International Journal of Geographical Information Science
Vol. 22, No. 4, April 2008, 463–478

Research Article

The effect of instructions on distance and similarity judgements in information spatializations

SARA IRINA FABRIKANT*† and DANIEL R. MONTELLO‡
†Department of Geography, University of Zurich, Winterthurerstrassse 190, CH-8057 Zurich, Switzerland
‡Department of Geography, University of California Santa Barbara, CA 93106, USA

(*Received 2 June 2006; in final form 13 June 2007*)

We investigate the relationship of perceived distances to judged similarities between document points in various types of spatialized displays. Our findings suggest that the distance–similarity relationship is not as self-evident to viewers as is commonly assumed in the information visualization literature. We further investigate how participants interpret instructions to judge distances when those instructions do or do not specify the type of distance. We find that in all types of spatialization displays, there is no significant difference between default and direct judgements of distance; people clearly interpret default distance instructions to refer to direct (straight-line) distance. These findings provide direct evidence on the conditions under which people employ distance when assessing similarity between data objects in various types of spatialized views and, when they do, which type of distance. They also give insight into how people explore the similarity of geographic features depicted in cartographic maps or GIS displays.

Keywords: Spatialization; Distance; Similarity; Human subject testing

1. Introduction

Information *spatializations* are visualizations that depict information-bearing documents such as books, websites, or news stories as spatial displays. The documents are commonly depicted as points, and the spatial arrangement of the points represents semantic relations among the documents. When inspecting a point-display spatialization of news stories, for instance, one would expect to find highly related news stories clustered together. We have coined this spatialization principle the 'distance–similarity metaphor' (Montello *et al.* 2003). The distance–similarity metaphor is reminiscent of Tobler's 'first law of geography' (Tobler 1970), which contends that one can predict the similarity of geographic features based on their distances to other features on the Earth's surface. (See a more recent discussion on Tobler's 1st law in a special forum in the *Annals of American Geographers*, Sui 2004) According to the first law, distance is correlated with similarity, in most cases because distance determines similarity. In essence, the distance–similarity metaphor is the inverse of the first law of geography, because similarity typically determines distance in spatializations. Thus, we have referred to the 'first law of cognitive geography' (Montello *et al.* 2003)—people believe that closer features are more

*Corresponding author. Email: sara@geo.unizh.ch

International Journal of Geographical Information Science
ISSN 1365-8816 print/ISSN 1362-3087 online © 2008 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/13658810701517096

similar than distant features. To the extent that this principle is true, it provides theoretical justification for the distance–similarity metaphor as a principle of spatialization design. Both the physical and cognitive versions of the first law of geography depend on distance. But the information visualization literature gives no insight as to which type of distance people do employ or should employ when assessing similarity between spatialized information objects.

2. Background

Our previous studies have produced empirical evidence of how distance operates in the context of two-dimensional point-display spatializations (Montello *et al.* 2003). We have shown that non-expert viewers interpret distance relationships among the points in these displays to reflect similarity relationships. Specifically, closer points are generally seen to represent more similar documents. This is the 'first law of cognitive geography'. Designers of visualizations implicitly base their application of the distance–similarity metaphor, in which more similar points are placed closer together in the visualization, on the first law of cognitive geography. A good example of this is the 'Galaxy' information visualization tool (Wise *et al.* 1995), shown in figure 1 (one of three InfoVis 2004 contest winners, available on the Web at: http://www.cs.umd.edu/hcil/InfovisRepository/contest-2004/3/unzip/PNNLstandardform2004.html). The designers of this spatialization labelled it Galaxy, as it employs 'visual metaphors that recapitulate the experiences of the night sky' (Wise *et al.* 1995, p. 58).

In a series of additional studies (Fabrikant *et al.* 2004, 2006), we have shown that various display types other than pure point displays lead to modified expressions of the first law of cognitive geography, suggesting that designers of information

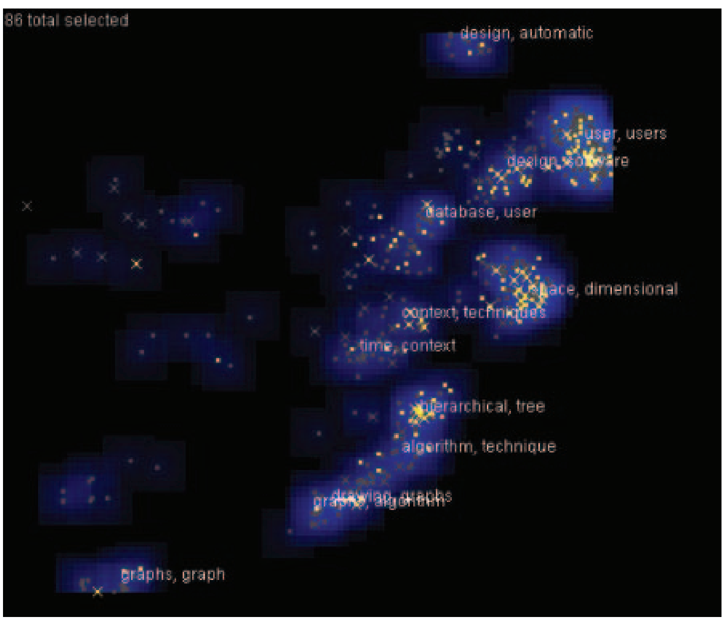


Figure 1. Point display spatialization of conference papers on information visualization (© Pacific Northwest National Laboratory, PNNL).

displays should apply the distance–similarity metaphor in a modified fashion. These other display types also show documents as points but embed them in a context of additional linear and areal display features. We have tested a variety of network displays (with monochromatic links and links of varying colour hue, width, and value) and region displays (with monochromatic regions and regions of varying unclassed or classed colour hues). In different display types, we have been able to draw conclusions about the effects of various display properties on similarity judgements, including the properties of direct distance (metric straight-line distance); network distance (metric distance along links); link hue, width, and value; and region membership and hue.

In all of our previous studies, however, research participants based their interpretations and judgements of similarity relationships on what we call 'default similarity' instructions. That is, they were simply told the points represented documents and asked to rate the relative similarity of two specific pairs of comparison document points. We did not tell them on what they should base their judgements of similarity. Thus, we have derived conclusions about the relationship of judged similarity to distances of various types on the basis of indirect evidence. Furthermore, we have inferred that various optical effects on perceived distance can influence similarity or not. A good example is the 'vertical illusion', wherein extents aligned vertically in the visual field are seen to be a little longer than those aligned horizontally (see discussion and demonstration of this in Montello *et al.* 2003). However, we have not demonstrated these optical effects on distance directly. We have varied *actual* distance properties of the displays but theorized about the effects of *perceived* distances. That is, given that we have only asked for similarity judgements, we have only been able to infer indirectly how various display properties affect perceived distances in the displays. Especially in network displays, we have hypothesized about the effects of different types of perceived distance on similarity judgements but never had any direct measurements of different perceived distances on the same displays for which we have measurements of perceived similarity. Our causal inferences have therefore not been as direct or solid as we would like. In the present research, we report an experiment with non-expert participants on the effect that instructions concerning similarity and distance have on the interpretation of distance and similarity relationships in different types of spatialization displays.

3. Experiment

In this experiment, we manipulated instructions to participants in one of three conditions. In the first condition, called the *similarity* condition, participants were asked to make default similarity judgements like those made in our previous studies; judgements were made between pairs of comparison document points depicted according to a variety of display metaphors, including points, monochromatic networks, and monochromatic and coloured regions. While viewing the same displays, a second group of participants judged relative distances instead of similarities between pairs of points. These participants were not told what type of distance to judge, so we call this the *default distance* condition. A third and final group also judged relative distances between the pairs of points, but they were explicitly told to judge the direct or straight-line distance between the points. We call this the *direct distance* condition.

3.1 *Methods*

3.1.1 Participants. Seventy-two students (32 males and 40 females) from undergraduate regional geography and introductory human geography classes took part in the experiment, with a mean age of 21.9 years. Based on the background data collected in the post-test questionnaire, the tested participants provided a good sample of the desired non-expert user population. The vast majority rated their map reading ability as average, claimed to use maps only occasionally, and had never had training in cartography, GIS, computer graphics, or graphic design. They received a small amount of course credit in return for their participation.

3.1.2 Materials. Participants viewed computer displays created using ESRI ArcMap. The displays consisted of black points. Each point was intended to represent a single document in a digital database. In each display, three points to be compared for similarity were labelled with red text as 'A', '1', and '2' (see figures 2 and 3). Four different display types were examined on different trials, all replications of the display types we have examined in the previous research cited above. Of the total of 71 trials, 18 were point displays (figure 3(*a*)), 21 were network displays (figure 3(*b*)), nine were monochromatic region displays (figure 3(*c*)), and 23 were classed coloured region displays (figure 3(*d*)).

Participants were randomly assigned to judge the 71 display types in one of three instruction conditions. In the *similarity* condition, a replication of the task used in our previous studies, they were told to 'compare these points in terms of their similarity'. No other information was provided on how to judge similarity. Participants rated similarity on a 9-point scale ranging left to right from '5' to '1' and then back up to '5'. On the left, '5' was labelled 'Documents A and 1 are much more similar to each other'. In the middle, '1' was labelled '1 and 2 are equally similar to A'. On the right, '5' was labelled '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 A and 2 as 'A:2'.

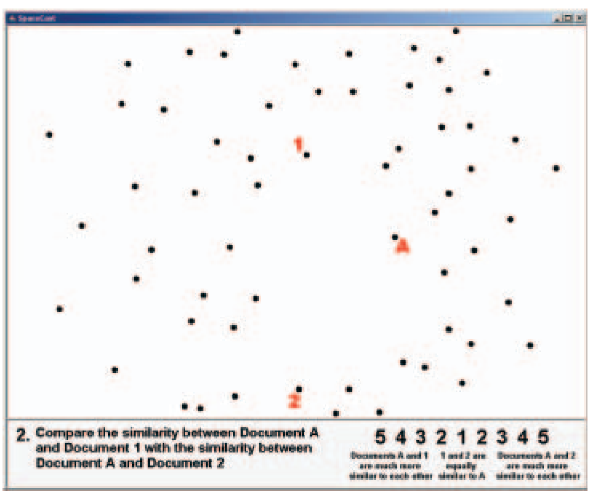


Figure 2. Sample screenshot showing display, similarity question, and rating scale as they appeared to participants.

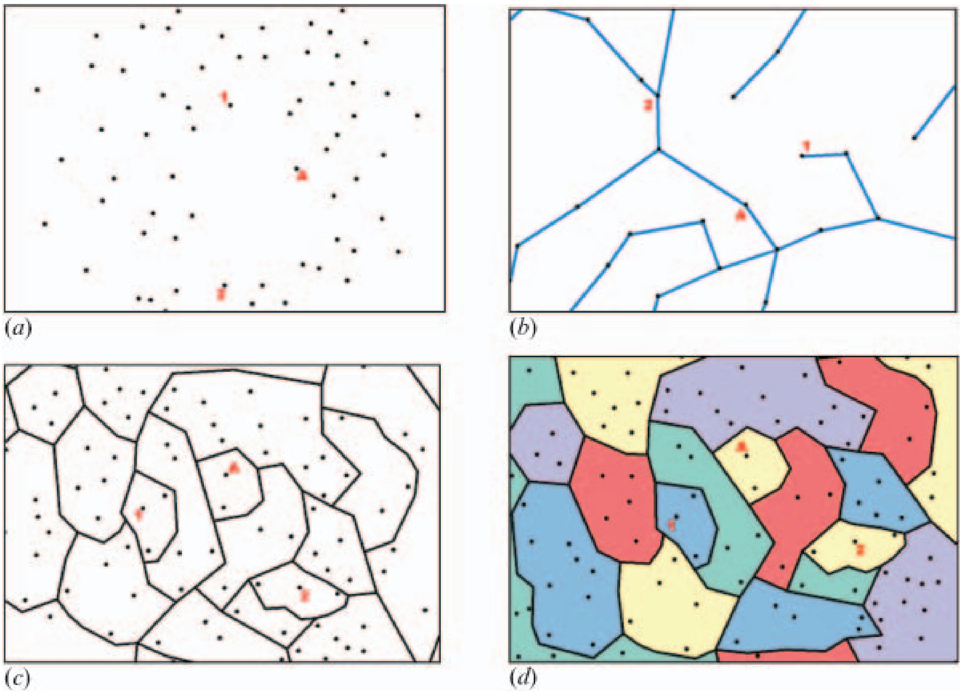


Figure 3. Sample displays types: (*a*) points, (*b*) networks, (*c*) monochromatic regions, (*d*) classed coloured regions.

In the *default distance* condition, participants were told to 'compare these points in terms of their distance from each other'. They were told nothing else about what was meant by distance. Distance was rated on a 9-point scale like that used to rate similarity, except '5' on the left was labelled 'Documents A and 1 are much closer to each other', '1' in the middle was labelled '1 and 2 are equally close to A', and '5' on the right was labelled 'Documents A and 2 are much closer to each other'.

Finally, in the *direct distance* condition, participants were told to 'compare these points in terms of their direct distance from each other. "Direct" distance is just the distance between the document points on the screen along the shortest path that would connect them. It is the "as-the-crow-flies" distance'. Direct distance was rated on the same 9-point scale used to rate default distance.

Participants were introduced to the concept of similarity, the style of the trials, and the format of the response scale through five practice trials at the beginning of the test. After the main test trials, participants answered 49 post-test questions, including 11 questions about their personal backgrounds, such as questions on age, gender, the presence of visual impairments (including specifically colour blindness), as well as their formal experiences in the areas of cartography, GIS, computer graphics, and graphic design. The other questions concerned, for example, how useful they thought each display type was for judgement 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. The monitor measured 16 $\times$ 12 inches; viewed from a distance of about 20 inches, the image subtended a horizontal viewing angle of approximately 40°. A standard mouse and keyboard were used to answer questions.

Answers were recorded automatically and stored digitally, including the time required to make similarity judgements. The response time was measured as the elapsed time in milliseconds between the trial display appearing on the screen and the participant proceeding to the next trial.

3.1.3 Procedure. Participants were first told 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. Depending on the condition to which they were assigned, participants were told to judge similarity, default distance, or direct distance. They were assured that there were no right or wrong answers, and they were asked not to waste time, as their answers would be timed.

Participants then performed the practice trials. Following that, they responded to the main test trials organized into four separate blocks, so that participants rated all trials of one display type before turning to another type. 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 were thanked for their participation.

3.2 *Results*

We transformed the similarity and distance judgements from 5-1-5 to 1–9 in order to analyse them as nine-point interval scales. Thus, a mean judgement less than 5.0 indicates that participants saw A:1 as more similar, while a mean judgement greater than 5.0 indicates they saw A:2 as more similar. Therefore, differences from equal similarity between A:1 and A:2 could be tested with t scores based on the difference of the mean similarity judgement from 5.0 (i.e. one-sample t test). Differences in judgements among the three conditions were tested with Analysis of Variance (ANOVA) using condition as a between-participant factor.

3.2.1 Point displays. The pattern of ratings on the 18 point displays was very clear. The main effect of instruction condition was statistically significant for all but four display trials. On every trial, pairwise comparisons showed that judgements in the *default distance* and *direct distance* conditions did not significantly differ. The two distance judgements never differed by more than half a judgement scale point and usually much less. For each of the 14 display trials that significantly differed across conditions, therefore, it was judgements in the *similarity* condition that were significantly different from judgements in the two *distance* conditions. These differences followed one of four patterns, corresponding to patterns in the displays:

1. On eight trials, one pair of comparison points is rated to be closer in distance and more similar than the other pair, but the similarity difference between the pairs is not as extreme as the distance difference (figure 4(*a*)). In the case of two of the trials, the distances were actually the same but were rated to be a little different because document point '2' was a little higher in the visual field than document point '1'; this produced a slight illusion of distance that affected distance estimates but was not strong enough to affect similarity ratings.

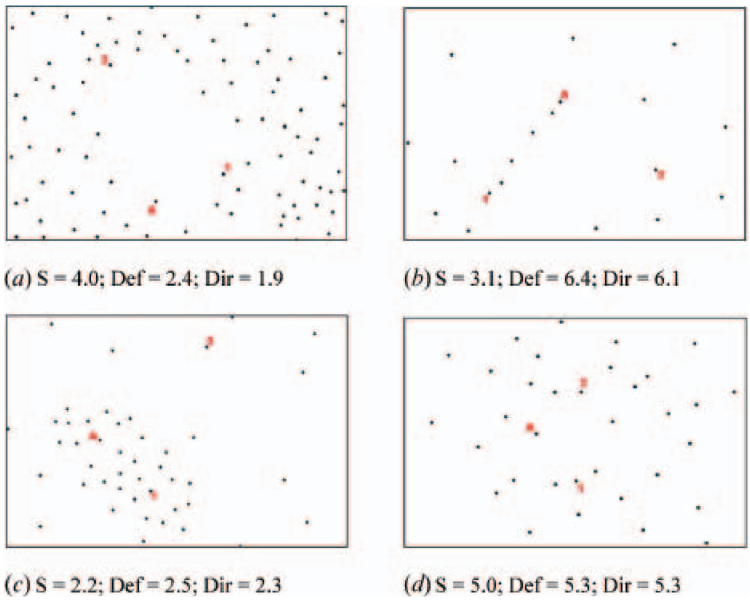


Figure 4. Example point displays and mean ratings. S=Similarity; Def=Default Distance; Dir=Direct Distance (5.0 is A:1 and A:2 are equal; <5.0 is A:1 closer or more similar; >5.0 is A:2 closer or more similar).

2. On two trials, the pairs are rated to be equal in distance, but because of an emergent linear or cluster effect, one of the pairs is rated to be more similar (figure 4(*b*)). Similarly, on four other trials, one pair of points is rated to be closer in distance, but because of an emergent linear or cluster effect, the more distant pair is rated to be more similar. The emergence and influence of such apparent linear or cluster features, given suitably arranged points, replicates findings we reported in Montello *et al.* (2003). Also replicating Montello *et al.*, linear features do not emerge unless displays have at least three or more intervening points. In fact, some of these trials show conclusively that emergent features strongly effect similarity through a mechanism other than distance reduction, because given enough intervening or cluster points, the apparent distance between points was seen to be relatively greater (an example of the 'filled-space illusion') even as their apparent similarity was also seen to be greater.
3. On one trial, one pair of points is rated to be closer in distance and more similar, and because of an emergent cluster that reinforced the distance differences, the similarity difference is seen to be just as extreme as the distance difference (figure 4(*c*)).
4. On three trials, the pairs are rated to be equal in distance and equal in similarity (figure 4(*d*)). These trials had insufficient emergent features to influence similarity judgements over and above the effects of distance.

We conducted an additional analysis to investigate further our finding that similarity followed direct distance on the 11 trials where no emergent feature was involved, but similarity differences were less extreme than distance differences, that is, the trials discussed above in figure 4(*a*) and (*d*). Treating each of the 11 trials as a unit of analysis, we calculated within-participant Pearson correlations of the mean similarity judgements (mean similarity difference between A:1 and A:2) with the mean direct distance judgements (mean direct distance difference between A:1 and A:2). This correlation averaged .89 over participants, a very large correlation (accounting for about 80% of the variance) that shows that changes in similarity judgements were sensitive to metric changes in distance, even if the similarity judgements were not as extreme as the distance judgements.

These results for point displays can be summed up concisely. Participants interpreted default (non-specific) instructions as requests to estimate direct or straight-line distances in point displays. Distance ratings followed actual direct distance relationships quite closely, except for optical illusions such as the filled-space illusion that can alter perceived distances a little. On displays without emergent features, similarity followed direct distance but not as severely. Given enough intervening or cluster points, an emergent feature largely dictated similarity judgements (points that were part of the same feature were seen to be more similar); the effect of emergent features on similarity held even when the feature caused an illusion in which points that were part of the feature were seen to be further away from each other.

3.2.2 Network displays. The pattern of results with the 21 network displays was also very clear. On these trials, the main effect of instruction condition was statistically significant for *only* four displays. On each of these four, pairwise comparisons again showed that judgements in the *similarity* condition differed from those in the two distance conditions, which never differed from each other. Again, the two distance judgements never differed by more than half a scale point and usually much less. The patterns of response across displays followed one of five patterns:

1. On six trials, both direct and network distances for the two pairs of comparison points were equal (figure 5(*a*)). On these trials, ratings of distance and similarity did not differ significantly from each other or from neutral (5.0).
2. On 12 trials, direct distances were equal, but network distances differed (figure 5(*b*)). Ratings of distances (whether default or direct) did not differ significantly from each other or from a neutral rating. Ratings of similarity did differ from neutral. Points connected by shorter network distances were rated as more similar (replicating Fabrikant *et al.* 2004), although this difference reached significance on only six of the 12 trials.
3. On one trial, direct distances differed for the two pairs, but network distances were equal (figure 5(*c*)). On this trial, ratings of distance significantly differed from neutral, but ratings of similarity did not.
4. On one trial, both direct and network distances differed for the two pairs, but in opposite directions (figure 5(*d*)). Ratings of distance and similarity both differed significantly, in opposite directions.
5. Finally, on one trial, both direct and network distances differed for the two pairs, and in the same direction (figure 5(*e*)). Ratings of distance and similarity both significantly differed, and in the same direction, but similarity differed less extremely than distance.

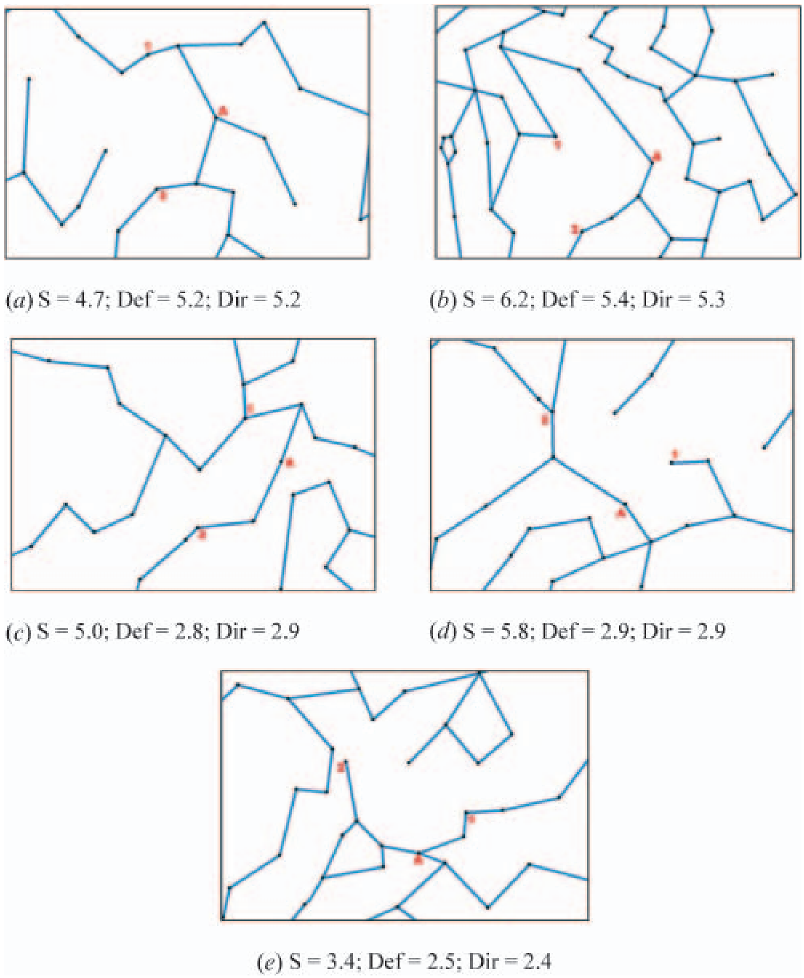
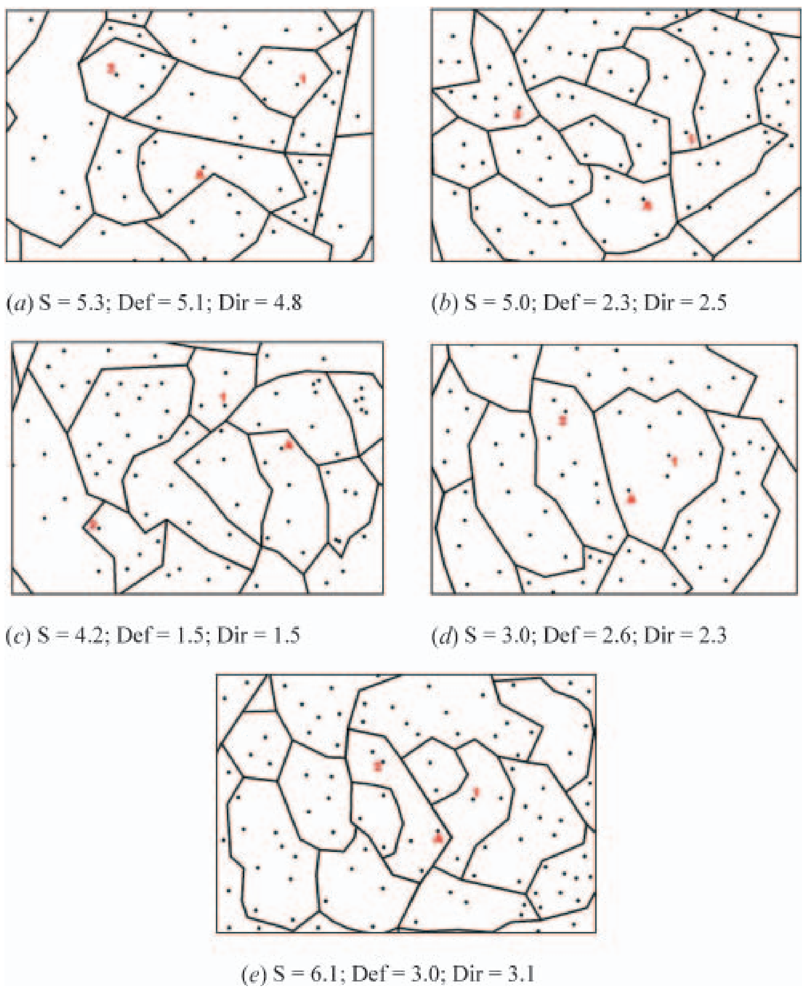


Figure 5. Example network displays and mean ratings. S=Similarity; Def=Default Distance; Dir=Direct Distance (5.0 is A:1 and A:2 are equal; <5.0 is A:1 closer or more similar; >5.0 is A:2 closer or more similar).

The results with network displays can be summed up as follows. Participants again interpreted default (non-specific) instructions as requests to estimate direct or straight-line distances, as in point displays. However, similarity judgements were based on network distances, as we found in previous studies (Fabrikant *et al.* 2004). As with the point-display trials, distance ratings generally followed actual direct distance relationships quite closely, except for the operation of optical illusions such as the vertical and filled-space illusions that alter perceived distances a little.

3.2.3 Monochromatic region displays. The pattern of results with the nine monochromatic region displays was also very clear. On these trials, the main effect of instruction condition was statistically significant on six of the nine display trials. On each of these six, pairwise comparisons again showed that judgements in the *similarity* condition differed from those in the two distance conditions, which did not differ from each other on any trial. The two distance judgements never differed by more than three-tenths of a scale point. The patterns of response across displays followed one of five patterns:

1. On two trials, distances were equal for the two pairs of comparison points, and their regional relationships with respect to 'A' were equivalent; that is, the number of regions intervening between A:1 was the same as for A:2 (figure 6(*a*)). On these trials, ratings of distances and similarity did not differ significantly from each other or from neutral.
2. On three trials, distances differed for the two pairs of comparison points, but their regional relationships with respect to 'A' were equivalent (figure 6(*a*)). For example, on one trial, '1' was closer to 'A' than was '2', but both were one region away from 'A'. On these trials, ratings of distance significantly differed from neutral, but ratings of similarity did not significantly differ from neutral.



(*a*) S = 5.3; Def = 5.1; Dir = 4.8

(*b*) S = 5.0; Def = 2.3; Dir = 2.5

(*c*) S = 4.2; Def = 1.5; Dir = 1.5

(*d*) S = 3.0; Def = 2.6; Dir = 2.3

(*e*) S = 6.1; Def = 3.0; Dir = 3.1

Figure 6. Example monochromatic region displays and mean ratings. S=Similarity; Def=Default Distance; Dir=Direct Distance (5.0 is A:1 and A:2 are equal; <5.0 is A:1 closer or more similar; >5.0 is A:2 closer or more similar).

3. On two trials, distances differed for the two pairs of comparison points, as did their regional relationships with respect to 'A'; that is, the number of regions intervening between the region of the comparison point ('1' or '2') and that of 'A' differed, but neither point shared a region with 'A' (figure 6(*c*)). On these trials, ratings of distance significantly differed from neutral, as did ratings of similarity, although much less than did ratings of distance.
4. On one trial, distances differed for the two pairs of comparison points, and the closer pair shared the same region, while members of the more distant pair were in two neighbouring regions (figure 6(*d*)). On this trial, ratings of distance differed significantly from neutral, as did ratings of similarity, equivalently to the difference in distance.
5. On one trial, distances differed for the two pairs of comparison points, but the more distant pair shared the same region, while members of the closer pair were in two neighbouring regions (figure 6(*e*)). On this trial, ratings of distance significantly differed from neutral, as did ratings of similarity, but in the opposite direction of distance.

The results with monochromatic region displays can be summed up as follows. As with point and network displays, participants viewing monochromatic region displays again interpreted default (non-specific) distance instructions as requests to estimate direct or straight-line distances. However, similarity judgements were based almost completely on region membership, as we found in previous studies (Fabrikant *et al.* 2006). In particular, points in the same region were rated as more similar than points in different regions, irrespective of the distance between the points. When both pairs consisted of points in different regions, the pairs were rated as equally similar, irrespective of the distance between their points. As with the point-display and network trials, default distance ratings generally followed actual direct distance relationships quite closely.

3.2.4 Coloured region displays. Finally, region colour had a clear effect on responses to the 23 coloured region displays. On these trials, the main effect of instruction condition was statistically significant on 18 of the displays. On each of these 18, pairwise comparisons again showed that judgements in the *similarity* condition differed from those in the two distance conditions, which did not differ from each other on any trial. The two distance judgements did not differ by more than nine-tenths of a scale point, which while larger than differences on the other sets of trials, still did not reach statistical significance. The patterns of response across displays followed one of eight patterns. Two of the patterns (figure 7) kept the distances between the comparison documents equal; thus, distance played no role in similarity judgements on these trials:

1. On three trials, distances were equal for the two pairs of comparison points, their regional relationships with respect to 'A' were equivalent, and region hues did not favour one pair over the other; that is, neither '1' nor '2' was in a region with the same colour as the region containing 'A' (figure 7(*a*)). On two of these trials, similarity and distance ratings were not significantly different from each other or from neutral. On the third, A:1 were rated a little closer, and significantly so, than A:2 (in both distance conditions). This was apparently because of the vertical illusion, discussed above. We avoided designing trials in the present study that contrasted vertical and horizontal extents, but this one trial was inadvertently included.

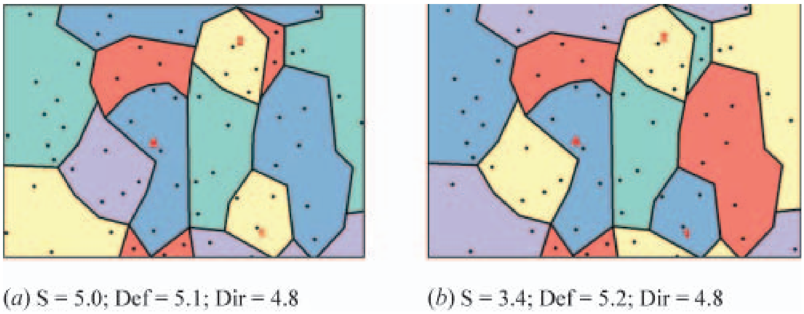


Figure 7. Example coloured region displays and mean ratings. Distances between comparison documents were equal in these trials and played no role in similarity judgments. S=Similarity; Def=Default Distance; Dir=Direct Distance (5.0 is A:1 and A:2 are equal; <5.0 is A:1 closer or more similar; >5.0 is A:2 closer or more similar).

2. On one trial, distances were equal for the two pairs of comparison points, and their regional relationships with respect to 'A' were equivalent, but region hues favoured one pair over the other (figure 7(*b*)). That is, 'A' and '1' were in regions of the same colour, while '2' was in a region of a different colour. On this trial, distance ratings were not significantly different from each other or from neutral, but A:1 were rated as significantly more similar than A:2.

The other six patterns of response (figure 8) occurred on trials in which the distances between the comparison documents were not equal. Thus, distance could play a role in similarity judgements on these trials, possibly strengthening hue or region effects, or working against them:

1. On five trials, distances differed for the two pairs of comparison points, but their regional relationships with respect to 'A' were equivalent, and region hues did not favour one pair over the other (figure 8(*a*)). On all five of these

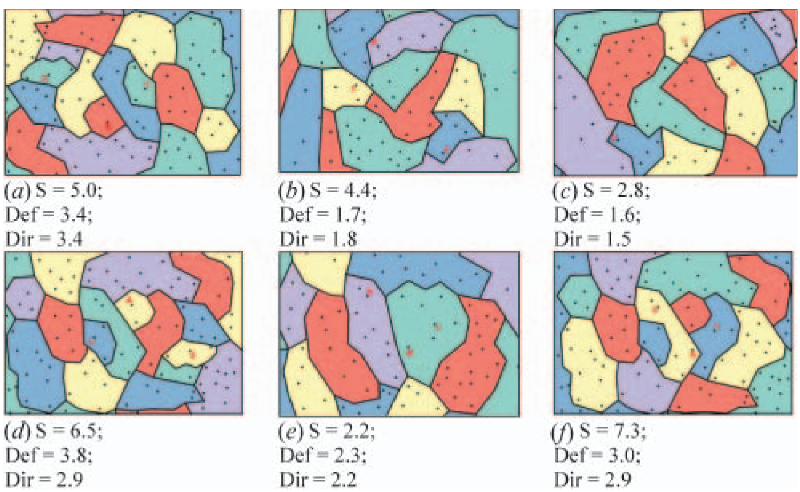


Figure 8. Example coloured region displays and mean ratings. Distances between comparison documents were unequal in these trials and could play a role in similarity judgments. S=Similarity; Def=Default Distance; Dir=Direct Distance (5.0 is A:1 and A:2 are equal; <5.0 is A:1 closer or more similar; >5.0 is A:2 closer or more similar).

trials, distance ratings were significantly different from neutral, while similarity ratings were not significantly different from neutral. That is, even though the two pairs of comparison points differed in distance on these trials, they were not seen as differing in similarity because neither regional relationships nor region colour suggested similarity differences to the participants.

2. On four trials, distances differed for the two pairs of comparison points, and region hues did not favour one pair over the other, as in the 8a trials (figure 8(*b*)). However, regional relationships of '1' and '2' to 'A' were not equivalent—whichever pair was closer also had fewer regions intervening between the two points. On these trials, distance ratings were significantly different from neutral. Similarity ratings differed weakly from neutral, in favour of the closer pair, a pattern that reached statistical significance on two of the trials (though it was clearly weaker than the difference from neutrality for the distance ratings). Thus, distance can affect similarity in coloured region displays when it coincides with a noticeable difference in the number of intervening regions.
3. On four trials, distances differed for the two pairs of comparison points, and their regional relationships with respect to 'A' were roughly equivalent, but region hues favoured the closer pair over the other (i.e. 'A' and '1' were closer *and* in regions of the same colour) (figure 8(*c*)). Both distance and similarity ratings were significantly different from neutral, though the effect on distance was generally larger than the effect on similarity.
4. On four trials, distances differed for the two pairs of comparison points, and their regional relationships with respect to 'A' were roughly equivalent, as in the 8c trials (figure 8(*d*)). However, on these 8d trials, region hues favoured the more distant pair over the closer pair; that is, 'A' and '1' were further apart but in regions of the same colour. Both distance and similarity ratings were significantly different from neutral, but in the opposite direction from each other. Clearly, region colour rather than distance determined similarity when the two visual variables 'contradicted' each other.
5. On one trial, '1' was within the same region as 'A' (and thus in a region of the same colour), while '2' was in a neighbouring region (of a different colour) (figure 8(*e*)). The pair within the same region, A:1, was also closer. This pair was rated as much closer and much more similar, both significant effects.
6. Finally, on one trial, '2' was within the same region as 'A' (and thus in a region of the same colour), while '1' was in a neighbouring region (of a different colour) (figure 8(*f*)). However, the pair within the same region, A:2, was also further apart. This pair was rated as significantly more similar but significantly further apart. Hence, region membership and colour, rather than distance, determined similarity.

The results with coloured region displays can be summed up as follows. As with all other display types we tested, participants viewing coloured region displays again interpreted default (non-specific) distance instructions as requests to estimate direct or straight-line distances. Like the monochromatic region displays, similarity judgements were again based almost completely on region membership. However, when pairs of comparison points did not share a region, colour (specifically hue) powerfully influenced ratings of similarity even as it had no discernible effect on distance ratings whatsoever. As with the other display types, distance ratings generally followed actual direct distance relationships quite closely, except for the operation of optical effects such as the vertical illusion.

4. Discussion

This study replicated our previous findings about the relationship of various types of distance and other visual variables to judgements of similarity for point (Montello *et al.* 2003), network (Fabrikant *et al.* 2004), and region display spatializations (Fabrikant *et al.* 2006). However, in the present study, we provide direct evidence for the relationship of similarity judgements to distance, in so far as we can compare judgements of similarity to judgements of relative distance made on the same spatialized displays. We also provided clear evidence that people estimate direct (straight-line) distance when they are not told specifically what form of distance to estimate (i.e. under default instructions). There are no differences between people's estimates of distance under default (no specified) and direct distance instructions for point, network, and region spatializations. Default distance instructions are interpreted as requests for estimates of direct distance in spatializations. Also, we find that well-known optical effects such as the vertical illusion (Gregory 1987) and the space-filling interval illusion (Thorndyke 1981) affect distance judgements in spatializations and can thus affect, at least to some degree, the operation of the first law of cognitive geography.

In point displays, judgements of similarity parallel judgements of direct distance (default distance too, of course), at least in the ordinal sense that greater differences in judged distances between two pairs of document points result in greater differences in judged similarity. However, we find that similarity judgements are much more variable than distance judgements. People do not judge relative similarities as consistently as they do relative distances, so that two participants who both see one pair of points as being much closer than another pair tend not to agree as much about the relative similarity of the pairs. This suggests that the distance–similarity metaphor is not as self-evident to people viewing information displays or as consistently applied as is commonly assumed in the information visualization literature (Card *et al.* 1999). We also confirm with point displays that emergent point features (i.e. apparent clusters or linear arrangement of points) act oppositely on distance judgements than they do on similarity judgements—emergent features make points appear (a little) further apart but more similar. The effect of features on distance reflects the perceptual effect of a dense aggregation of points in visual space. In contrast, we believe the effect of features on similarity reflects a *con*ceptual effect. Points belonging to a common emergent feature are interpreted to be more similar, even if they are seen as further apart, essentially a type of region effect. In figure 1, for example, the emergent cluster is especially noticeable, as it is visually highlighted with a blue, cloud-like shading (but given suitable point densities, the effect is strong even without such highlighting, as we showed in Montello *et al.* 2003). If we adhere to Wise *et al.*'s (1995) contention that spatializations should be designed to 'reveal relationships between documents in a manner similar if not identical to the way the natural world is perceived' (p. 52), the points in the cluster labelled 'drawing, graphs' (lower middle) should be made part of the cluster 'graph, graphs' (lower left) or at least moved closer to it than to the nearby cluster 'algorithm technique', with which it is presumably less similar in content.

In network and region displays, judgements of similarity mostly diverge from judgements of direct distance, as we expected from our previous research. Replicating our earlier studies, similarity judgements in network displays correspond to distance along links rather than direct distance across links. In contrast, distance judgements themselves reflect direct distance across links. These findings provide some support for the validity of geographic network transportation maps (the London underground map being a classic example), in so far as designers readily distort the absolute locations of stations, and thus their direct distance interrelationships, for aesthetic reasons. At the same time, our findings also cast doubt on the validity of the designers' assumption that metric distance along network links is irrelevant to people, that only link topology matters (Ruggles and Armstrong 1997, Fabrikant *et al.* 2004).

Similarity judgements in region displays mostly correspond to region membership rather than direct distance within or across regions, and are strengthened or weakened appropriately by region hues. This finding suggests a redesign solution for point spatializations such as that shown in figure 1, if distance relationships cannot be adequately controlled through the spatialization algorithm. Colour hues could be systematically applied to point display spatializations to visually highlight semantically related cluster regions that contain points representing similar documents that are not near each other. For example, the clusters 'drawing, graphs' and 'graph, graphs' in figure 1 could be shaded with the same colour hue to visually emphasize their semantic overlap, even though they are farther apart from each other than from the less related cluster 'algorithm technique'.

This result should not be surprising for geographers used to working with choropleth maps (i.e. statistical maps) or area class maps (e.g. land use zones, or geologic maps). However, it might have some design implications for widely used political reference maps. In those maps, a limited set of hues (typically four or five) is applied to visually distinguish countries, mostly for aesthetic reasons. Naive map users might wrongly infer semantic relationships based on identical colour hues where there are no such relationships intended.

5. Conclusions and outlook

All the results taken together confirm and refine a previously proposed theoretical spatialization framework for designing cognitively adequate spatialization displays presented by Skupin and Fabrikant (2003) and Fabrikant and Skupin (2005). The empirically validated framework is firmly grounded in GIScience theory, including cognitive and experiential principles, and relies on longstanding cartographic design practice.

Our findings also provide much needed empirical evidence on the conditions under which people employ distance when assessing similarity between data objects in two-dimensional spatialized views and, when they do, which type of distance. That is, the findings provide evidence for the effectiveness and operation of the distance–similarity metaphor in the design of different types of spatialization. They may also give insight into how the first law of cognitive geography operates when people explore the similarity of geographic features depicted in cartographic maps and other information displays.

Our ongoing research aims to empirically evaluate the distance–similarity metaphor for 3D spatialized displays. We are adapting our well-established experimental design framework developed for 2D experiments to the 3D desktop environment. In these 3D studies, we distinguish 2D from 3D in terms of the apparent dimensionality of the image on the computer screen (i.e. depiction of a 2D pattern vs. a 3D pattern, on the 2D screen). We also distinguish monoscopic from stereoscopic 3D images; the former get depth from perceptual cues such as shading and motion parallax (information available to one eye), while the latter gets depth also from stereopsis, the 'true 3D' visual sense that derives from the integration of the offset images from each eye. Finally, we distinguish static images from dynamic images that are interactive (the movement of the image may be actively controlled by the viewer). We will present the results from these experiments in follow-up publications.

Acknowledgements

Funding by the National Geospatial-Intelligence Agency (NMA-201-00-1-2005) is greatly appreciated. We thank Amy Linker, Richard Middleton, Jeff Onsted, Marco Ruocco, and Ben Turner for their assistance in preparing materials and running the experiments. We are also grateful for David Mark's continued insightful input, discussion, and brainstorming since the inception of this project.

References

CARD, S.K., MACKINLAY, J.D. and SHNEIDERMAN, B., 1999, *Readings in Information Visualization: Using Vision to Think* (San Francisco: Morgan Kaufmann).

FABRIKANT, S.I., MONTELLO, D.R. and MARK, D.M., 2006, The distance–similarity metaphor in region-display spatializations. *IEEE Computer Graphics & Application*, **26**, pp. 34–44.

FABRIKANT, S.I., MONTELLO, D.R., RUOCCO, M. and MIDDLETON, R.S., 2004, The distance–similarity metaphor in network-display spatializations. *Cartography and Geographic Information Science*, **31**, pp. 237–252.

FABRIKANT, S.I. and SKUPIN, A., 2005, Cognitively plausible information visualization. In *Exploring Geovisualization*, M.J. Kraak (Ed.), pp. 667–690 (Amsterdam: Elsevier).

GREGORY, R.L. (Ed.), 1987, *The Oxford Companion to the Mind* (Oxford: Oxford University Press).

MONTELLO, D.R., FABRIKANT, S.I., RUOCCO, M. and MIDDLETON, R.S., 2003, Testing the first law of cognitive geography on point-spatialization displays. In *Spatial Information Theory: Foundations of Geographic Information Science, Conference on Spatial Information Theory (COSIT) 2003, Lecture Notes in Computer Science 2825*, W. Kuhn, M.F. Worboys and S. Timpf (Eds), pp. 316–331 (Berlin: Springer).

RUGGLES, A.J. and ARMSTRONG, M.P., 1997, Toward a conceptual framework for the cartographic visualization of network information. *Cartographica*, **34**, pp. 33–48.

SUI, D.Z., (Guest Ed.), 2004, Forum on Tobler's First Law of Geography. *Annals of the Association of American Geographers*, **92**, pp. 269–310.

SKUPIN, A. and FABRIKANT, S.I., 2003, Spatialization methods: A cartographic research agenda for non-geographic information visualization. *Cartography and Geographic Information Science*, **30**, pp. 95–119.

THORNDYKE, P.W., 1981, Distance estimation from cognitive maps. *Cognitive Psychology*, **13**, pp. 526–550.

TOBLER, W.R., 1970, A computer movie simulating population growth in the Detroit region. *Economic Geography*, **42**, pp. 234–240.

WISE, T.A., THOMAS, J.J., PENNOCK, K.A., LANTRIP, D.B., POTTIER, M., SCHUR, A. and CROW, V., 1995, Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In *Proceedings, IEEE Information Visualization '95*, 30–31 October, Atlanta, GA, pp. 51–58.