(2003). In W. Kuhn, M. Worboys, & S. Timpf (Eds.), Spatial information theory: Foundations of geographic information science, pp. 316-331. Berlin: Springer.

Testing the First Law of Cognitive Geography on Point-Display Spatializations

Daniel R. Montello, Sara Irina Fabrikant, Marco Ruocco, and Richard S. Middleton

Department of Geography, University of California at Santa Barbara
Santa Barbara, CA 93106 USA
{montello,sara,ruocco,richard}@geog.ucsb.edu

Abstract. Spatializations are computer visualizations in which nonspatial information is depicted spatially. Spatializations of large databases commonly use distance as a metaphor to depict semantic (nonspatial) similarities among data items. By analogy to the "first law of geography", which states that closer things tend to be more similar, we propose a "first law of cognitive geography," which states that people *believe* closer things are more similar. In this paper, we present two experiments that investigate the validity of the first law of cognitive geography as applied to the interpretation of "point-display spatializations." Point displays depict documents (or other information-bearing entities) as 2- or 3-dimensional collections of points. Our results largely support the first law of cognitive geography and enrich it by identifying different types of distance that may be metaphorically related to similarity. We also identify characteristics of point displays other than distance relationships that influence similarity judgments.

1 Introduction

Information visualization may be defined as "the art, science, and technology of making visual representations for knowledge discovery" (by analogy to a definition of cartography offered by [1]). The design and implementation of information visualizations has attracted a great deal of attention from various disciplines working on human-computer interaction [e.g., 2, 3]. However, most of this development has proceeded relatively uninformed by relevant behavioral and cognitive science on human perception and cognition, and with little empirical evidence to support claims that interactive visualization tools indeed amplify people's cognition (good examples of the application of perceptual and cognitive science to cartography and information visualization may be found in [4, 5, 6]). We believe basic-science cognitive research is necessary to establish solid theoretical foundations for the design of effective information visualizations. Here we report research that contributes to this objective.

We are especially interested in a class of information visualizations known as "spatializations" [7, 8]. Spatializations are information visualizations in which nonspatial information is depicted as a spatially extended entity, such as a natural or urban landscape. The use of such spatial metaphors is very common in information visualization. Spatialization involves a two-step process of transformation [9, 10, 11]. First, a geometric and semantic generalization procedure transforms large complex data do-

mains into their basic information components. Mathematical transformations create a semantically defined coordinate system that rearranges a set of data items based on their content and functional interrelationships. Secondly, the spatialized data abstractions are graphically (sometimes *carto*graphically) rendered for visual examination and exploration.

In the experiments we report below, we focus on perhaps the most basic principle for rendering information spatializations—the distance-similarity metaphor. This metaphor states that elements closer together on information displays will be understood by users to be more similar. The *distance-similarity metaphor* reflects a basic principle we call *the first law of cognitive geography,* as expressed in the context of information displays. The first law of cognitive geography states that people believe closer things to be more similar than distant things. We identify this law by analogy to the geographical principle that closer places on the earth's surface actually tend to be more similar, whether in rainfall amounts, landforms, linguistic dialects, or foods consumed. This occurs because closer places tend to interact more with each other (movement of matter, energy, information). As a descriptive generalization and predictive heuristic (even occasionally an explanatory mechanism), this principle is so fundamental to analytic geographic thinking that Tobler [12] dubbed it *the first law of geography.* The first law may extend to information spatializations insofar as spatializations afford experiencing information spaces as if they were geographic spaces (or spaces more generally). Our research may thus be understood as an empirical investigation of the first law of cognitive geography in the context of information displays—are closer features on information displays perceived or conceived to be more similar? If so, what aspect of "closeness" corresponds to similarity, and what is the function relating the two?

Although the distance-similarity metaphor is very appealing intuitively, empirical evidence of its validity in information displays is scarce [see 13]. The few usability studies that do exist are not placed within the context of a coherent theoretical spatial framework. Commonly, semantic similarity is mapped metaphorically onto distance (relative location) in the graphic information space, usually metric distance of the straight-line or direct Euclidean type. The widespread and uncritical application of this metaphor may be problematic when information spaces resemble geographic spaces, such as landscapes [14], natural terrain [15], or urban spaces [16]. Geometry in geographic space is not just Euclidean, and in fact, it is not just metric. In other words, similarity can be graphically suggested in terms of several types of "distance," especially when distance is understood broadly to include a variety of expressions of separation (temporal, topological, etc.). Information spaces will be more usable if they are based on a sound theoretical and empirical framework, including those concerning cognitive aspects of space and place.

Our work ties research on information visualization to research on the psychology of distance perception and cognition [e.g., 17, 18, 19]. Distance research has characterized the psychophysical functions describing the relation of physical to psychological distance, identified factors that influence the extent of psychological distance, modeled processes used to estimate distances during perception or from memory, and more. Below we discuss some distance research that bears on the way graphical appearance may affect perceived distance in information displays. Our work also connects information visualization to a large body of research on the psychology of simifactors that influence judgments of similarity, modeled how different factors are combined or given differential attention as a function of context or training, modeled the role of similarity relationships in the categorical organization of stimuli, and more. Although this research has much to say about similarity judgments made from information displays, it does not specifically address in detail how people interpret different forms of distance in terms of representing similarities.

Alternatively, one may consider information displays from the perspective of graphic perception. The displays may be meant to suggest or represent landscapes, but they are small graphic displays, either static or dynamic. Phenomena that influence the perception of distance or similarity among graphic elements in pictorial displays might be expected to influence judgments about the entities represented metaphorically by the graphic elements. One such phenomenon that has been studied systematically for well over a century is the *filled-space illusion* [23, 24]. Intervals separating graphical elements that are empty (blank) appear shorter than do intervals containing intervening elements. If distance is interpreted by users to signify similarity in an information display, we expect filled intervals to appear longer and lead to judgments of lower similarity. However, though the effect is fairly reliable, it is modest in strength and would not be expected to affect similarity much in situations where the interval distances being compared are fairly different. Another relevant perceptual phenomenon is the *vertical illusion*, also called the *horizontal-vertical illusion* [25, 26]. This refers to the fact that vertically-oriented lines in the visual field appear longer than horizontal lines of the same length. Again, if distance is interpreted to signify similarity, we expect the vertical illusion to bias judgments of similarity accordingly.

Perceptual phenomena may influence judged similarity not just through an influence on perceived distance but through an influence on perceptual grouping. As discussed by the Gestalt school of perceptual psychologists, including Wertheimer and Koffka [reviewed in 27, 28], the arrangement of features in a picture or graphical image will influence the perceived thematic or group membership relations of elements. The *laws of perceptual organization* describe stimulus conditions under which separate elements in graphical displays will be grouped into coherent higher-order features. Under certain conditions, for example, a concentration of points will be seen as a cluster or object. Those conditions include various aspects of graphical similarity, proximity, continuation, familiarity, and so on. If graphical elements are perceived to be part of a higher-order feature, users may make a categorical or typological inference that the separate elements are more similar (they may or may not perceive them to be closer). We refer to this as the *emergent-feature effect.*

In this paper, we present results from two experiments on perhaps the simplest of spatialization metaphors—an area or cloud of points, each point meant to signify an information-bearing entity such as a book or news story. We call this a *point-display spatialization.* In both experiments, we examined the way people interpret information displays consisting of points. We wanted to find out how people would interpret the graphical layout of the displays in order to arrive at judgments of semantic similarity among the documents. On different trials, participants viewed and made similarity judgments while viewing different point displays; the different displays varied the spatial relationship between two pairs of comparison points (documents) and varied the context provided by other points in the display. In addition to the point-display

works, regions, and network-region hybrids. But in order to give the present paper a clear focus and manageable size, we focus exclusively on the point displays here. Results for the additional display types will be reported in a forthcoming paper.

2 Experiment 1

In our first experiment, we investigated how users interpret point-display spatializations. In addition to an initial exploration of how users interpret such displays, we wanted to test our data-collection methods, including our similarity comparison request and rating scale. We showed research participants computer displays of simple point displays, explaining that the points represented documents. Three of the document points were labeled 'A,' '1,' and '2;' participants were asked to compare the similarity of A and 1 to the similarity of A and 2. A 9-point scale was provided for them to express their judgments of similarity on an interval scale. Thus, the main task of the experiment involves comparing the relative similarity of two pairs of comparison document points. We varied the point displays across trials with respect to the direct distances between the comparison points and the possible emergence of a cluster feature.

2.1 Methods

2.1.1 Participants. Forty-four students (25 males and 19 females) from an undergraduate regional geography class took part in the experiment, with a mean age of 21.0. They received a small amount of course credit in return for their participation.

2.1.2 Materials. Participants viewed computer displays composed of black points, created using ESRI ArcMap®. Each point was intended to represent a single document in a digital database. In each display, three points to be compared for similarity were labeled with red text as 'A,' '1,' and '2' (see Figure 1). Participants were prompted to "compare the similarity between document A and document 1 with the similarity between document A and document 2." They rated similarity on a 9-point scale ranging left to right from '5' to '1' and then back up to '5' (5-1-5 was transformed to 1-9 for analysis). On the left, '5' was labeled "Documents A and 1 are much more similar to each other." In the middle, '1' was labeled "1 and 2 are equally similar to A." On the right, '5' was labeled "Documents A and 2 are much more similar to each other." In this paper, we refer to the pair of documents A and 1 as 'A:1,' and that of A and 2 as 'A:2.'

Participants viewed 10 different point-display trials in a block (as mentioned above, they also viewed 30 additional trials involving other display metaphors). The point displays were varied to allow comparisons of the effects of different visual aspects of the point displays on judgments of similarity, specifically (1) straight-line distance, (2) vertical illusion, and (3) emergent cluster of points (the potential phenomenon that an aggregation of neighboring points would be perceived as a cluster of related documents). Graphical elements that were not thought to affect similarity judgments between the comparison points (such as the absolute location of the point on the screen) were varied non-systematically.

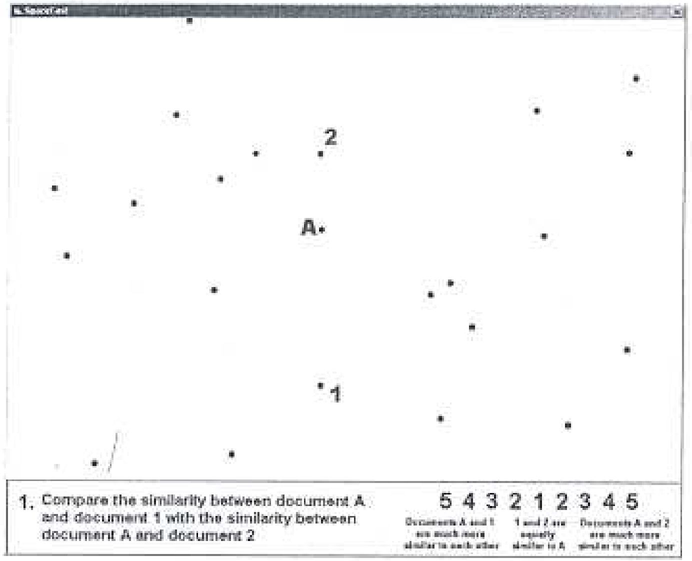


Fig. 1. Sample screenshot from an Experiment 1 trial, showing display, similarity question, and rating scale as they appeared to participants. The 5-1-5 scale shown to participants was transformed into a standard 1-9 scale for data analysis and description

Participants were introduced to the concept of similarity, the style of the trials, and the format of the response scale through three practice trials at the beginning of the test. To avoid priming any particular equivalence between distance and similarity, the practice trials prompted judgments of non-distance similarity (e.g., by asking for a comparison of the similarity of images of a pet dog, a domestic cat, and a tiger). Participants also responded to 11 pre-test questions asking about their personal background, including questions on age, sex, the presence of visual impairments (including specifically color blindness), as well as their formal experience in particular areas such as cartography and GIS. After the main test questions, participants responded to 28 post-test questions, such as how useful they thought each display type was for rating similarity and how easy it was to judge similarities for each display type. Participants also indicated how they had judged similarity and whether the displays reminded them of anything.

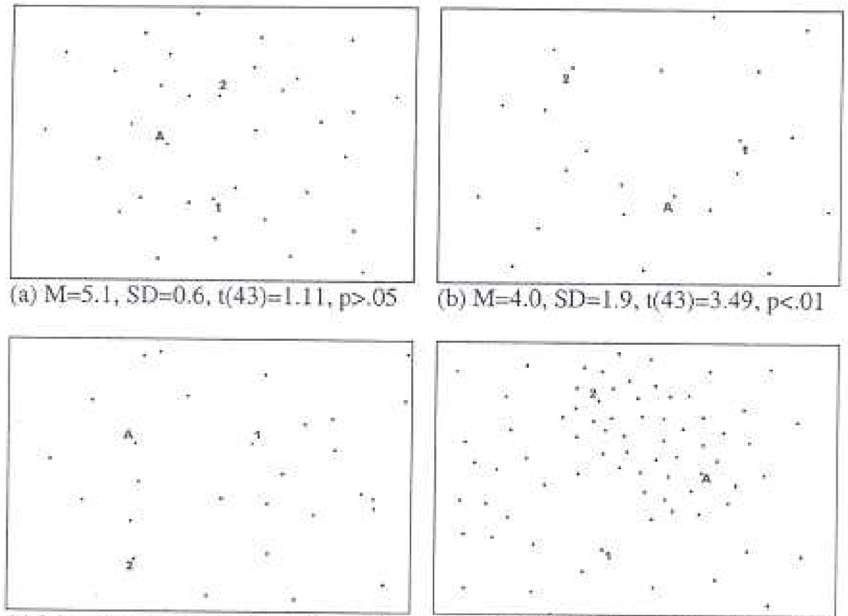
The experiment was administered using a Windows 2000 Pentium III personal computer. The interface was programmed with Microsoft Visual Basic 6.0. Images were projected onto a back-projection screen using an RGB color projector, generating an image size of 1.8 meters wide and 1.4 meters high, at 0.6 meters above the floor. Participants sat at a viewing table 2.7 meters in front of the screen, resulting in a horizontal viewing angle of approximately 37°. A standard mouse and keyboard were used to answer questions. Answers were recorded automatically and stored digitally, including the time required to make similarity judgments. Response time was measured as the elapsed time in milliseconds between when the trial display appeared on the screen and the moment the participant proceeded to the next trial.

2.1.3 Procedure. Prior to each test, participants were briefed that they would be presented with a series of trials about "diagrams that show an information collection from our computer database. The database contains documents such as news stories, books, and journal articles." Participants were told that each document would be shown as a single point. No information was provided on how to judge similarity. Participants were assured that there were no right or wrong answers and were asked not to waste time, as their answers would be timed. Participants then answered the pre-test questions and performed the practice trials. Following that, participants responded to the main test trials organized into four separate blocks (the block of point displays plus three blocks for the other metaphor types mentioned above), so that participants rated all trials of one display type before turning to another type. The block of point trials was presented in a counterbalanced order with the network and region blocks, so that an equal number of participants responded to point trials first, second, or third. All participants responded to network-region trials last. Trials within each block were presented in a different randomized order for each participant. After completing the main test trials, participants answered the post-test questions, were marked down for credit, and thanked for their participation.

2.2 Results and Discussion

Similarity ratings were treated as 9-point interval scales, by scoring a response of '5' to the far left ("A and 1 much more similar") as a '1,' a response of '5' to the far right ("A and 2 much more similar") as a '9,' and a response of '1' in the middle ("1 and 2 equally similar to A") as a '5.' Thus a mean rating less than 5 indicates that participants saw A:1 as more similar, while a mean rating greater than 5 indicates they saw A:2 as more similar. Participants apparently did equate direct distance with (dis)similarity, rating closer document pairs as more similar. Figure 2a shows an example where the actual direct distance between A:1 equals that between A:2. Participants gave this a mean rating of 5.1, which is not significantly different from 5.0, a rating of exactly equal similarity between A:1 and A:2. In contrast, Figure 2b shows an example where the actual direct distance between A:1 is much less than between A:2. On this trial, participants rated the relative similarity as 4.0, which indicates that A:1 is seen as significantly more similar than A:2. In fact, on all four of the point-display trials in which the direct distances between A:1 and A:2 differed, participants rated the closer pair as significantly more similar.

On three of the six point-display trials in which the direct distances between A:1 and A:2 were equal, participants rated the two pairs as not significantly different in similarity. On the other three, therefore, participants rated the two pairs that were actually equally distant as being significantly different in similarity. These trials contradict the basic principle of direct distance equaling similarity. Two perceptual phenomena explain these trials. First is the vertical illusion, in which vertical separations appear longer than horizontal separations. Figure 2c shows an example where the actual direct distance between A:1 equals that between A:2, but A:2 is vertically displayed while A:1 is horizontally displayed. As we anticipated, participants rated the relative similarity as 4.4, which indicates that A:1 is seen to be significantly more similar than A:2. The second relevant perceptual phenomenon is the emergent-feature effect. Figure 2d shows an example where the direct distance between A:1 again equals that between A:2, but A and 2 are members of an emergent cluster of points. In this case, participants rated the relative similarity of the two pairs as 6.7, which indicates that A:2 is seen to be significantly more similar than A:1.



(a) M=5.1, SD=0.6, t(43)=1.11, p>.05 (b) M=4.0, SD=1.9, t(43)=3.49, p<.01

(c) M=4.4, SD=1.8, t(43)=2.21, p<.05 (d) M=6.7, SD=2.2, t(43)=5.13, p<.0001

Fig. 2. Sample displays and similarity ratings for four trials of Experiment 1. The t-scores are tests of significant differences from 5.0, a neutral rating ("1 and 2 are equally similar to A") on the transformed 1-9 scale (rating <5 indicates A:1 are more similar; >5 indicates A:2 are more similar). A neutral rating for equal comparison distances is demonstrated in (a), while the closer document point is rated as more similar in (b). The vertical illusion occurs in (c), while an emergent cluster occurs in (d). (M=arithmetic mean, SD=standard deviation, t=t-score [degrees of freedom], p=probability difference is due to chance)

To interpret these data more systematically, we calculated mean correlations of the ratio of direct distance between A:1 and A:2 *with* rated similarity. A Pearson's correlation coefficient was calculated separately for each participant, with the number of pairs of data points equal to the number of trials (10 in this case). These correlations were positive for 37 of the 44 participants. The correlations were normalized by Fisher's r-to-z transformation [29]. The arithmetic mean of the resulting z-scores was calculated by aggregating across participants. The mean z-score, transformed back into a correlation coefficient, equaled .43 (given how the data were coded, a positive correlation indicates that the closer pair is seen to be more similar). Based on a t-test calculated on the z-scores, this correlation is significantly greater than 0, $t(43) = 6.31$, $p<.0001$. This correlation was equally strong for female and male participants, not significantly differing as a function of participant gender, $F(1, 42) = 0.24$, $p>.05$. The correlation also did not significantly differ as a function of the order in which participants viewed network and/or region displays before viewing point displays did not much influence their interpretation of those point displays.

important, because it demonstrates that whether participants viewed network and/or region displays before viewing point displays did not much influence their interpretation of those point displays.

Participants took a mean of 12.2 seconds (SD = 7.2 s) to respond to each point-display trial (the fastest response was 4.9 s, the slowest was 40.6 s). Female and male participants answered nearly equally quickly; response-time not significantly differing as a function of participant gender, $F(1, 42) = 0.28$, $p>.05$. Response-time did vary significantly as a function of the order in which participants saw the block of point-display trials, $F(2, 41) = 5.58$, $p<.01$. Participants who saw point-display trials third (after the network and region trials) were fastest (8.2 s), those who saw point-display trials second were slowest (16.5 s), while those who saw point-display trials first were intermediate in their speed of response (13.2 s). In general across all types of displays, most participants responded slowest near the beginning of the experiment and sped up after that. For some reason, a few participants answered exceptionally slowly to point trials during the second block.

Finally, we examined participants' self-reports of how they interpreted the point displays. In response to the post-test question "How much did you consider the distance between documents in your answers?," participants marked a mean response of 2.4 on a 3-point scale that included "not at all," "somewhat," and "very much" (one participant did not answer). In order to find out if these self reports actually related to how participants interpreted the displays, we correlated participants' self reports with their distance-similarity correlations (in z-score form). This correlation was .32, a modest yet significant relationship ($p<.05$), suggesting that participants were able to judge their own interpretation strategies with at least some accuracy.

3 Experiment 2

In our second experiment, we attempted to replicate and extend the results we obtained with point-display spatializations in our first experiment. In particular, we examined the distance-similarity relationship by systematically varying distances between the two pairs of comparison documents; the range of variation was rather restricted and varied haphazardly in Experiment 1, with the comparison distances actually equal in 6 of the 10 trials. We also wanted to replicate and further explore the emergent-feature effect we found in Experiment 1, wherein two document points within a cluster of points were seen to be part of a common structure and therefore seen to be more similar. In addition to this emergent cluster, however, we unexpectedly found possible evidence for an emergent-line effect in Experiment 1.Two trials had one or two points intervening between either A:1 or A:2. In at least one case, the similarity ratings suggested that participants may have seen a linear sequence of points as an emergent linear feature, which might have influenced similarity ratings as did the emergent cluster. Finally, in this experiment we varied the size of the viewed displays. If participants compared distances between the document points in order to arrive at similarity judgments, we wondered if only relative distance was important, or whether absolute distance might play a role. The possible role of spatial scale in perceptual and cognitive tasks has been discussed in recent literature [e.g., 30]. Does the scale of the display influence the impact of differing relative distances?

3.1 Methods

3.1.1 Participants. Forty-eight students (27 males and 21 females) from an undergraduate introductory human geography class took part in the experiment, with a mean age of 21.5. They received a small amount of course credit in return for their participation. None had participated in Experiment 1.

3.1.2 Materials. As in Experiment 1, participants viewed computer displays composed of different graphical elements. However, only point and region displays were tested in this experiment (and both black-and-white and colored regions were included). All displays again included black points with three points to be compared for similarity labeled as 'A,' '1,' and '2.' Participants performed the same similarity judgments using the same scale as in Experiment 1.

The block of point displays consisted of 16 trials in this experiment. This block was presented in a counterbalanced order with three other blocks (involving region trials), for a total of 94 test trials. The point displays were again systematically varied to allow comparisons of the effects of different visual metaphors on judgments of similarity. In this experiment, the point displays were varied to allow examination of the effects of (1) straight-line distance, (2) emergent clustering of points, and (3) the number of intervening points between documents (i.e. a possible emergent linear feature). Straight-line distances among points were varied systematically so that the ratio of the distance between A:2 to the distance between A:1 varied from 1.0 (equal) to 1.5, 2.0, 2.5, or 3.0 times as long. To avoid possible confounding by the vertical illusion, comparisons between A:1 and A:2 were matched in orientation on the screen (i.e., both vertical, horizontal, or diagonal).

Participants responded to five practice trials in Experiment 2, similar to those from Experiment 1 but including a couple dealing with color. After the main test trials, participants answered 56 post-test questions including the same 11 questions used in Experiment 1 about their personal backgrounds. The additional post-test questions were adapted from Experiment 1 to account for the new display types.

We used the same equipment and setup as in Experiment 1, with one important addition. In Experiment 2, we varied the size of the projected image of the displays. Two sizes were used, *Small* and *Large*. Half of the participants viewed the small image from a distance of 1.6 meters, projected to be .6 meters wide and .4 meters high, at a height of 1.2 meters above the floor; this resulted in a horizontal viewing angle of approximately 20°. The other half of the participants viewed the large image, which was projected the same as in Experiment 1: 1.8 meters wide, 1.4 meters high, at 0.6 meters above the floor. Viewed from 2.7 meters in front of the screen, the horizontal viewing angle was approximately 37°.

3.1.3 Procedure. Participants were tested exactly as in Experiment 1. The main test trials were again organized into blocks of the same display type, so that participants responded to all trials of one type before proceeding to another type. Block order was fully counterbalanced, and trial orders within blocks were randomized for each participant.

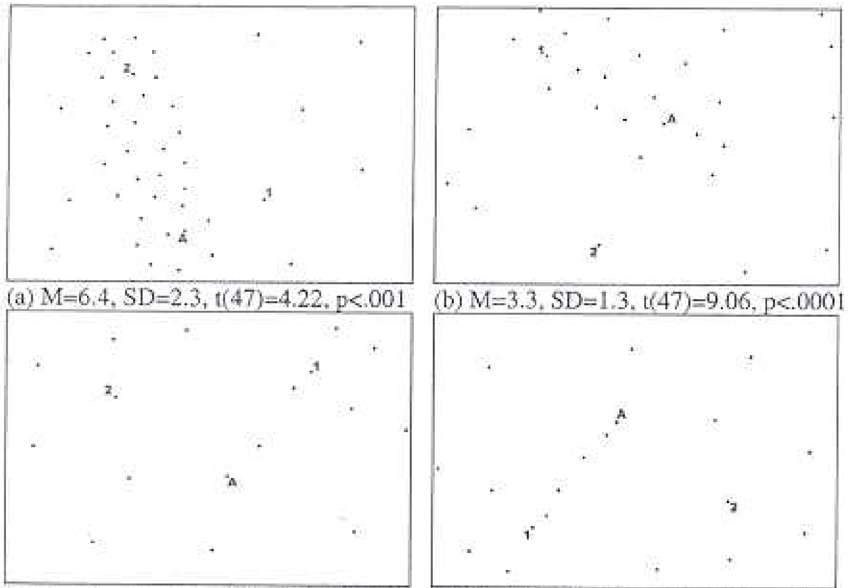
3.2 Results and Discussion

As in Experiment 1, similarity ratings were scored so that a mean rating less than 5 indicates that participants saw A:1 as more similar, while a mean rating greater than 5 indicates that participants saw A:2 as more similar. Because several of the trials in this experiment were specifically designed to examine emergent-feature effects, which we discuss below, we begin here by restricting our examination to the five trials designed only to vary direct distance over relatively unfilled space. On these trials, we systematically varied the ratio of the distances between A:2 and A:1 from a ratio of 1/1 (equal distances) to a ratio of 3/1 (A:2 3 times as far apart as A:1). On the trial where this ratio equaled 1.0, participants in fact rated the relative similarity of the two pairs as 4.9, which is not significantly different from 5.0. For the remaining trials, where the ratio got systematically larger (A:1 got relatively closer), the similarity ratings declined accordingly (approached greater similarity for A:1): 4.4, 3.7, 3.3, 3.4. These were all significantly less than 5.0. Interestingly, the pattern supports a linear relationship between direct distance and similarity across the scale of the experiment, with the exception that the most extreme distance difference between A:2 and A:1 (ratio of 3/1) led to A:1 being rated as no more similar than the next most extreme difference (ratio of 2.5/1). This small difference in relative distance was either imperceptible to participants, or exceeded some threshold at which distance differences no longer influence perceived similarity.

As in Experiment 1, we calculated mean correlations for each participant of direct distance ratios A:1/A:2 with rated similarity across trials, restricting the correlations just to the five trials described above. These correlations were positive for 37 of the 48 participants, again indicating that the closer pair is seen to be more similar. (Correlations for two participants could not be calculated, as they rated all five trials as having equal similarity). The mean correlation (based on z-score transformed values) equaled a robust .70, which is statistically significantly greater than 0, $t(45) = 6.06$, $p<.0001$. This correlation was nearly identical for female and male participants, $F(1, 44) = 0.04$, $p>.05$. The correlation also did not significantly differ as a function of the order in which participants saw the block of point-display trials, $F(3, 42) = 0.37$, $p>.05$. As in Experiment 1, whether participants viewed region displays (with or without colors or black borders) before viewing point displays did not influence their interpretation of those point displays.

We next examined the seven trials specifically designed to test for emergent-cluster effects. These trials placed two of the three comparison document points within a cluster of points in the display. Figure 3a shows an example where the direct distance between A:1 is much less than between A:2, but A and 2 are inside a cluster of points. Participants gave this a mean rating of 6.4, indicating significantly greater similarity of A:2 than A:1. On all of the trials where A shared a cluster with either 1 or 2, but not both, the pair of documents sharing the cluster were rated as significantly more similar, even when they were not as close together. As Figure 3b shows, this emergent cluster effect occurred even when the cluster was formed by a fairly sparse concentration of points.

The final four trials were designed to test for emergent-feature effects of a different kind, namely emergent linear features produced by points intervening between the comparison points. In all four trials, the direct distances between A:1 and A:2 were equal. In one trial, a single point intervened between A and 1. That did not produce an emergent-feature effect, as participants rated the relative similarity of the two pairs as 4.9. Two intervening points in another trial produced no effect either, being rated as 5.1 (Figure 3c). But both three and four intervening points (four shown in Figure 3d) produced a noticeable emergent linear feature: The mean similarities for the two trials were 3.8 and 3.4, both significantly less than 5.0.



(a) M=6.4, SD=2.3, t(47)=4.22, p<.001 (b) M=3.3, SD=1.3, t(47)=9.06, p<.0001

(c) M=5.1, SD=1.9, t(47)=0.36, p>.05 (d) M=3.4, SD=1.9, t(47)=5.83, p<.0001

Fig. 3. Sample displays and similarity ratings for four trials of Experiment 2. The t-scores are tests of significant differences from 5.0, a neutral rating ("1 and 2 are equally similar to A") on the transformed 1-9 scale (rating <5 indicates A:1 are more similar; >5 indicates A:2 are more similar). Emergent-cluster effects are demonstrated in (a) and (b). An emergent-linear effect does not occur in (c) but does in (d); (M=arithmetic mean, SD=standard deviation, t=t-score [degrees of freedom], p=probability difference is due to chance)

Participants took a mean of 8.5 seconds (SD = 3.8 s) to respond to a point-display trial in this experiment (the fastest response was 3.3 s, the slowest was 21.5 s). This is 3-4 seconds faster than in Experiment 1. Female and male participants answered nearly equally quickly, $F(1, 46) = 0.07$, $p>.05$. As in Experiment 1, response times varied significantly as a function of the order in which participants saw the block of point-display trials, $F(3, 44) = 6.12$. Participants who saw point-display trials first were slowest (11.9 s), followed by those who saw them second (8.1 s), while those who saw them third or fourth were nearly identically fast in their responses (6.9 and 7.0 s, respectively). This fits a general pattern in which participants get faster as the experiment progresses, and they get accustomed to the displays and similarity rating task.

In this experiment, participants were assigned to view the displays at one of two sizes, Small or Large. Mean similarity ratings averaged over all trials did not signifi-

Small, 8.6 s for Large). However, display size did influence the strength of the distance-similarity relationship on the five trials that varied direct distance systematically. For these trials, participants viewing Small displays had a mean distance-similarity correlation of .52, while those viewing Large displays had a mean distance-similarity correlation of .83. This difference is statistically significant, $F(1, 44)=4.93$, $p<.05$. Display size influenced the operation of the distance-similarity metaphor; larger images strengthened the correspondence between direct distance and perceived similarity. Display size apparently did not influence the emergence of clusters or linear features; only one of the 11 trials that tested these effects revealed a significant difference in similarity ratings for the two viewing conditions (which is not unlikely, even given only chance variation).

Finally, as in Experiment 1, we examined self-reports of how participants interpreted the point displays. In response to the post-test question "How much did you consider the distance between documents in your answers?," participants marked a mean response of 2.9 on an expanded 5-point scale that included "not at all," "slightly," "somewhat," "very much," and "exclusively." We again correlated participants' self reports with their correlation of direct distance and similarity for the five appropriate trials. This correlation was only .16, a weaker relationship than in Experiment 1 and statistically nonsignificant. However, in response to the post-test question "How much did you consider the presence of documents grouped or clustered around the documents A, 1 and A, 2?," participants marked a mean response of 3.3 on the 5-point scale; these responses correlated .57 ($p<.0001$) with the similarity rating given to the trial shown in Figure 3c, a trial that led to a strong emergent cluster. Similarly, in response to the post-test question "How much did you consider the presence of intervening documents between documents A, 1 and A, 2?," participants marked a mean response of 2.7 on the 5-point scale (one participant did not answer). Responses to this question correlated -.35 ($p<.05$) with the similarity rating given to the trial shown in Figure 3d, a trial that led to a strong emergent linear feature. Both correlations of post-test self-report questions with similarity ratings are in the direction indicating that participants were able to judge their own interpretation strategies with some accuracy.

4 General Discussion and Conclusions

The two experiments reported here investigate how users interpret point-display spatializations to infer similarity relationships. These displays represent information-bearing entities, such as library documents, as a collection of points. The spatial and graphical relations among points are intended metaphorically to represent the semantic similarities among the entities. Probably the most common way this has been done is to locate points so that more similar points are closer together. We have dubbed the notion that people will think closer things are more similar the *first law of cognitive geography*, by analogy to the well-known *first law of geography*. In the context of information displays, this law operates as the *distance-similarity metaphor*, wherein viewers of information displays equate direct (straight-line) distance with the dissimilarity between the reference entities supposed to be represented by the graphical elements

The results of our two experiments provide qualified support for the validity of the first law of cognitive geography and its operation as the distance-similarity metaphor in the interpretation of information spatializations. When people view point representations of documents that appear against a fairly uniform background, either blank or evenly filled with other points, they equate metric direct distance to semantic similarity. This relationship is approximately linear over the ranges we varied in these experiments, though we did uncover evidence of a threshold at which further exaggeration of the contrasts between two comparison distances will not further enhance similarity contrasts. In cases such as these where people apply the distance-similarity metaphor, graphical characteristics that produce phenomena in which perceived distance varies from actual distance will influence judged similarity. An example of such a phenomenon is the vertical illusion, in which vertically-oriented linear features will be perceived as relatively longer than horizontally-oriented features.

However, when the background context against which people view comparison points is not uniform, having uneven groupings of points, visual organization of the points can negate the normal operation of the distance-similarity metaphor. A higher-density collection of points against a sparser background of points is seen as a cluster; if two comparison document points are seen to be inside this cluster, they are seen as more similar, part of a common feature, even if another comparison point is in fact much closer in direct distance. Similarly, a string of points against a background of unorganized points is seen as a linear feature; if two comparison document points are part of this linear feature, they are seen as more similar, again part of a common feature (we showed in Experiment 2 that a set of at least three intervening points, a string of five points, is required to produce the perception of a line). We call these *emergent-feature effects.* They can specifically override the distance-similarity metaphor, leading to a *feature-similarity metaphor* wherein elements that are part of the same feature share greater similarity. In terms of the mechanism of the emergent-feature effect, it is interesting to note that the perceived cluster or linear feature is a phenomenon of visual perception as the Gestalt psychologists described [e.g., 31]. The fact that people viewing the display assign greater similarity to points that are part of the same feature may reflect higher-level *conceptual* organization, however. In particular, such an interpretation by users is consistent with an expression of categorical (particularly regional) cognition [see 32, 33, 34].

Another perceptual phenomenon that may be expected to influence the interpretation of graphical displays is the *filled-interval illusion,* wherein the interval separating comparison points has other points within it and thus appears longer. Either of our emergent features, clusters or lines, might be expected to produce this effect; although we do not yet have data substantiating it, a look at the sample displays in Figures 1 and 2 suggests that this does happen (in 2d, A:2 looks further apart than A:1, even though they are equally distant). It is important to note that the implication of the filled-interval illusion for similarity judgments should be exactly opposite from the implications of the emergent-feature effect. Thus, concentrations of points may produce the appearance of greater distance separating points, but they apparently also lead to the impression of greater similarity, demonstrating a clear dissociation of distance and similarity relationships.

In Experiment 2, we found a previously unreported effect of display size on the relationship of judged similarities to relative distances between the comparison points. For baseline trials where direct distance determined participants' judgments of simi-

larity, the distance-similarity metaphor operated more strongly for participants who viewed the Large display than for those who viewed the Small display; the mean correlation of distance ratios with similarity was a considerable .52 for the Small display but a considerably larger .83 for the Large display. This difference occurred even though the size difference between Small and Large was not as great as it could be—the Large display was just three times wider and filled less than twice the width of the visual field than did the Small display. The result leaves open several interesting questions about its precise mechanism. Is it the difference in sizes between the displayed images that matters (distal size), or the difference between the portions of the visual field filled by the images (proximal size)? Does the larger display exaggerate the perceived distance differences between the points, or does it simply increase the strength of correspondence between relative distances and similarities? These questions call for further research. In any case, this constitutes a novel empirical demonstration that scale matters to the psychology of spatial information [30], and it potentially has important implications for information visualization and communication.

As noted above, the two experiments discussed in this paper included displays based on other spatialization metaphors, specifically networks and regions. In these experiments, along with a third that included only network displays, we have examined the operation of spatial metaphors used to represent similarity relationships. These experiments have also examined the possible role of several graphical variables that are not spatial *per se,* namely various aspects of color (hue, value, and saturation) and line thickness. These results will appear in a forthcoming paper. They are interesting because they demonstrate how nonspatial (nongeometric) aspects can influence the interpretation of information displays. They are also interesting because they allow a comparison of different types of distance other than direct metric distance, such as network distance, that may stand for similarity in information displays.

We are currently finishing a fourth experiment that addresses the distance-similarity metaphor directly by having participants estimate distance instead of just similarity. In this experiment, as in our other experiments, three groups of participants view point, network, and region displays. One group of participants is randomly assigned to rate similarity as in the experiments described in this paper; because they are not told what to base those similarity judgments on, we call this "default similarity." A second group of participants is randomly assigned to estimate distances between points; because they in turn are not told what type of distance to estimate, we call this "default distance." The third group of participants is also randomly assigned to estimate distances between points, but in this condition, they are told to estimate direct or straight-line distance; we call this "direct distance." A comparison of ratings in these three conditions will address some of the issues brought up by the research in this paper, such as the possibility that emergent features can increase perceived similarity among documents at the same time they increase perceived distance •a clear exception to the first law of cognitive geography. These results should help us contribute further to the newly-developing area of scientific research on the use of information spatializations. Research in this area will help address basic questions about the cognitive processing of information displays and help inform the design of human-information interfaces that are more effective, efficient, accessible, and enjoyable to use.

Acknowledgements

We thank Amy Linker and Heather Alexander for help with data collection, and the UCSB students who provided data by serving as research participants in the experiments. We also thank David Mark and Corinne Jörgensen, who have collaborated on this research program since its inception. Three anonymous reviewers provided feedback that improved the paper. The National Imagery and Mapping Agency generously provided funding to support this research.

References

1. International Cartographic Association (ed.): Multilingual Dictionary of Technical Terms in Cartography. ICA, Stuttgart (1973)
2. Card, S.K., Mackinlay, J.D., Shneiderman, B.: Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann, San Francisco (1999)
3. Spence, R.: Information Visualization. Addison Wesley, Boston (2001)
4. Fabrikant, S.I.: Evaluating the Usability of the Scale Metaphor for Querying Semantic Spaces. In: Montello, D.R. (ed.): Spatial Information Theory: Foundations of Geographic Information Science. Lecture Notes in Computer Science, Vol. 2205. Springer, Berlin (2001) 156-172
5. MacEachren, A.M.: How Maps Work: Representation, Visualization, and Design. Guilford Press, New York (1995)
6. Slocum, T.A., Blok, C., Jiang, B., Koussoulakou, A., Montello, D.R., Fuhrmann, S., Hedley, N.R.: Cognitive and Usability Issues in Geovisualization. Cart. & Geog. Infor. Sci. 28 (2001) 61-75
7. Couclelis, H.: Worlds of Information: The Geographic Metaphor in the Visualization of Complex Information. Cart. & Geog. Infor. Sys. 25 (1998) 209-220
8. Kuhn, W., Blumenthal, B.: Spatialization: Spatial Metaphors for User Interfaces. Depart. Geoinfor., Tech. Univ. Vienna (1996)
9. Fabrikant, S.I., Buttenfield, B.P.: Formalizing Semantic Spaces For Information Access. Annals Assoc. Amer. Geog. 91 (2001) 263-280
10. Shepard, R.N.: Representation of Structure in Similarity Data—Problems and Prospects. Psychometrika 39 (1974) 373-422
11. Skupin, A.: From Metaphor to Method: Cartographic Perspectives on Information Visualization. In: IEEE Symposium on Information Visualization, InfoVis 2000, Salt Lake City, UT (2000) 91-97
12. Tobler, W.R.: A Computer Movie Simulating Urban Growth in the Detroit Region. Econ. Geog. 46 (1970) 234-240
13. Fabrikant, S.I.: Spatial Metaphors for Browsing Large Data Archives. Unpublished Ph.D. Diss., Depart. Geog., Univ. Colorado-Boulder (2000)
14. Chalmers, M.: Using a Landscape Metaphor to Represent a Corpus of Documents. In: Frank, A.U., Campari, I. (eds.): Spatial Information Theory: A Theoretical Basis for GIS. Lecture Notes in Computer Science, Vol. 716. Springer-Verlag, Berlin (1993) 377-390
15. Wise, T.A.: The Ecological Approach to Text Visualization. J. Amer. Soc. Infor. Sci. 53 (1999) 1224-1233
16. Dieberger, A., Frank, A.U.: A City Metaphor for Supporting Navigation in Complex Information Spaces. J. Visual Lang. Comp. 9 (1998) 597-622
17. Berendt, B., Jansen-Osmann, P.: Feature Accumulation and Route Structuring in Distance Estimations - An Interdisciplinary Approach. In: Hirtle, S.C., Frank, A.U. (eds.): Spatial Information Theory: A Theoretical Basis for GIS. Lecture Notes in Computer Science, Vol.

18. Hartley, A.A.: Mental Measurement in the Magnitude Estimation of Length. J. Exp. Psych.: Human Perc. Perf. 3 (1977) 622-628
19. Montello, D.R.: The Perception and Cognition of Environmental Distance: Direct Sources of Information. In: Hirtle, S.C., Frank, A.U. (eds.): Spatial Information Theory: A Theoretical Basis for GIS. Lecture Notes in Computer Science, Vol. 1329. Springer-Verlag, Berlin (1997) 297-311
20. Goldstone, R.L.: Similarity, Interactive Activation, and Mapping. J. Exp. Psych.: Learn., Mem., Cog. 20 (1994) 3-28
21. Medin, D.L., Goldstone, R.L., Gentner, D.: Respects for Similarity. Psych. Rev. 100 (1993) 254-278
22. Tversky, A.: Features of Similarity. Psych. Rev. 84 (1977) 327-352
23. Buffardi, L.: Factors Affecting the Filled-Duration Illusion in the Auditory, Tactual, and Visual Modalities. Perc. & Psychophys. 10 (1971) 292-294.
24. Thorndyke, P.W.: Distance Estimation from Cognitive Maps. Cog. Psych. 13 (1981) 526-550
25. Amstrong, L., Marks, L.E.: Differential Effects of Stimulus Context on Perceived Length: Implications for the Horizontal-Vertical Illusion. Perc. & Psychophys. 59 (1997) 1200-1213
26. Gregory, R.L.: Eye and Brain: The Psychology of Seeing (3rd edn.). McGraw-Hill, New York (1978)
27. Goldstein, E.B.: Sensation and Perception (3rd edn.). Wadsworth, Belmont, CA (1989)
28. Gregory, R.L. (ed.): The Oxford Companion to the Mind. Oxford University Press (1987)
29. Cohen, J., Cohen, P.: Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Lawrence Erlbaum Ass., Hillsdale, NJ (1975)
30. Montello, D.R.: Scale and Multiple Psychologies of Space. In: Frank, A.U., Campari, I. (eds.): Spatial Information Theory: A Theoretical Basis for GIS. Lecture Notes in Computer Science, Vol. 716. Springer-Verlag, Berlin (1993) 312-321
31. Masin, S.C.: Absolute and Relative Effects of Similarity and Distance on Grouping. Perc. 31(2002) 799-811
32. Friedman, A., Brown, N.R.: Reasoning about Geography. J. Exper. Psych.: Gen. 129 (2000) 193-219
33. Hirtle, S.C., Jonides, J.: Evidence of Hierarchies in Cognitive Maps. Mem. & Cog. 13 (1985) 208-217
34. Huttenlocher, J., Hedges, L.V., Duncan, S.: Categories and Particulars: Prototype Effects in Estimating Spatial Location. Psych. Rev. 98 (1991) 352-376