increases with increased differences between the color-concept association distributions for the set of concepts. The series of studies in this section emphasize that people's inference about the meanings of colors are highly context specific, depending on the other colors and concepts in the scope of the encoding system.

Although much has been learned from research on color semantics for categorical information, many open questions are yet to be answered. Here, we highlight two such questions.

Cultural effects? Color-concept associations serve as input to assignment inference, which result in interpretations of the meanings of colors in visualizations [36]. If this input differs due to cultural differences in color-concept associations [13, 14, 49], then the output (interpretation of the meanings of colors) should also differ. However, if the process underlying assignment inference is a general cognitive mechanism, and the input is known, then it should be possible to predict cultural differences in the output. Cross-cultural experiments are needed to test if assignment inference is actually a culturally general cognitive mechanism, and if the current model of assignment inference [39, 22] can predict cultural similarities/differences in inference about the meanings of colors in information visualizations.

This logic extends to semantic discriminability theory. The theory implies that distribution difference will predict capacity for semantic discriminability in any culture, as long as the association distribution data reflect the associations held by a given culture. But, if the color-concept associations collected from one culture are used to predict capacity for another culture that has different color-concept associations, then the predictions might be misleading and the palettes generated might not be semantically discriminable for those who are a part of that second culture. Future research is needed to test whether cultural variations in color-concept association distribution differences predict cultural variations in capacity for semantic discriminability.

Extension to other perceptual features? The work described in this section focused on color, but semantic discriminability theory is broadly defined to apply to other perceptual features in vision (e.g., shape, visual texture, orientation, size), and features in other modalities (e.g., sounds, odors, tactile textures) [22]. However, questions remain as to how to effectively sample perceptual features in these other domains to test this hypothesis, and which other kinds of perceptual features will have systematic and distinct enough associations with concepts to support semantic discriminability.

3 Color semantics for continuous data

In Section 2, we focused on color semantics for visualizations representing categorical information. In Section 3, we turn to factors that contribute to color semantics for visualizations of continuous data, like the colormap data visualizations ("colormaps" for short) from Figure 2B. In colormaps, gradations of color are mapped to gradations of quantities across a spatial representation [12]. The spatial representation could take a variety of forms depending on the type of data, including geographical maps to show climate data across regions of the world, a brain map to show neuroimaging data across different regions of the human brain, or a matrix to show gene expression co-occurrences in different samples of organisms.

Traditionally, the literature has drawn a distinction between the kinds of factors that influence inferred mappings for categorical information and continuous data. For categorical information, the emphasis has been on "direct" color-concept associations (Section 2), whereas for continuous data the emphasis has been on "relational" associations. Direct color-concept associations (or direct associations for short) are just the color-concept associations we discussed in Section 2, but here we call them "direct" associations to distinguish them from "relational associations." Unlike direct associations, which are the degree to which an individual color is associated with an individual concept, relational associations are correspondences between relational properties of visual features and relational properties of concepts [40]. For example, observers have a dark-is-more bias, inferring that darker colors map to larger quantities [21, 8, 4, 37, 47, 40]. The dark-is-more bias is relational because it concerns the relative lightness within a sequence of colors, rather than the lightness of any individual color alone.

Although previous work distinguished factors relevant for visualizations of categorical information and continuous data, recent work by Schoenlein et al. [40] suggests that inferred mappings for continuous data visualized in colormaps are influenced both by direct and relational associations. The relative contribution of these different factors can be understood as different sources of merit in assignment inference. In the following sections, we will first discuss different kinds of relational associations for colormaps, and then explain how relational and direct associations can all be considered as sources of merit in assignment inference for colormap data visualizations.

3.1 Relational associations for colormaps

Several types of relational associations can contribute to inferred mappings for colormap data visualizations (Table 1). The effects of relational associations on inferred mappings are governed by at least two main principles:

1. **Applicability principle:** A relational association can only be activated if it is applicable to the visualization, given the perceptual properties of the visualization.
2. **Combination principle:** If multiple relational associations are activated, they will combine to produce the inferred mapping. Sometimes relational associations work together and sometimes they conflict. When they conflict, they may cancel each other or some relational associations may dominate others, depending on their relative strength.

In the following sections we will discuss empirical evidence for each type of relational association listed in Table 1. In doing so, we will consider perceptual properties that determine whether each relational association applies to a given visualization, and how relational associations combine when multiple are activated at the same time.

Table 1 Types of relational associations between visual features and quantity.

Association type	Description	Related References
Structure Preservation	Structure among perceptual features corresponds to structural properties among concepts to which they are mapped.	[27, 46, 3, 20, 12, 11, 50]
Dark-is-more bias	Regions that appear darker map to larger quantities.	[8, 21, 37, 47, 4, 40]
Opaque-is-more bias	Regions that appear more opaque map to larger quantities.	[37, 1, 35]
Hotspot-is-more bias	Regions closer to the center of “hotspots” map to larger quantities.	[47, 42]
High-is-more bias	Colors higher up on vertically oriented legends map to larger quantities.	[50, 12, 37, 47]

3.1.1 Structure preservation

Structure preservation is a relational association in which structure among perceptual features corresponds to structural properties among concepts to which they are mapped [27, 46, 3, 20, 12, 11, 50]. One example of such structure is the progression of lightness (light to dark) and the progression of quantity (small to large). For example, Figure 12 shows a colormap and four accompanying legends specifying encoded mappings that could correspond to the colormap. The left two encoded mappings are structure preserving because gradations of lightness align with gradations of quantity. From the perspective of structure preservation, both of these encoded mappings (dark-more and light-more encodings) are equally good. However, the right two encoded mappings are not structure preserving because lightness is scrambled with respect to quantity.

Structure preservation is applicable whenever there is structure among the concepts that can be preserved by the visual features that represent those concepts. Structure preservation is always applicable when discussing continuous data because the data have graded structure. Structure preservation is assumed in all of the rest of the relational associations we will discuss next.

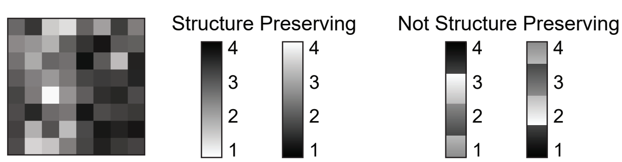

Fig. 12 Example colormap assigning lightness (light to dark) to quantities (1-4) with legends that maintain structure preservation (left) and legends that do not maintain structure preservation (right). Figure adapted from [40].

3.1.2 Dark-is-more bias

The dark-is-more bias leads to the inference that darker colors map to larger quantities [21, 8, 4, 37, 47, 40]. It is applicable when colors in the color scale vary in lightness. When we say “lightness” we mean the perceptual dimension of lightness, going from dark to light (e.g., L* in CIELAB space). We note that in HSB color space, both the “saturation” (S) and ”brightness” (B) dimensions vary in perceptual lightness, so when some discuss color scales defined by saturation variation, there is still lightness variation. Although it is possible for color scales to have no lightness variation (e.g., vary only in hue or perceptual saturation), in practice, color scales tend to vary in lightness, which helps perceive spatial structure in data [34, 52, 15]. Thus, the dark-is-more bias is almost always applicable to inferred mappings for colormaps.

Early evidence for the dark-is-more bias comes from studies in which participants were shown colormaps without legends, and were asked to indicate which regions

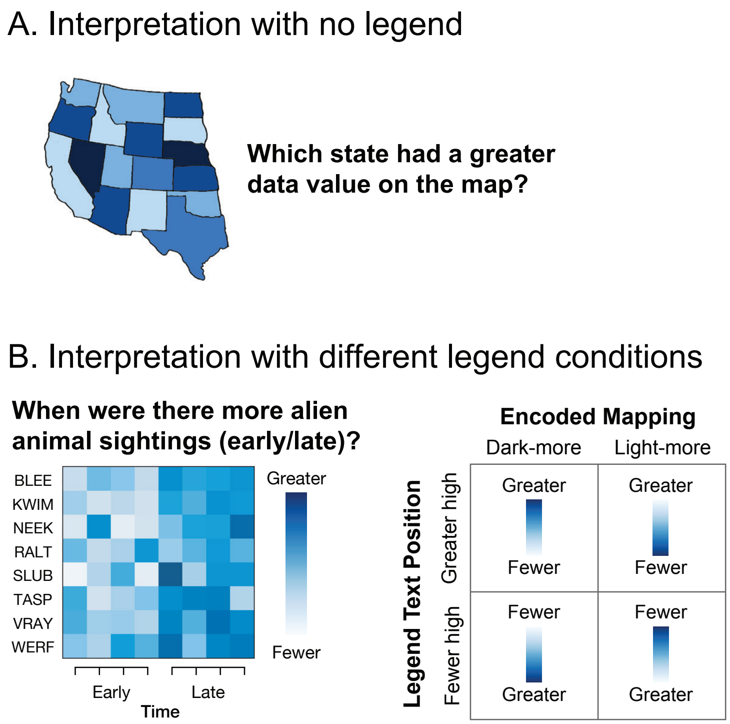

Fig. 13 Types of tasks for assessing inferred mappings for colormaps. (A) Interpretations are made based on the colormap alone, with no legend to specify the encoded mapping (as in [21]). Inferred mappings are assessed by examining the proportion of times each option is chosen. (B) Interpretations are made by reporting the correct answer based on the legend (as in [37]). Inferred mappings are assessed by determining which encoded mappings facilitate faster response times (RTs) to make accurate responses (i.e., encoded mappings facilitate faster RTs if they better match inferred mappings).

represented "more" (Figure 13A) [8, 21]. Participants systematically chose the darker regions, suggesting they inferred that darker colors mapped to larger quantities.

More recent evidence comes from studies in which participants were shown colormaps with legends specifying the encoded mapping. Participants were asked to correctly interpret the colormap according to the legend [37]. On half of the trials, the legend specified dark-more encoding, and on the other half the legend specified light-more encoding (Figure 13B). Also, on half of the trials, "greater" was at the top of the legend and on half of the trials, it was at the bottom. Participants therefore had to read the legend on every trial to determine the encoded mapping. Participants were faster at correctly interpreting the colormap when the legend specified dark-more encoding, providing further evidence for the dark-is-more bias.

We will discuss what happens when the dark-is-more bias combines with each of three other relational associations in the following sections.

3.1.3 Opaque-is-more bias

The opaque-is-more bias leads to the inference that regions appearing more opaque map to larger quantities. This bias is only applicable when regions of the colormap appear to vary in opacity. The percept of opacity variation can be achieved by starting with a colored region and decreasing its alpha in a series of steps so that more and more of the background becomes visible through the region's surface [35]. Functionally, this amounts to interpolating between the color of that region and the color of the background (Figure 14). This interpolation can vary along the perceptual dimensions of lightness, as described above in Section 3.1.2, hue, chroma, or any combination therein.

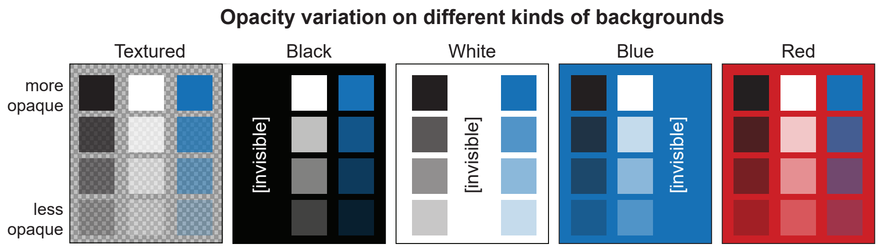

Fig. 14 Black, white, and blue squares are displayed on different backgrounds to show how their appearance changes with opacity variation. The squares in the top row are opaque, and they decrease in opacity in each sequential row below. Colored squares are rendered invisible when they match the color of the background, but they are included in the diagram for completeness.

Apparent opacity variation therefore depends not only on properties of the color scale used to create the colormap, but also properties of the background. Schloss et al. [37] developed a metric for quantifying apparent opacity variation, called the *opacity variation index*. It is computed for a given color scale and background by (1) identifying the endpoint of the color scale that contrasts most with the background, (2) drawing a line between the color of that region and the color of the background region in CIELAB space, (3) calculating the distance between each color on the color scale and and its projection onto the line, and (4) computing the root mean squared error of those distances[3] (Figure 15B and C). This method is only an initial approach to quantifying apparent opacity variation in colormaps, and likely can be improved upon in future work. Nonetheless, it was effective at predicting human performance, as we will discuss next.

Researchers have long considered that the background could have an effect on people's inferred mappings for colormaps, but this notion was framed in terms of contrast with the background [21]. For example McGranaghan [21], presented

[3] As specified in [37], the *opacity variation index* is defined as $log(z+1)$, where z is the root mean squared error between each point on the color scale (square markers in Figure 15B and C) and the line between the highest-contrast color and the background (circle markers in Figure 15B and C). Smaller values correspond to greater perceptual evidence for opacity variation.

participants with partial maps of the United States, with states colored in various shades of blue (Figure 13A). Maps were shown on a white, gray, or black background. McGranaghan hypothesized that participants would infer dark meant more on a light background, but light meant more on a dark background, in a *contrast-is-more* bias. The results showed that participants inferred dark meant more on all three backgrounds, though the effect was weaker on the black background. This was taken as evidence against the existence of a potential contrast-is-more bias.

In a subsequent study examining the effects of the background, Schloss et al. [37] presented participants with colormaps of fictitious data about alien animal sightings on white or black backgrounds. The color scales were standard scales used in visualization (Autumn, Hot, and Gray from MATLAB, and ColorBrewer Blue). As described in Section 3.1.2, each colormap had a legend, and participants were asked to interpret the colormap by reading the legend and indicating whether there were more alien animal sightings early or late in the day (Figure 13B).

The effect of the background lightness depended on the color scale (Figure 15A). For Autumn and Hot, the background had no effect and responses were consistent with a dark-is-more bias on both white and black backgrounds. For ColorBrewer Blue, the background had a moderate effect but responses were still consistent with a dark-is-more bias on both the black and white background (similar to what McGranaghan [21] reported). For Gray, the background had a larger effect that trended toward inferences that lighter colors meant more. The authors were initially puzzled by why the background mattered for some color scales and not others, until they realized that the colormaps differed in how much the regions appeared to vary in opacity. Thus, they developed the opacity variation index described above to test whether these effects could be predicted by apparent opacity variation. Overall, there was a bias for participants to be faster when the legend specified dark-more encoding than light-more encoding (dark-is-more bias), but this was modulated by opacity variation in a manner consistent with an opaque-is-more bias.

This brings us to our first consideration of the combination principle. On a white background, the dark-is-more bias and opaque-is-more bias work together–the darker region is also the more opaque region, so response times were especially fast for dark-more encoding than light-more encoding. On a black background, these two biases conflict–the darker region is the less-opaque, more transparent region. Under such conflicts, if the opacity variation index was strong (Gray color scale), the opaque-is-more bias tended to override the dark-is-more bias when combining to produce the inferred mapping. When the index was moderate (ColorBrewer Blue color scale), the opaque-is-more bias dampened the effect of the dark-is-more bias, but did not cancel it out. This finding aligns with the results reported by McGranaghan [21]. Finally when the index was weak (Autumn and Hot), and therefore not applicable, there was no opaque-is-more bias activated to influence the inferred mapping. One can avoid conflicts between the dark-is-more bias and opaque-is-more bias by either (1) presenting colormaps on light backgrounds, such that the two biases work together, or (2) avoid colormaps that appear to vary in opacity when displayed on a dark background.

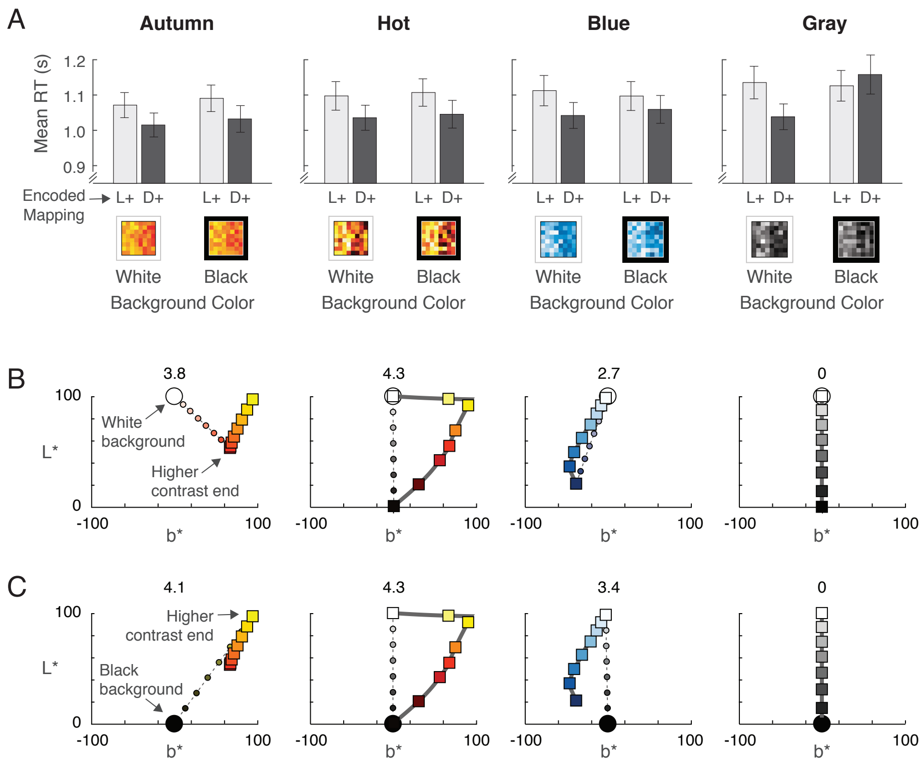

Fig. 15 Opacity variation in colormap visualizations. (A) Mean response times (RT) to correctly interpret dark-more vs. light-more encodings of colorscales varying in opacity when presented on a white vs. black background. Error bars represent standard error of the means. (B) Plots in CIELAB space, showing the colors from each color scale (squares) and the interpolation between the highest-contrast color and the white background (circles). Plots are shown on the plane of L* (lightness) and b* (yellowness/blueness), and the axis for a* (redness/greenness) is not shown. The number above each plot is the opacity variation index. (C) The same as (B), but for a black background. Figure adapted from [37]

3.1.4 Hotspot-is-more bias

The hotspot-is-more bias leads to the inference that regions closer to the center of "hotspots" map to larger quantities [47, 42]. It is applicable when there is spatial structure in the data that looks like a hotspot (e.g., concentric rings), such as in Figure 16.

Until now, in this section we have discussed colormaps in which there was little spatial structure in the data to provide a cue to the locus of larger quantities (e.g., grids of randomly colored squares [37]). However, Schott [42] raised the possibility that color-based biases (e.g., dark-is-more bias) may not influence interpretations of colormaps when there are strong spatial cues to the locus of large quantities, such as hotspots. Hotspots are properties of datasets in which the region with extreme values (very high or very low values) are surrounded by roughly concentric regions with less and less extreme values. These patterns are characteristic of fMRI and EEG signals from neuroimaging data, and storm patterns in meteorological data.

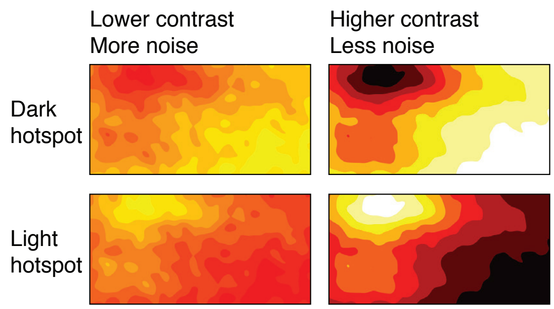

Fig. 16 Colormaps with dark (top) and light (bottom) hotspots. Colormaps on the right have higher lightness contrast and less noise in the underlying dataset than colormaps on the left.

Sibrel et al. [47] tested whether a hotspot-is-more bias exists, and if so, whether it overrides the influence of the dark-is-more bias. They asked participants to interpret colormaps containing hotspots, like those in Figure 16, left. The participants were told the colormaps represented data about alien animal sightings in different regions of a planet, and their task was to press the left/right arrow key to indicate whether there were more sightings on the left or right of the region based on the legend. On half of the trials the hotspot was light, and on half the trials the hotspot was dark (hotspot location and darker region location were left/right balanced across trials). In this initial experiment, participants were faster at responding when the legend indicated dark was more (dark-is-more bias), with no effect of whether the hotspot was light or dark (no hotspot-is-more bias).

Surprised by this result, Sibrel et al. [47] conducted a series of subsequent experiments to see if they could find evidence for a hotspot-is-more bias and to see if they could make it strong enough to override the dark-is-more bias. First they modified the trial structure such that the hotspot was a reliable cue to the locus of the larger quantity. That is, rather than the legend specifying that the colors in the hotspot mapped to more on 50% of the trials, the legend was biased to indicate that the hotspot mapped to more on 77% of the trials. Here, they found evidence for a hotspot-is-more bias. When the hotspot was dark, RTs were faster for dark-more encoding than light-more encoding, consistent with both the dark-is-more bias and the hotspot-is-more bias. However, when the hotspot was light, causing a conflict between the dark-is-more bias and hotspot-is-more bias, the difference in RTs was significantly weaker. Still the hotspot-is-more bias did not override the dark-is-more bias. To get the hotspot-is-more bias to slightly, but significantly, override the dark-is-more bias, it was necessary to not only have the hotspot be a reliable cue, but also to make it even more perceptually salient through increasing lightness contrast and reducing visual noise in the image (Figure 16, right).

These results suggest that color-based biases are powerful contributors to inferred mappings, which cannot be merely dismissed when there is strong spatial structure in the data.

3.1.5 High-is-more bias

The high-is-more bias leads to the inference that colors positioned higher up on a vertically oriented legend map to larger quantities. The high-is-more bias is only applicable when colormaps have vertically oriented legends, which is not always the case in experiments [21] or in practice, as documented by [6]. The high-is-more bias is part of a more general expectation that larger amounts will be displayed higher in space [50, 12].

Evidence supporting the high-is-more bias comes from studies showing that response times to correctly interpret colormaps are faster when "greater" is at the top of the legend than at the bottom [37, 47] (Figure 13B). Moreover, the dark-is-more bias has a larger influence when "greater" is at the top of the legend than at the bottom. One way to view this finding is that when these two biases work together (i.e., the darker region encodes "more" and "more" is represented at the top of the legend) inferences are clearer and interpretation is especially easy, but once these biases conflict, inferences become muddled and interpretation is generally harder.

In Section 3.1, we have highlighted several kinds of relational associations that can contribute to inferred mappings, when they are applicable. We also described what can happen to inferred mappings when different sources of relational associations combine, and which types of relational associations tend to dominate when different types conflict. Ultimately, a goal in this line of work is to construct a comprehensive model to predict people's inferred mappings for information visualizations, while accounting for all applicable factors for a given type of visualization. Next, we discuss an initial step towards such a model.

3.2 Assignment inference for visualizations of continuous data

Until now in this chapter we have discussed distinct factors that contribute to inferred mappings for different kinds of visualizations: direct color-concept associations for visualizations about categorical information and relational associations for visualizations of continuous data. However, recent work by Schoenlein et al. [40] has bridged these areas by extending the framework of assignment inference previously established with visualizations of categorical information (Section 2.2) to visualizations of continuous data. Their approach is illustrated in Figure 17.

During their study, participants were presented with colormaps like those in Figure 17 (left). The colormaps represented fictitious data about environmental concepts, such as the amount of ocean water in different counties. The task was to indicate

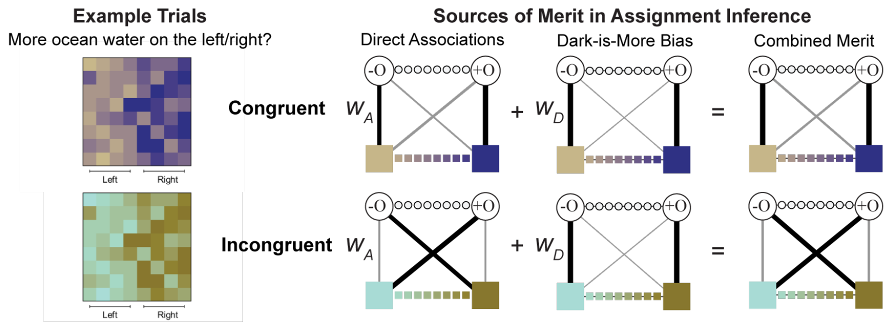

Fig. 17 Example trials from Schoenlein et al. [40] in which participants inferred which region of colormaps (left/right) represented more of the domain concept ocean water. Inferences can be predicted by simulating assignment inference using a weighted combination of multiple sources of merit (direct associations and dark-is-more bias), in cases when they are congruent (top row) and incongruent (bottom row). Figure reproduced from [40].

where there was more of the concept, on the left or right side of the map. There was no legend, so participants responded according to their inferred mappings. For both colormaps in Figure 17, the dark-is-more bias implies participants should infer the darker side represents more ocean water. However, direct associations imply different responses for the top and bottom colormaps. For the top colormap, direct association imply they will choose the darker side because ocean water is more associated with dark blue than with light brown (congruent with the dark-is-more bias). For the bottom colormap, direct associations imply that participants will choose the *lighter* side, because ocean water is more associated with light blue than with dark yellow (incongruent with the dark-is-more bias). How will participants respond?

This problem can be considered through the framework of assignment inference. Direct and relational associations are distinct sources of merit, and inferred mappings are computed over the weighted combination of these two sources of merit. Figure 17 (right) illustrates this scenario using separate bipartite graphs to represent merit from direct associations and the dark-is-more bias. The concepts are the two endpoints of the conceptual dimension (a lot of ocean water, $+O$ and no ocean water, $-O$). The two colors are the two endpoint colors from the color scale used to create the colormap. Although the colormaps included gradations of colors and quantities, the problem was reduced to the two endpoint colors and concepts. This simplification was possible because in their stimuli, association strength and lightness both varied monotonically between the two endpoint colors. Given that there were only two colors and two concepts, merit from direct associations could be treated as association strength between each endpoint color and each endpoint concept (as described for categorical data in Section 2.2). Merit for the dark-is-more bias put greater value on dark-more/light-less edges than light-more/dark-less edges (see [40] for details). The question was, how much weight should be put on direct associations (W_A) vs. the dark-is-more bias (W_D) when combining these sources of merit?

Schoenlein et al. [40] addressed this question by systematically varying the amount of weight put on each source while simulating assignment inference, and determined which weighting best predicted participant's inferred mappings. They found that the best combination of weights placed a .7 weight on direct associations and a .3 weight on dark-is-more bias. This combined weighting was better for predicting participant judgments than weighting on each source of merit alone. With greater weight on direct associations, direct associations overrode the dark-is-more bias when they were in strong conflict. As such, participants inferred that lighter colors mapped to more ocean water in the incongruent example in Figure 17.

This study has set up a method for combining multiple sources of merit to predict inferred mappings in assignment inference. Of course, direct associations and the dark-is-more bias are only two potential sources of merit in assignment inference. But, Schoenlein et al.'s [40] approach can be extended to account for all known direct and relational sources of merit, plus new sources of merit that are yet to be discovered.

3.3 Summary and open questions for visualizations of continuous data

In Section 3 we have discussed multiple factors that influence inferred mappings for colormap data visualizations: structure preservation, dark-is-more bias, opaque-is-more bias, hotspot-is-more bias, high-is-more bias, and direct associations. We have also presented a framework of understanding how to combine multiple (sometimes competing) sources of merit to predict inferred mappings from assignment inference.

Still, many questions remain about the nature of inferred mappings for continuous data, especially with regard to the kind of data being represented and the observers' knowledge about the domain. These questions have been raised in previous work [37, 47, 6, 40], and we summarize them here.

More of what? When colormaps use color to encode quantities, "more" could refer to more of the concept being represented, or more of the numerical values used to measure the concept. For example, when discussing data about response time, researchers often refer to instances in which people were *faster* (i.e., when there was more speed), which corresponds to smaller numbers (i.e., amount of milliseconds). Under such instances, people may infer that darker colors are mapped to faster response times, which correspond to smaller numbers. The question is whether the relational associations reported above, all focusing on what maps to "more", operate at the conceptual or numeric level.

Affects of domain expertise? Some people have expertise working with colormaps in particular domains (e.g., neuroscientists, climate scientists, epidemiologists). Within these domains, conventions arise, which sometimes violate the biases reported above. For example, in neuroimaging, there is a convention to use light-more encodings [6], violating the dark-is-more bias. Questions remain concerning whether domain experts have qualitatively different inferred mappings from novices, and if so, whether those differences are constrained to colormaps in their domain, or generalize to other colormaps on data outside their area of expertise.

Relative contributions of different sources of merit? Schoenlein et al. [40], established the relative weighting to be placed on direct associations and the dark-is-more bias when simulating assignment inference when considering only those two sources of merit. Open questions concern how to construct a comprehensive model that places appropriate weight on each source of merit that is applicable for any given kind of visualization.

Addressing these questions will deepen our understanding of inferred mappings for colormaps, and this knowledge will help design colormaps that facilitate interpretability.

4 Conclusion

A central goal in the psychology of information visualization is understanding people's inferences about the meanings of visual features in visualizations. If visualizations are designed in a manner that aligns with people's expectations, then people can spend less cognitive resources on figuring out what the visual features mean, and focus their effort on figuring out how to use the information presented in visualizations to think about and act on the world around them.

It may be tempting to seek out prescriptive rules for how to use color to convey meaning (e.g., use color x to always mean y). However, as discussed in this chapter, inferences about the meanings of colors are context dependent, contingent on the other colors and concepts in the encoding system, as well as spatial properties (e.g., hotspot structure, height in space). Thus, fully anticipating people's expectations about the meanings of colors in visualizations will require a comprehensive model that accounts for all factors influencing inferred mappings. Initial steps toward this end are showing promising results, but there is much more exciting work to be done.

Although we do not yet have a comprehensive model, designers can still use the results discussed in this chapter to inform their designs. For example, evidence suggests that for visualizations of categorical information, it is better to use color palettes that maximize association difference rather than association strength. Ultimately, when selecting colors for visualizations, we advocate for learning as much as possible about the various factors that can influence people's expectations about the meanings of colors. Then, use critical thought to consider which factors are most relevant for a particular visualization, and how to leverage them in a manner that makes sense for the design as a whole.

By deepening the understanding of color semantics, this field of research is providing insight into the human ability to translate perceptual input into knowledge about the world, while providing insight into how to design visualizations that facilitate visual communication.

References

1. A. N. Bartel, K. J. Lande, J. Roos, and K. B. Schloss. A holey perspective on venn diagrams. *Cognitive Science*, 46(1):e13073, 2021.
2. J. Bertin. *Semiology of graphics: diagrams, networks, maps*. University of Wisconsin Press, Madison, 1983.
3. J. Blachowicz. Analog representation beyond mental imagery. *The Journal of philosophy*, 94(2):55–84, 1997.
4. C. A. Brewer. Color use guidelines for mapping and visualization. In A. M. MacEachren and D. R. F. Taylor, editors, *Visualization in Modern Cartography*, pages 123–148. Elsevier Science Inc., Tarrytown, 1994.
5. R. Burkard, M. Dell'Amico, and S. Martello. *Assignment problems: revised reprint*. SIAM, 2012.
6. M. Christen, D. A. Vitacco, L. Huber, J. Harboe, S. I. Fabrikant, and P. Brugger. Colorful brains: 14 years of display practice in functional neuroimaging. *NeuroImage*, 73:30–39, 2013.
7. W. S. Cleveland and R. McGill. Graphical perception and graphical methods for analyzing scientific data. *Science*, 229(4716):828–833, 1985.
8. D. J. Cuff. Colour on temperature maps. *The Cartographic Journal*, 10(1):17–21, 1973.
9. A. R. Fox. A psychology of visualization or (external) representation? *arXiv preprint arXiv:2009.13646*, 2020.
10. R. L. Goldstone, F. Pestilli, and K. Börner. Self-portraits of the brain: cognitive science, data visualization, and communicating brain structure and function. *Trends in cognitive sciences*, 19(8):462–474, 2015.
11. G. P. Goodwin and P. Johnson-Laird. Reasoning about relations. *Psychological review*, 112(2):468, 2005.
12. M. Hegarty. The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science*, 3(3):446–474, 2011.
13. D. Jonauskaite, A. M. Abdel-Khalek, A. Abu-Akel, A. S. Al-Rasheed, J.-P. Antonietti, Á. G. Ásgeirsson, K. A. Atitsogbe, M. Barma, D. Barratt, V. Bogushevskaya, et al. The sun is no fun without rain: Physical environments affect how we feel about yellow across 55 countries. *Journal of Environmental Psychology*, 66:101350, 2019.
14. D. Jonauskaite, J. Wicker, C. Mohr, N. Dael, J. Havelka, M. Papadatou-Pastou, M. Zhang, and D. Oberfeld. A machine learning approach to quantify the specificity of colour–emotion associations and their cultural differences. *Royal Society open science*, 6(9):190741, 2019.
15. G. Kindlmann, E. Reinhard, and S. Creem. Face-based luminance matching for perceptual colormap generation. In *IEEE Visualization, 2002. VIS 2002.*, pages 299–306. IEEE, 2002.
16. S. M. Kosslyn. Understanding charts and graphs. *Applied cognitive psychology*, 3(3):185–225, 1989.
17. H. W. Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.
18. S. Lin, J. Fortuna, C. Kulkarni, M. Stone, and J. Heer. Selecting semantically-resonant colors for data visualization. In *Computer Graphics Forum*, volume 32, pages 401–410. Wiley Online Library, 2013.
19. A. Lindner, N. Bonnier, and S. Süsstrunk. What is the color of chocolate?–extracting color values of semantic expressions. In *Conference on Colour in Graphics, Imaging, and Vision*, volume 2012, pages 355–361. Society for Imaging Science and Technology, 2012.
20. C. J. Maley. Analog and digital, continuous and discrete. *Philosophical Studies*, 155(1):117–131, 2011.
21. M. McGranaghan. Ordering choropleth map symbols: The effect of background. *The American Cartographer*, 16(4):279–285, 1989.
22. K. Mukherjee, B. Yin, B. E. Sherman, L. Lessard, and K. B. Schloss. Context matters: A theory of semantic discriminability for perceptual encoding systems, 2021.
23. J. Munkres. Algorithms for the assignment and transportation problems. *Journal of the society for industrial and applied mathematics*, 5(1):32–38, 1957.

24. S. Murthy, R. Hawkins, and T. L. Griffiths. Shades of confusion: Lexical uncertainty modulates ad hoc coordination in an interactive communication task. *Cognition*, in press.
25. D. Norman. *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books (AZ), 2013.
26. L.-C. Ou, M. R. Luo, A. Woodcock, and A. Wright. A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *Color Research & Application*, 29(3):232–240, 2004.
27. S. Palmer. Fundamental aspects of cognitive representation. in roach, e. and lloyd, b. b., editors, cognition and categorization. pages 259–303, 1978.
28. S. Pinker. A theory of graph comprehension. *Artificial intelligence and the future of testing*, 73:126, 1990.
29. R. Rathore, Z. Leggon, L. Lessard, and K. B. Schloss. Estimating color-concept associations from image statistics. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1226–1235, 2020.
30. H. Ritchie and M. Roser. Biodiversity. *Our World in Data*, 2021. https://ourworldindata.org/biodiversity.
31. H. Ritchie and M. Roser. Clean water and sanitation. *Our World in Data*, 2021. https://ourworldindata.org/clean-water-sanitation.
32. H. Ritchie and M. Roser. Farm size. *Our World in Data*, 2021. https://ourworldindata.org/farm-size.
33. H. Ritchie, M. Roser, and P. Rosado. Co_2 and greenhouse gas emissions. *Our World in Data*, 2020. https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions.
34. B. E. Rogowitz, L. A. Treinish, and S. Bryson. How not to lie with visualization. *Computers in Physics*, 10(3):268–273, 1996.
35. R. E. Roth, A. W. Woodruff, and Z. F. Johnson. Value-by-alpha maps: An alternative technique to the cartogram. *The Cartographic Journal*, 47(2):130–140, 2010.
36. K. B. Schloss. A color inference framework. In . G. V. P. L. MacDonald, C. P. Biggam, editor, *Progress in Colour Studies: Cognition, Language, and Beyond*. John Benjamins, Amsterdam, 2018.
37. K. B. Schloss, C. C. Gramazio, A. T. Silverman, M. L. Parker, and A. S. Wang. Mapping color to meaning in colormap data visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):810–819, 2019.
38. K. B. Schloss, Z. Leggon, and L. Lessard. Semantic discriminability for visual communication. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1022–1031, 2021.
39. K. B. Schloss, L. Lessard, C. S. Walmsley, and K. Foley. Color inference in visual communication: the meaning of colors in recycling. *Cognitive Research: Principles and Implications*, 3(1):5, 2018.
40. M. A. Schoenlein, L. Campos, J, K. J, Lessard, L, Schloss, and K. B. Unifying effects of direct and relational associations for visual communication. *IEEE Transactions on Visualization and Computer Graphics*, in press.
41. M. A. Schoenlein and K. B. Schloss. Colour-concept association formation for novel concepts. *Visual Cognition*, pages 1–23, 2022.
42. G. D. Schott. Colored illustrations of the brain: some conceptual and contextual issues. *The Neuroscientist*, 16(5):508–518, 2010.
43. V. Setlur and M. C. Stone. A linguistic approach to categorical color assignment for data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):698–707, 2016.
44. P. Shah and P. A. Carpenter. Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General*, 124(1):43, 1995.
45. P. Shah and J. Hoeffner. Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, 14(1):47–69, 2002.
46. R. N. Shepard and S. Chipman. Second-order isomorphism of internal representations: Shapes of states. *Cognitive psychology*, 1(1):1–17, 1970.
47. S. C. Sibrel, R. Rathore, L. Lessard, and K. B. Schloss. The relation between color and spatial structure for interpreting colormap data visualizations. *Journal of vision*, 20(12):7–7, 2020.

48. C. Soriano and J. Valenzuela. Emotion and colour across languages: implicit associations in spanish colour terms. *Social Science Information*, 48(3):421–445, 2009.
49. D. S. Y. Tham, P. T. Sowden, A. Grandison, A. Franklin, A. K. W. Lee, M. Ng, J. Park, W. Pang, and J. Zhao. A systematic investigation of conceptual color associations. *Journal of Experimental Psychology: General*, 149(7):1311, 2020.
50. B. Tversky. Visualizing thought. *Topics in Cognitive Science*, pages 499 – 535, 2011.
51. B. Tversky, J. B. Morrison, and M. Betrancourt. Animation: can it facilitate? *International Journal of Human-Computer Studies*, 57(4):247–262, 2002.
52. C. Ware. Color sequences for univariate maps: Theory, experiments and principles. *IEEE Computer Graphics and Applications*, 8(5):41–49, 1988.
53. C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, 2^{nd} edition, 2004.