The Distance-Similarity Metaphor in Region-Display Spatializations

Sara Irina Fabrikant
University of Zurich

Daniel R. Montello
University of California at Santa Barbara

David M. Mark
University at Buffalo, SUNY

Region-display spatializations represent documents metaphorically as points within regions. Semantic interrelatedness is expressed by some combination of interpoint distance and region membership. In two experiments, the authors investigate judgments of document similarity as a function of these variables. Distance matters, but region membership largely determines judged similarity; hue further modifies it.

Spatial metaphors are a popular approach to visualizing information in such forms as information worlds, information spaces, and cyberspaces. Spatial metaphors can also describe user interactions with these products—for example, navigating the WWW, traversing document spaces, and flying over information landscapes. These expressions reflect an intuition that users can explore and understand abstract information spaces as if they were real geographic spaces.

According to the distance-similarity metaphor[1]—one of the most popular spatial metaphors in information visualization—similar entities in a display should be placed closer together because users will interpret closer entities as being more similar. Explicit or implicit belief in the distance-similarity metaphor justifies the notion that more similar documents should be placed near each other in a spatialized document archive. Figure 1 illustrates this principle. It depicts a portion of the Open Directory Project (ODP), a large human-edited Web site directory that uses a treemap spatialization method.[2] A treemap's purpose is to help people visually explore hierarchically organized data, such as file and directory structures on computer operating systems, or various hierarchical databases available through the Internet. Using the distance-similarity metaphor, Web sites depicted in Figure 1 (red and white dots) whose contents are more similar are closer together in the display, whereas less similar sites are farther apart.

Another spatial metaphor at work in Figure 1 is the region metaphor, which reflects the data's hierarchical nature. Grouping similar documents into homogeneous thematic zones or clusters emphasizes them visually. The clusters, spatially semicontiguous zones in various color shades, show the themes at a particular hierarchical level. Thematically different regions are shaded in different hues (blue and green, for example). We call any spatialized display that applies the region metaphor a *region-display spatialization*. Cartographers use color hue to depict categorical differences in geographic data, such as in soil and election maps (see the "Background in Spatialization" sidebar on page 36). However, as Figure 1 shows, categorically different region pairs—"computers" and "society," for example—share the same color hues. This violates the cartographic principle for thematic maps that unique symbols should denote category membership.

Our previous studies showed how the distance-similarity metaphor operates in the context of point,[1] network,[3] and surface display spatializations.[4] This article reports results of our empirical investigations of the metaphor's effectiveness in region-display spatializations. We try to shed light on how the spatial metaphor of region membership (in monochrome regions) and the nonspatial metaphor of color hue (that is, colored regions) can affect the distance-similarity metaphor's operation in spatialized region displays.

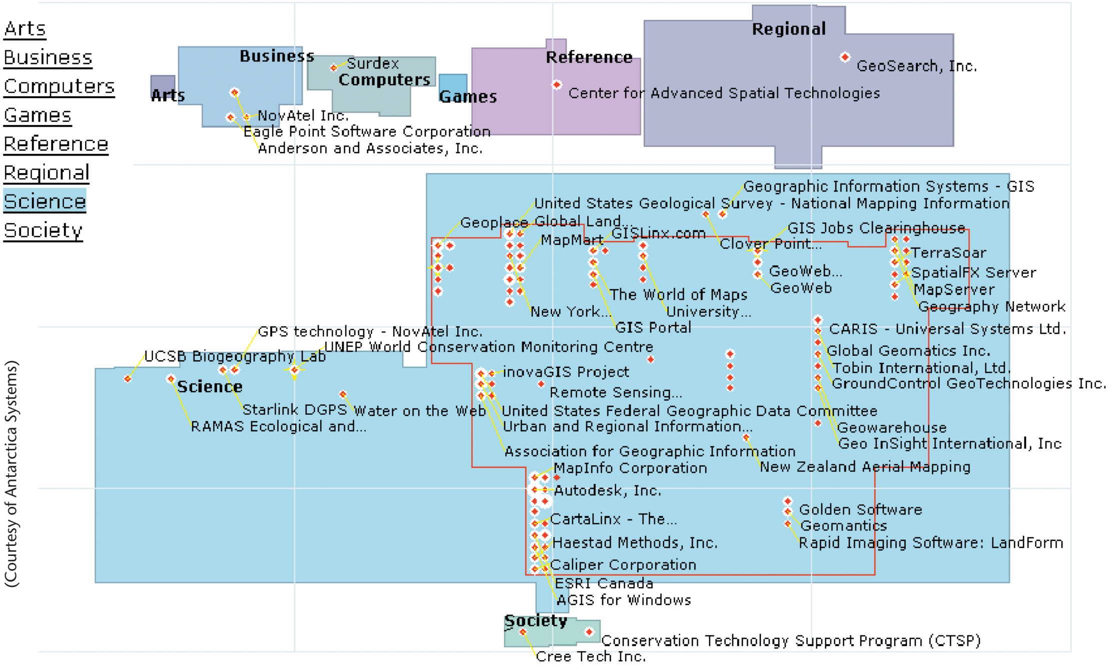
Experiments

We conducted two experiments on nonexpert users' interpretation of the distance-similarity metaphor in region-display spatializations such as those shown in Figure 1. Each point in the display represents an information-bearing entity such as a book, Web site, or news story. Their organization within regions might or might not suggest something to users about their interrelationships. In different trials, participants made similarity judgments while viewing region displays that varied the distance relationship of two pairs of comparison points (documents), the context provided by the display's region structure, and the visual characteristics of the region boundaries and polygons. We were specifically interested in how viewers balance or coordinate the implications of distance relationships in the displays with the implications of region membership relationships to infer similarity between documents. We also investigated how hue—a nonspatial visual variable—influences people's similarity assessments.

Experiment 1

Our first experiment investigated how nonexpert users interpret simple black-and-white region-display spatial-

Published by the IEEE Computer Society



(Courtesy of Antarctica Systems)

1 A region-display spatialization.

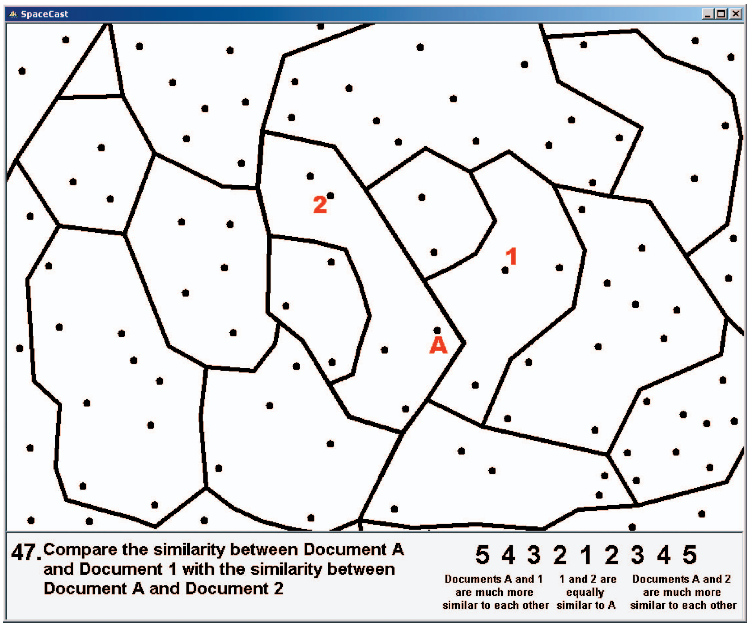
izations. We showed research participants computer displays of points overlaid with planar-enforced polygon coverage so each point was within exactly one 2D polygonal region. We explained that the points represented documents, with three of the document points labeled A, 1, and 2 (see Figure 2). We asked participants to compare the similarity of A and 1 to the similarity of A and 2 using a 9-point scale. That is, we asked participants to compare the relative similarity of two pairs of document points with various regional membership relations.

Independently of regional membership, the relative locations of the three comparison points varied so that the direct (straight-line) metric distances between the two pairs (A and 1 and A and 2) were equal, or differed to varying degrees. In addition to the region trials, this experiment also tested other spatialization metaphors, namely points without regions and points within networks. Other articles report results from trials involving point spatializations[1] and results involving network spatializations.[3] Here we focus exclusively on results from the region displays.

Method. Participating in the experiment were 44 students (25 male and 19 female) from an undergraduate regional geography class, with a mean age of 21.0 years. The test sample represented the desired novice user population: most participants weren't geography majors; rated their map reading ability as average; had used maps only occasionally; and had no training in cartography, geographic information systems (GISs), computer graphics, or graphic design.

The participants viewed computer displays created using ESRI ArcMap and composed of black points contained within regions formed by surface tessellation into 2D polygons (both convex and concave) with black boundaries. These were inspired by region displays similar to that in Figure 1 but weren't actual treemap outputs. Each point represented a single document in a digital database. In each display, we used red text to label three points (A, 1, and 2) for participants to compare for similarity (see Figure 2). The display prompted participants to "compare the similarity between document A and document 1 with the similarity between document A and document 2." Participants rated similarity on a 9-point scale ranging left to right from 5 to 1 and then back up to 5 (see Figure 2). In this article, we refer to the pair of documents A and 1 as A:1 and A and 2 as A:2.

Participants viewed 10 region trials in a block (they also viewed 30 additional trials involving other display metaphors). We varied the region displays to allow comparisons of distance relationship effects on judged



47. Compare the similarity between Document A and Document 1 with the similarity between Document A and Document 2

5 4 3 2 1 2 3 4 5

Documents A and 1 are much more similar to each other

1 and 2 are equally similar to A

Documents A and 2 are much more similar to each other

2 Sample screenshot from a trial in the first experiment showing display, similarity question, and rating scale as they appeared to participants.

Background in Spatialization

Few nonexpert spatialization users know how spatializations are created, and, because spatializations rarely include legends or other traditional map marginalia, they don't provide these users with information on how to interpret spatialized display aspects such as distance, region, or scale.

A fundamental assumption in information visualization is that spatialized displays work because users can understand them intuitively.[1,2] If this assumption is generally true, understanding the fundamentals of geographic space (the metaphors' source domain) as understood by display users will help you construct cognitively adequate information displays based on meaningful spatial metaphors (target domains). Location and distance on the Earth's surface are among the most fundamental geographic primitives, and both are reflected in the distance-similarity metaphor.

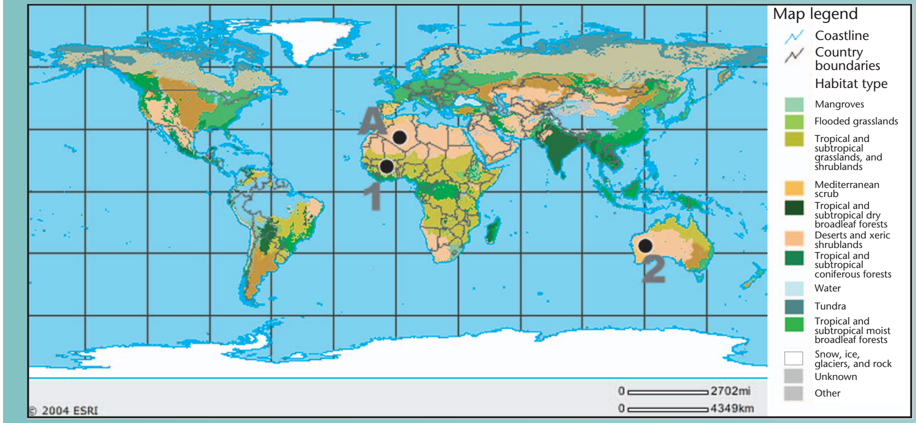
The distance-similarity metaphor is essentially the inverse of the empirical principle of the first law of geography.[3] The first law of geography suggests that you can predict the similarity of geographic features based on their relative locations on the Earth's surface. You can measure the law's effect using spatial autocorrelation indices—that is, indices showing the strength with which the characteristics of a particular location in space correspond to surrounding locations' characteristics.

Geographic regionalization—the discontinuous partitioning of space—is a fundamental task that has occupied geographers for centuries. Some geographic phenomena vary more or less smoothly over space while others exhibit extreme discontinuities, which appears to violate the first law.[4] In fact, regionalization typically depends on the decay of similarity over distance suggested by the first law, insofar as regions consist of spatially proximate and thematically similar locations.[5]

Figure A shows a common geographic region display in which a discrete graphic model renders the more continuous reality. Area-class or categorical-coverage maps can depict regions.

These maps categorize each point in geographic space, and the regions visually emerge. The map in Figure A classifies the land area into terrestrial ecoregions representing zones of relatively homogeneous land cover properties.

In this map, same-colored zones identify areas on the Earth's surface that share similar ecological characteristics, as the map legend indicates. If the first law were the only principle operating, we'd expect the land-cover type "deserts and xeric shrublands" at place A to be more similar to the type "tropical and subtropical grasslands" at place 1 than to the identical type "deserts and xeric shrublands" at the more distant place 2. Geographers resolve this apparent contradiction by considering two additional spatial primitives: scale and aggregation. The scale at which one is exploring geographic space will determine how regular or irregular certain geographic phenomena appear on Earth's surface. As a general rule, geographic data exhibit increased variability with increasing distance.[4] Even if distance remains constant, depending on the context, there are many ways to aggregate individual locations into geographic areas. Consequently, you can't make (error-free) inferences about individual locations' characteristics based only on regional relationships—an example of the cross-level fallacy.



(Map used with permission of the National Geographic Society)

A Terrestrial ecoregions.

similarity to region membership effects. Either one pair was closer than the other pair, or they were the same distance apart. At the same time, either one pair was in the same region and the third point was in a neighboring region, or all three points were in different regions. We varied graphical elements that we didn't expect to affect similarity judgments (such as the absolute location of a point on the screen) nonsystematically.

Three practice trials at the beginning of the test introduced participants to the concept of similarity, the trial style, and the response scale format. To avoid priming any particular equivalence between distance and similarity, the practice trials prompted nondistance similarity judgments (for example, by asking participants to compare the similarity of images of a dog, cat, and tiger). Participants also responded to 11 pretest questions about their personal backgrounds, including questions on age, gender, and any visual impairments (including color blindness), as well as their formal experiences in particular areas such as cartography and GIS. After the main test questions, participants responded to 28 posttest questions that asked, for example, 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.

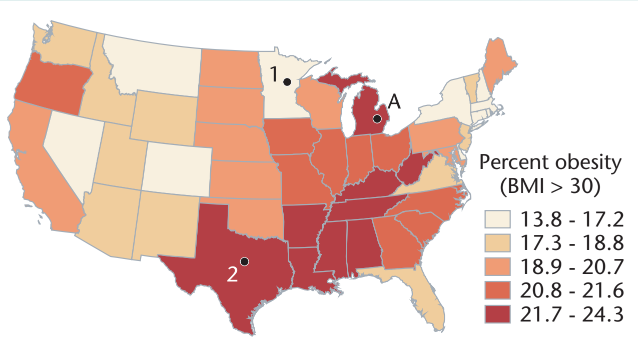
We administered the experiment using a Windows 2000 Pentium III personal computer. We programmed the interface using Microsoft Visual Basic 6.0, and projected images onto a back-projection screen using an

The scale issue, combined with the aggregation problem, gives rise to the modifiable areal unit problem (MAUP).[4] For example, land cover and surface characteristics measured on a soil patch of a few square meters in size (that is, high resolution or fine spatial scale) might not be evident at a coarser (that is, lower-resolution) scale, because the phenomenon might not be distributed evenly over the domain (in other words, it might exhibit high spatial heterogeneity).

A choropleth map is another type of discrete region display, arguably one of the most commonly used (and abused) methods of mapping. Contrary to area-class maps, the boundaries delineating regions in a choropleth map aren't data derived. The boundaries typically separate administrative regions such as census enumeration units or countries. Choropleth maps often depict intangible, abstract themes of the environment, such as health statistics, economic activities, or political events. In terms of abstraction level, they're more closely related to information visualization displays than to area-class maps (see Figure A).

The choropleth map in Figure B depicts the proportions of obese people per state in the US in 2000. Obesity rates are rendered with varying color shades—the darker the shade, the higher the ratio of obese people in the state. Spatial autocorrelation (that is, the first law) appears to be evident in this purely human health and dietary phenomenon, when mapped at the aggregation level of US states. Contiguous states in the South seem to have higher rates than adjacent states in the Midwest, for example. Despite this fact, the obesity rate for Michigan (A) is more similar to Texas (2) than to Minnesota (1), which is in fact closer.

Humans have great difficulty conceptualizing the *n*-dimensional reality of, for example, health phenomena. Aggregation and categorization are fundamental organizational principles of human cognition[6] that extend to geographical phenomena.[5] This might explain the popularity of area-class and choropleth maps, and why people seem to have no problem resolving spatial contradictions when using these maps. In fact, we could even conjecture that choropleth maps' popularity is due to their high cognitive adequacy—that is, they represent continuous properties of space discretely because this is how humans make sense of the environment. The question arises as to whether the interpretation of metaphorical region displays (such as in Figure 1 in the main article) might differ from the interpretation of area-class or choropleth maps of real-world regions or statistical phenomena. Could people interpret metaphorical displays more variably because they're less tied to assumptions about real space? Considering that people can somehow resolve apparent distance-similarity contradictions on choropleth maps, how do they resolve potential contradictions between the distance-similarity metaphor and other spatial metaphors such as aggregation (region membership), or nonspatial metaphors such as color hue?



B **Obesity rates per US state in 2000.**

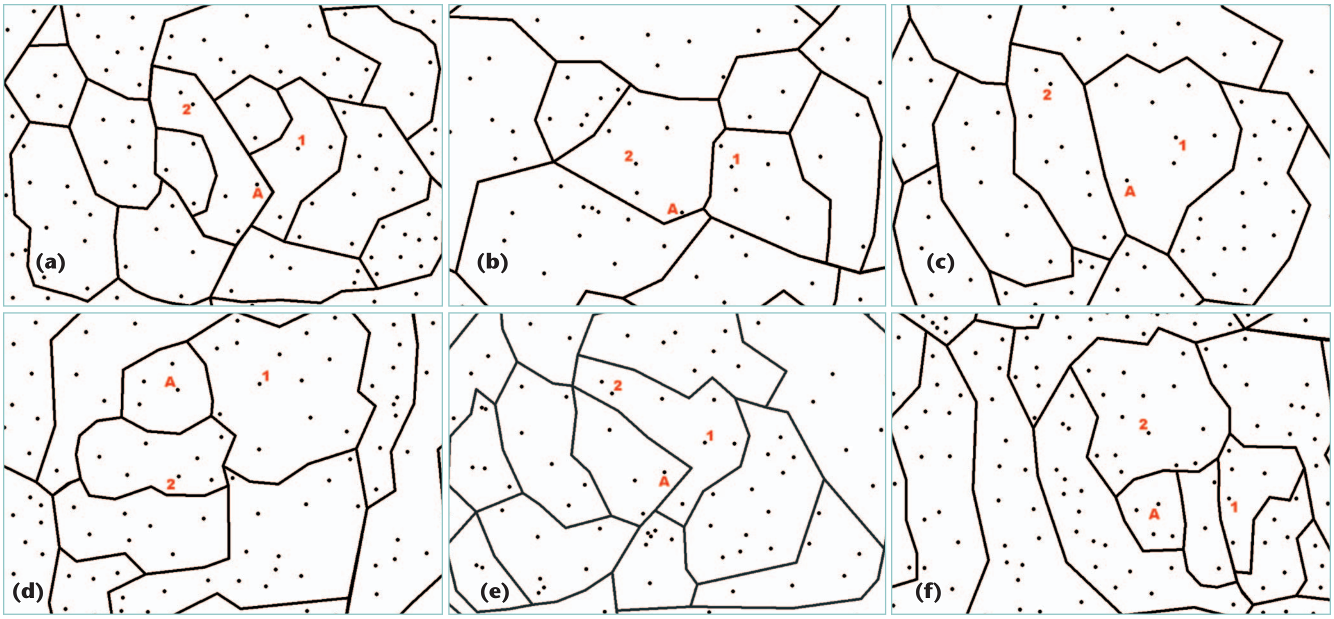
References

1. S. Card, K. Mackinlay, and B. Shneiderman, "Readings in Information Visualization," *Using Vision to Think*, Morgan Kaufmann, 1999.
2. T.A. Wise, "Visualizing the Non-Visual: Spatial Analysis and Interaction with Information from Text Documents," *Proc. IEEE Information Visualization*, IEEE CS Press, 1995, pp. 51-58.
3. W.R. Tobler, "A Computer Movie Simulating Urban Growth in the Detroit Region," *Economic Geography*, vol. 46, no. 2, 1970, pp. 234-240.
4. P.A. Longley et al., *Geographic Information Systems and Science*, Wiley, 2005.
5. D.R. Montello, "Regions in Geography: Process and Content," *Foundations of Geographic Information Science*, M. Duckham et al., eds., Taylor & Francis, 2003, pp. 173-189.
6. E.E. Smith and D.L. Medin, *Categories and Concepts*, Harvard Univ. Press, 1981.

RGB color projector, generating a 1.8-meter-wide and 1.4-meter-high image at 0.6 meters above the floor. Participants sat at a viewing table 2.7 meters away from the screen, resulting in an approximately 37-degree horizontal viewing angle. They used a standard mouse and keyboard to answer questions. The system recorded answers automatically and stored them digitally, including the time required to make similarity judgments. We measured response time as the elapsed time in milliseconds between the trial display appearing on the screen and the participant proceeding to the next trial.

We told participants that the display would present a series of trials about "diagrams that show an information collection from our computer database," which contains documents such as news stories, books, and journal articles, and that the display would present each document as a single point. We gave them no information on how to judge similarity, and attached no meaning to the graphical elements other than the points. We assured participants that there were no right or wrong answers, and asked them not to waste time, as we would time their answers.

After answering the pretest questions and performing the practice trials, participants responded to the main test trials organized into blocks (the block of region displays plus blocks of the other display types), rating all trials of one display type before turning to another type. Trials within each block appeared in a different randomized order for each participant. Finally, participants answered the posttest questions.



3 Example display types for our first experiment. The mean ratings were (a) 5.6 (not significant); (b) 6.1 ($p < 0.0001$); (c) 3.4 ($p < 0.0001$); (d) 5.2 (not significant); (e) 4.4 ($p < 0.05$); and (f) 5.2 (not significant).

Results. We treated similarity ratings as 9-point interval scales by scoring a response of 5 to the far left (A and 1 are much more similar) as a 1, a response of 5 to the far right (A and 2 are much more similar) as a 9, and a response of 1 in the middle (1 and 2 are equally similar to A) as a 5 (see Figure 2). Thus, a mean rating less than 5.0 indicates that participants saw A:1 as more similar, whereas a mean rating greater than 5.0 indicates that they saw A:2 as more similar. We then tested differences from equal similarity between A:1 and A:2 with t-scores based on the difference of the mean similarity rating from 5.0.

To examine the effects of direct distance and region membership on similarity judgments for the two pairs of comparison documents, A:1 and A:2, we examined several subsets of trials:

- One trial depicted the relative distance relationships as working against region membership relationships —that is, the document in the same region as A was further from A, while the other document was closer but in a neighboring region (see Figure 3a).
- Two trials depicted the relative distance relationships as equal but not the region membership relationships —that is, the documents were the same distance from A, but one was in the same region while the other was in a neighboring region (see Figure 3b).
- One trial depicted the relative distance relationships as reinforcing the region membership relationships —that is, one document was both closer to and in the same region as A, while the other document was in a different region (see Figure 3c).
- Three trials depicted both distance and region membership relations as equal (see Figure 3d).
- One trial examined whether distance would affect similarity when region membership was equal—that is, one document was closer to A but region membership relationships were equal (see Figure 3e).
- Finally, two trials examined whether region membership proximity would affect similarity—that is, both documents were equally close to A and in different regions from A, but one neighbored A's region and the other was one region removed (see Figure 3f).

We aggregated these trial subsets (when there was more than one trial), reverse-scoring trials when appropriate so that a mean score above 5.0 would reflect region membership's effect on similarity for all trials.

The participants gave the one trial in subset 1 a mean similarity rating of 5.6, $t(43) = 1.74$ (not significant). When region membership contradicted distance, it weakened but didn't eliminate distance's effect on similarity.

In contrast, for subset 2, participants gave the two trials a mean similarity rating of 6.1, $t(43) = 4.89$ ($p < 0.0001$). When distance relationships didn't differentiate the pairs of documents, region membership determined similarity.

Participants gave the one trial in subset 3 a mean similarity rating of 3.4, $t(43) = -5.60$ ($p < 0.0001$). In this subset, distance and region membership relationships apparently reinforced each other.

For subset 4, participants gave the three trials a mean similarity rating of 5.2, $t(43) = 1.45$ (not significant). When distance and region membership relationships were equal, similarity ratings were equal. Ostensibly no other visual variable was available on which to judge similarity.

They gave the one trial in subset 5 a mean similarity rating of 4.4, $t(43) = -2.50$ ($p < 0.05$). In this subset, distance across region boundaries affected similarity ratings when region membership relationships were equal.

Finally, participants gave the two trials in subset 6 a mean similarity rating of 5.2, $t(43) = 0.88$ (not significant). When distance relationships were equal and neither document shared a region with the comparison document, similarity ratings were again equal. Apparently being one region removed was no different from being more than one region removed; topological proximity between regions didn't operate according to a proximity/similarity metaphor.

We compared response times and similarity ratings as a function of trial block order and gender. Participants responded significantly more slowly during the first block of region trials (mean of 16.5 seconds per trial) than during the second (10.2 seconds per trial) or third blocks (11.4 seconds per trial)—$F(2, 41) = 4.75$ ($p < 0.05$). As we expected, however, mean similarity ratings didn't differ as a function of trial block order—$F(2, 41) = 2.07$ (not significant). Response times also didn't significantly differ as a function of participant gender: women responded in an average of 11.9 seconds per trial, and men an average of 11.7 seconds per trial—$t(42) = 0.11$. Participant gender didn't influence mean similarity ratings either: women rated the pairs of comparison documents a 5.1, on average, while men rated them a 4.9, $t(42) = 0.97$ (not significant).

Experiment 2

Our second experiment attempted to replicate and extend the results we obtained with region-display spatializations in our first experiment. First, we wanted to find out if our finding from the first experiment about region membership's effect on similarity judgments, over and above direct distance's effect, would replicate with a new set of black-and-white displays viewed by a new set of participants. We also wanted to examine how color hue affected judgments of similarity in region displays. We expected that region hue could compound or weaken region membership's effects. For instance, we expected that documents in separate regions might still be seen as similar if the two regions were of the same or even similar hues.

We varied hue in two ways. In one type of trial, we colored all regions in one of four hues, with the proviso that neighboring regions were always of a different hue. Thus, several regions were the same hue in these displays. We call these classed hues because we suspected that the hues might suggest to users membership in a thematic class. In the other type of hue trial, we colored each region a (somewhat) different hue, so the number of hues in the display equaled the number of regions. We call these unclassed hues.

In addition to region hue, we also used a black border around regions. On half of the hue trials, the colored regions were surrounded by a black border (like the borders in the black-and-white trials); on the other half, the colored regions simply abutted each other, with the color transitions signaling region boundaries. This way, we could determine whether clear borders (like those in the black-and-white trials) were necessary for the region effect to occur, or at least whether clear borders would strengthen or otherwise influence any region or hue effects.

Finally, we varied the size of the viewed displays. We reported elsewhere[1] that participants who viewed point displays projected at the relatively large size used in the first experiment rated similarity differences between the two pairs of documents more strongly than did participants who viewed displays projected at a smaller size. Here we examine whether display size influences rated similarity in region displays.

Method. Participants in this experiment were 48 students (27 male and 21 female) from an undergraduate introductory human geography class, with a mean age of 21.5. None had participated in the first experiment. They also received a small amount of course credit in return for their participation. We again judged the test sample to be a good sample of the desired novice user population.

As in the first experiment, participants viewed computer displays composed of different graphical elements. However, this experiment tested only region and point displays (other work discusses the point display trials[1]). All displays included black points, with three points (labeled A, 1, and 2) described as documents to be compared for similarity. Participants performed the same similarity judgments using the same scale as in the first experiment.

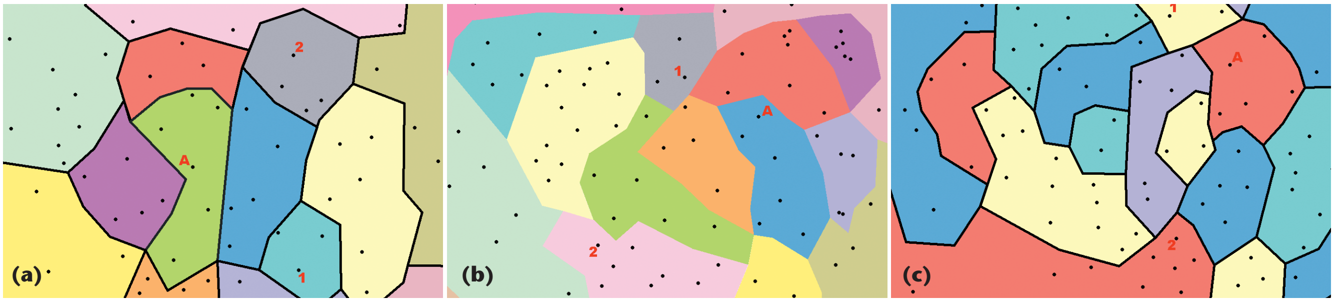
We varied the region displays according to three variables:

- hue presence (black-and-white versus colored regions),
- border (black borders versus no borders, applied only to colored regions), and
- hue class (classed versus unclassed hues, applied only to colored regions).

We presented participants with 78 region trials, which, with the 16 point trials reported elsewhere, gave us a total of 94 trials. We divided the 78 region trials into 10 trials of black-and-white regions (like those in the first experiment) and 68 trials of colored regions. We divided the 68 colored regions into 34 with borders and 34 without borders; and the two sets of 34 regions into 24 with classed hues and 10 with unclassed hues. We used more classed-hue than unclassed-hue trials because the classed trials offered more characteristics to vary. In particular, we could contrast trials with comparison points in different regions of the same hue to trials with points in regions of different hue in various ways. Figure 4a (next page) shows a bordered unclassed display; Figure 4b shows an unbordered unclassed display; and Figure 4c shows a bordered classed display.

We organized the region trials into three blocks according to the hue presence and border variables. We didn't use the hue class to structure trial blocks; rather, we randomly intermixed classed and unclassed hue trials within their respective blocks (bordered or unbordered).

Participants responded to five practice trials in this experiment, similar to those from the first experiment but including two dealing with color. After the main test trials, participants answered 56 posttest questions, including the 11 questions from the first experiment about their personal backgrounds. We adapted the additional posttest questions from the first experiment to account for the new display types.



4 Example display types for the second experiment: (a) bordered unclassed display, (b) unbordered unclassed display, and (c) bordered classed display.

We used the same equipment and setup as in the first experiment with one important addition. In this experiment, we varied the displays' projected image size. Half of the participants viewed a small image from a distance of 1.6 meters, projected to be 0.6 meters wide and 0.4 meters high at a height of 1.2 meters above the floor. This resulted in a horizontal viewing angle of approximately 20 degrees. The other half of the participants viewed a large image, projected (as in the first experiment) to be 1.8 meters wide and 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 degrees. We chose these sizes to provide a reasonable size contrast.

We tested participants as in the first experiment, presenting the three region blocks (black-and-white, bordered colored, and unbordered colored), along with the block of point trials, in counterbalanced orders. We presented trials within blocks in different random orders for each participant.

Results. We again treated similarity ratings as 9-point interval scales, so that a mean rating less than 5.0 indicates that participants saw A:1 as more similar, while a mean rating greater than 5.0 indicates that they saw A:2 as more similar. We first examined the black-and-white trials to determine whether we replicated our findings from the first experiment on the effects of direct distance and region membership on similarity judgments. We aggregated trials into subsets as in the first experiment's analysis.

For subset 1, one trial depicted the relative distance relationships as working against region membership relationships. Participants gave the trial a mean similarity rating of 6.5, $t(47) = 6.54$ ($p < 0.0001$). In this experiment, when region membership contradicted distance, region membership's effect weakened the distance effect so much that it essentially eliminated it.

For subset 2, one trial depicted the relative distance relationships of the two pairs as equal, but only one of the comparison documents was in the same region as A. Participants gave this trial a mean similarity rating of 6.6, $t(47) = 6.54$ ($p < 0.0001$). Region membership again determined similarity when distance relationships didn't differentiate the pairs of documents.

For subset 3, one trial depicted the relative distance relationships as reinforcing region membership relationships. Participants gave this trial a mean similarity rating of 6.9, $t(47) = 7.85$ ($p < 0.0001$). Again, distance and region membership relationships reinforced each other.

For subset 4, two trials depicted both distance and region membership relations as equal. Participants gave these trials a mean similarity rating of 4.9, $t(47) = -1.77$ (not significant). Again, when distance and region membership relationships were equal, similarity ratings were equal.

For subset 5, five trials examined whether distance would affect similarity when region membership was equal. We used several trials to examine this effect here because in the first experiment, we based our finding that distance mattered when region membership was equal on only one trial. Participants gave the five trials a mean similarity rating of 3.9, $t(47) = -8.71$ ($p < 0.0001$). Clearly, distance across region boundaries affected similarity ratings when region membership relationships were equal.

Finally, this experiment had no subset 6 trials contrasting region proximity while holding distance equal.

In sum, results for the monochrome region trials in the second experiment are largely consistent with the results of the first. That is, we again show that region membership in region-display spatializations largely, but not completely, overrides distance's effects on similarity.

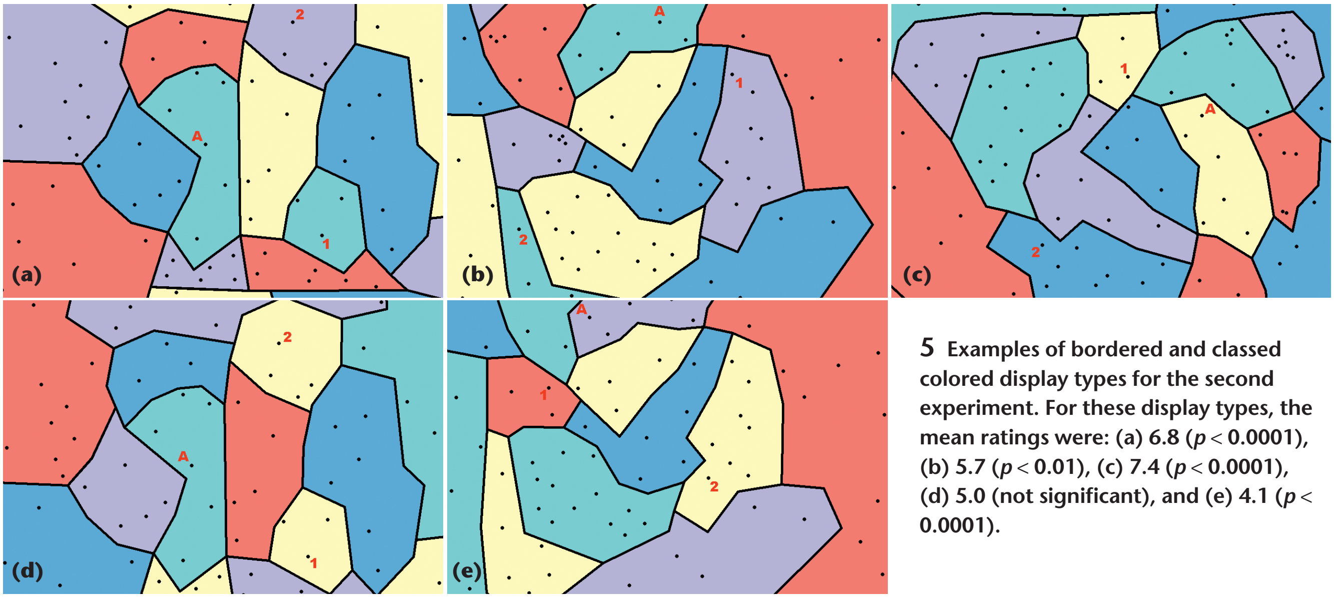
We next analyzed the colored region displays. We started with the unclassed trials because they presented the same 10 display patterns as the black-and-white trials, with the addition of unique colors to each region. We first analyzed bordered trials, as they were most similar to the black-and-white trials. We aggregated bordered unclassed trials into the same subsets as the black-and-white trials.

Participants gave one subset 1 trial a mean similarity rating of 6.7, $t(47) = 5.55$ ($p < 0.0001$). As with black-and-white displays (at least in the second experiment), when region membership contradicted distance, it eliminated distance's effect on similarity.

Participants gave one subset 2 trial a mean similarity rating of 7.3, $t(47) = 11.51$ ($p < 0.0001$). Again, region membership determined similarity when distance relationships didn't differentiate the pairs of documents.

The one trial in subset 3 depicted the relative distance relationships as reinforcing the region membership relationships. Participants gave that trial a mean similarity rating of 7.8, $t(47) = 14.86$ ($p < 0.0001$).

The two trials in subset 4 depicted distance and region membership relations as equal. Participants gave these trials a mean similarity rating of 5.0, $t(47) = -0.53$ (not



5 Examples of bordered and classed colored display types for the second experiment. For these display types, the mean ratings were: (a) 6.8 ($p < 0.0001$), (b) 5.7 ($p < 0.01$), (c) 7.4 ($p < 0.0001$), (d) 5.0 (not significant), and (e) 4.1 ($p < 0.0001$).

significant). Again, when distance and region membership relationships were equal, similarity ratings were equal.

Finally, for subset 5, five trials examined whether distance would affect similarity when region membership was equal. Participants gave the five trials a mean similarity rating of 4.5, $t(47) = -3.57$ ($p < 0.001$). Clearly, distance across region boundaries again affected similarity ratings when region membership relationships were equal.

Directly comparing these effects to those found in the black-and-white trials revealed that filling the regions with colors strengthened region membership's effects on similarity ratings. Although trial subset 1 didn't differ much between the black-and-white and colored versions, subsets 2 and 3, which revealed significant region effects with black-and-white regions, did so even more strongly with unclassed colored regions. Repeated-measures comparisons of black and white with bordered unclassed colored trials were significant for subsets 2 and 3: $t(46) = 3.05$ ($p < 0.01$) and $t(46) = 3.63$ ($p < 0.001$). Trial subset 5 showed a distance effect on similarity when region membership was equal; however, the effect was significantly weaker with bordered unclassed colored regions: $t(46) = 4.77$ ($p < 0.0001$). Essentially, colored regions overrode distance relationships more than did black-and-white regions, perhaps because color provided a stronger differentiation between regions.

The absence of borders around the unclassed colored regions didn't weaken or otherwise modify the region effects. Mean similarity ratings for unbordered trials were close to those for bordered trials and didn't significantly differ for any trial subset. (Because of a design error, one unbordered trial was displayed with different color hues than its corresponding bordered trial, but we excluded these trials from this analysis.) A repeated-measures analysis of all unclassed trials revealed that a border had no main effect on similarity ratings ($F(1, 47) = 1.92$ (not significant)), nor did an interaction exist between the border's presence and specific questions ($F(8, 40) = 0.85$ (not significant)).

Finally, we analyzed the results for the classed colored region displays. We again analyzed bordered trials first. The spatial patterns of five of the classed trials were copies of five of the unclassed trials. We could directly compare these trials because they displayed two of the three comparison documents as being within the same region and the third as being in another region of a different hue. In fact, these classed trials produced virtually the same effect pattern as did the corresponding unclassed trials. There was neither a main effect of classed-unclassed ($F(1, 47) = 2.50$ (not significant)) nor an interaction between classed-unclassed and specific questions ($F(4, 44) = 0.92$ (not significant)). Hence, the use of a classed color scheme per se apparently didn't alter similarity judgments, as compared to an unclassed color scheme. As we'll show, however, classed color schemes did make a difference when documents being compared were in different regions that could be of the same hue.

The remaining 19 classed trials with borders displayed the three comparison documents in three regions. (We didn't test pairs of documents within the same region in these trials, because documents in the same region would necessarily be in regions of the same hue.) In some of these trials, two of the three regions were the same hue; in others, all three were different hues. As we did with the black-and-white and unclassed trials, we aggregated these remaining classed trials into subsets of trials that displayed the two pairs of comparison documents—A:1 and A:2—according to similar relationships (in this case, relationships of relative distance and relative region hue). We used five subsets:

- One trial depicted the relative distance relationships as equal, but only one comparison document's region hue matched A's region hue (see Figure 5a).
- Four trials depicted the relative distance relationships as working against region hue relationships—that is, the document closer to A was in a region of a different hue, whereas the other document was in a region of the same hue as A (see Figure 5b).

- Four trials depicted the relative distance relationships as reinforcing the region hue relationships—that is, one document was both closer to A and in a region of the same hue (see Figure 5c).
- Two trials depicted distance and region hue relations as equal (see Figure 5d).
- Eight trials examined whether distance would affect similarity when region hue was equal—that is, one document was closer to A but neither document was in a region of the same hue as A (see Figure 5e).

We aggregated these trial subsets (when there was more than one trial), reverse scoring trials when appropriate so that a mean score above 5.0 would reflect the effect of region hue on similarity for all trials.

Participants gave one subset 1 trial a mean similarity rating of 6.8, $t(47) = 8.51$ ($p < 0.0001$). When we held distance across regions equal, participants interpreted matched region hues as indicating greater similarity.

The relationship between distance and similarity in point displays was significantly stronger with the large display than with the small display.

For subset 2, four trials depicted the relative distance relationships as contradicting region hue relationships. Participants gave those trials a mean similarity rating of 5.7, $t(47) = 2.80$ ($p < 0.01$). Thus, they were more likely to interpret matched region hues as indicating similarity than they were closer distance (across regions), although an examination of individual responses showed that several participants saw distance as equally or more relevant to similarity in this condition.

Participants gave the four trials in subset 3 a mean similarity rating of 7.4, $t(47) = 15.43$ ($p < 0.0001$). Distance and region hue relationships clearly reinforced each other.

Participants gave the two trials in subset 4 a mean similarity rating of 5.0, $t(47) = -0.52$ (not significant). When distance and region hue relationships were equal, similarity ratings were equal.

Finally, for subset 5, participants gave the eight trials a mean similarity rating of 4.1, $t(47) = -7.70$ ($p < 0.0001$). Distance across region boundaries affected similarity ratings when region hue relationships were equal.

As with the unclassed colored regions, the absence of borders around the classed colored regions had no observable effect on similarity ratings. Mean similarity ratings for unbordered trials were similar to those for bordered trials, and didn't significantly differ for any of the trial subsets. (Because of a design error similar to that made on the unclassed trials, one unbordered classed trial was displayed with different color hues than its corresponding bordered trial. We excluded both trials from this analysis.) A repeated-measures analysis of all classed trials revealed that the border's presence had no main effect on similarity ratings—$F(1, 47) = 0.22$ (not significant)—nor an interaction between the border's presence and specific questions—$F(22, 26) = 1.17$ (not significant).

Response time and similarity as a function of block order and gender. As in the first experiment, we compared response times as a function of trial block order and gender. Participants responded to black-and-white region trials significantly more slowly when they occurred during the first trial block (mean of 15.8 seconds per trial) than during the second (9.2 seconds per trial), third (8.1 seconds per trial), or fourth blocks (7.0 seconds per trial)—$F(3, 44) = 8.38$ ($p < 0.001$). As we expected, mean similarity ratings didn't differ as a function of trial block order, $F(3, 44) = 1.05$ (not significant). Response times also didn't significantly differ as a function of participant gender: women responded in an average of 10.6 seconds per trial, men in an average of 9.4 seconds per trial, $t(46) = 0.75$ (not significant). Participant gender didn't influence mean similarity ratings either: on average, women rated the pairs of comparison documents a 4.5, whereas men rated them a 4.6, $t(46) = -0.74$ (not significant).

On colored regions, whether bordered or unbordered, participants responded to trials significantly more slowly when they occurred during the first trial block (mean of 10.5 seconds per trial, both for bordered and unbordered regions) than during the second (8.3 and 9.9 seconds per trial), third (8.6 and 7.7 seconds per trial), or fourth blocks (7.2 and 7.2 seconds per trial). These differences didn't reach statistical significance, however—$F(3, 44) = 1.78$ and 2.14. As we expected, mean similarity ratings didn't differ as a function of trial block order for either bordered or unbordered regions—$F(3, 44) = 0.51$ and 1.29 (not significant).

Response times also didn't significantly differ as a function of participant gender: women responded in an average of 8.6 seconds per trial on bordered regions and 9.5 seconds per trial on unbordered regions; men responded in an average of 8.8 seconds per trial on bordered regions and 8.3 seconds per trial on unbordered regions. Neither of these reached statistical significance, $t(46) = -0.22$ and 1.00.

Participant gender didn't influence mean similarity ratings either. On average, both men and women rated the pairs of comparison documents on bordered regions a 4.5, $t(46) = 0.03$ (not significant). Women rated the pairs of comparison documents on unbordered regions a 4.4, on average, whereas men rated them a 4.6, $t(46) = -1.00$ (not significant).

Display scale. As we report elsewhere,[1] the relationship between distance and similarity (as expressed by a correlation) in point displays was significantly stronger with the large display than with the small display. To test the possible effect of display size on judged similarity in region displays in the second experiment, we compared the size of mean differences reported in the two display size conditions. We tested each group of trials (black-and-white, unclassed colored, and classed colored) in repeated-measures analyses with display size as a between-participant factor, and question and border (when appropriate) as repeated factors. The main effect of display size didn't reach significance for any trial group, nor did any size interactions with question or border.

In point-display spatializations, similarity is essentially a matter of metric distance, with some exceptions such as clusters. Changes in display scale directly affect apparent metric size and distance relationships. However, as our empirical results suggest, image scale seems less important in region-display spatializations because region membership, region hue, and so on largely account for similarity, and not metric distance.

Region enclosure

Region enclosure relations influence participants' judgments of the relative similarity of document points in an information display. When interpoint distances were equal but one pair was in the same region and the other pair spanned a region boundary, most participants used region enclosure as a basis for making relative similarity judgments. This was the case for monochrome displays and also for multihued displays, whether or not border lines explicitly marked the region boundaries. When region enclosure relations contradicted interpoint distances, the region relations weakened, cancelled, or even reversed similarity judgments that participants likely would have made based on interpoint distances alone. Also, in colored displays, when all three points were in different regions, points in regions of the same color were judged to be more similar, other things being equal. These results aren't that surprising given knowledge of established principles of cartographic design. However, the results we report here are the first to confirm these design principles empirically in the context of point similarity judgments on diagrams divided into regions.

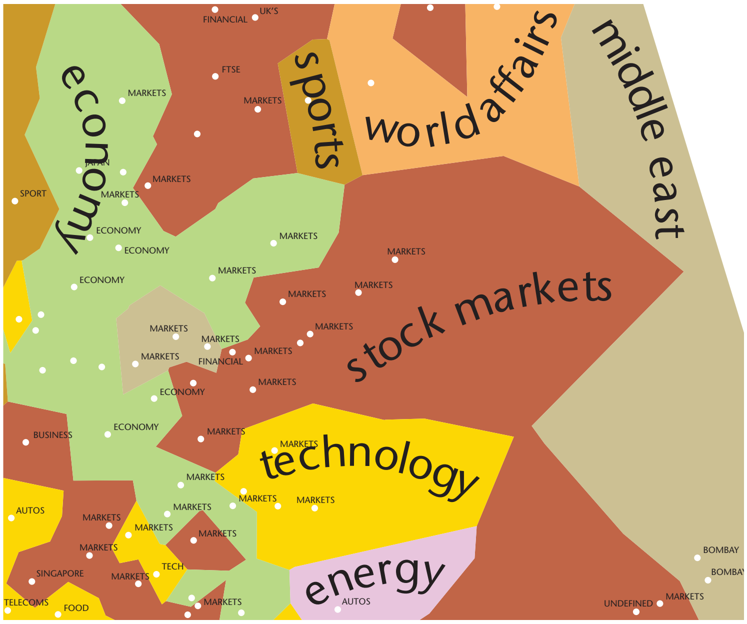
In cartographic displays of real geographic information (that is, information about Earth's surface), geographic reality at least partially determines the locations of points and region boundaries. Cartographers have some flexibility in their choice of colors, line weights, and other design elements, but not much flexibility in feature or boundary positions. The results reported here thus both confirm cartographic design practice and suggest possible unintended consequences of some design choices. For example, users might interpret similar or identical colors of regions on political maps as indicating semantic similarity, even if that isn't the designer's intention. The design implications of this study are more substantial in the context of information displays where nonspatial information is spatialized for presentation, and where networks, regions, and even positions are adjustable design elements rather than representations of features with positions in the real world.

Design implications for region display spatializations

Together with empirical findings from point, network, and surface display spatializations, our experimental results on region displays fully support a theoretical spatialization framework grounded on GIScience, including cognitive, perceptual, and experiential principles.[5] Our results confirm that the spatial metaphor region is a useful concept for depicting clusters of similar documents in an information space. The results also suggest that the interpretation of metaphorical region displays operates akin to regions depicted in area-class maps or choropleth maps of real-world phenomena. If interpoint distance is meant to fully capture and present document similarity, display designers should avoid adding regions to the display, because, as this study's results clearly demonstrate, region enclosure relations and region color similarity will modify judgments that participants would have made based on interpoint distances alone. If, however, similarity indicated by the display is meant to be of a more qualitative nature, adding regions around groups of similar points and coloring semantically similar regions with similar hues will reinforce the similarity judgments that viewers are likely to make from point displays alone. Regions help structure space and can help users navigate through a hierarchically organized information space. However, if similarity is meant to be communicated entirely via regional membership, our results suggest that interpoint distances continue to influence perceived similarity even in region displays.

Our results confirm that the spatial metaphor region is a useful concept for depicting clusters of similar documents in an information space.

Our results for the nonspatial color hue metaphor confirm the pattern found for colored network-display spatializations[3] in suggesting the utility of long-established cartographic design principles for spatialization. Cartographers use color hues to depict qualitative information about geographic features. Hue differences are mapped onto categorical differences (that is, at the nominal level of measurement) such as soil types, land-use zones, or animal habitats. Haphazard use of color could lead to misinterpretation—for example, if the same color represents different categories (as Figure 1 shows). Designers of the region spatialization in Figure 1 might have attempted a color scheme typically found on political maps, where a small number of hues (typically four or five) are assigned to countries such that no two adjacent countries have the same color. In this example, hue is simply used for aesthetics and perceptual distinctiveness, and doesn't signify a semantic category membership. According to our findings, we suggest redesigning the display using a set of distinctly different color hues for the hierarchy's top-level categories and varying color saturation or value within the chosen hue to show different depth levels within the hierarchy branches. A color selection tool such as the ColorBrewer (see http://www.colorbrewer.org) provides cartographically sound color schemes.



6 Region spatialization of news stories with region labels.

Designers can also use region labels as graphic symbols to reinforce category membership, as the extract of a news-article spatialization in Figure 6 shows. Cartographically informed labeling can help reduce the need for additional map legends.

Conclusion and outlook

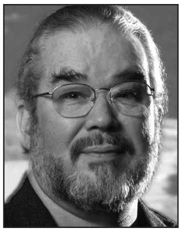
Taken with our earlier results, our findings for points within regions provide a comprehensive picture of how other information display design elements influence judgments of the similarities of pairs of points in such displays. Interestingly, in point displays, clustering or alignment of groups of points has some effect on similarity judgments, introducing into the visual field what amount to implicit lines or regions. Thus, avoiding networks or regions won't necessarily avoid network and region effects, because point patterns can induce line-like and region-like effects. Color hue can enhance the region effect or hinder it. These examples demonstrate how the use of cartographic design principles provides a sound design framework for better controlling perceptual and cognitive effects of design choices for spatialization displays. ■

References

1. D.R. Montello et al., "Spatialization: Testing the First Law of Cognitive Geography on Point-Spatialization Displays," *Spatial Information Theory: Foundations of Geographic Information Science, Conf. Spatial Information Theory* (COSIT), W. Kuhn, M.F. Worboys, and S. Timpf, eds., LNCS 2205, Springer-Verlag, 2003, pp. 335-351.
2. B. Johnson and B. Shneiderman, "Treemaps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures," *Proc. 2nd Int'l IEEE Visualization Conf.*, IEEE CS Press, 1991, pp. 284-291.
3. S.I. Fabrikant et al., "The Distance-Similarity Metaphor in Network-Display Spatializations," *Cartography and Geographic Information Science*, vol. 31, no. 4, 2004, pp. 237-252.
4. S.I. Fabrikant, "Distanz als Raummetapher für die Informationsvisualisierung [Distance as a Spatial Metaphor for the Visualization of Information]," *Kartographische Nachrichten*, vol. 52, no. 6, 2003, pp. 276-282.
5. S.I. Fabrikant and A. Skupin, "Cognitively Plausible Information Visualization," *Exploring Geovisualization*, J. Dykes, A.M. MacEachren, and M.J. Kraak, eds., Elsevier, 2005, pp. 667-690.

Sara Irina Fabrikant *is an associate professor and head of the Geographic Visualization and Analysis Division in the Department of Geography at the University of Zurich. Her research interests include geographic information visualization, GIS science and cognition, and graphical user-interface design, including dynamic cartography and hypermedia. Fabrikant has a PhD in geography from the University of Colorado-Boulder. Contact her at sara@geo.unizh.ch*

Daniel R. Montello *is a professor in the Department of Geography and an affiliated professor in the Department of Psychology at the University of California at Santa Barbara. His research interests include spatial and geographic perception, cognition, and behavior, including cognitive issues in cartography and GIS. Montello has a PhD in environmental psychology from Arizona State University.*



David M. Mark *is a professor in the Department of Geography at the University at Buffalo, SUNY, where he also is director of the Buffalo office of the US National Center for Geographic Information and Analysis. His research interests include geospatial ontology, spatial cognition, geographic information science, and indigenous GIS. Mark has a PhD in geography from Simon Fraser University.*