Klippel, A., Hardisty, F., & Li, R. (forthcoming, 2011). Interpreting spatial patterns: An inquiry into formal and cognitive aspects of Tobler's first law of geography. *Annals of the Association of American Geographers*.

Interpreting Spatial Patterns -

An Inquiry into Formal and Cognitive Aspects of Tobler's First Law of Geography

Abstract

The characterization, identification, and understanding of spatial patterns are central concerns of geography. Deeply rooted in the notion that geographic location matters, one testable assumption is that near things are more related than distant things—a concept often referred to as Tobler's first law of geography. One means of quantifying this assumption is using measures of spatial autocorrelation. Several such measures have been developed to test whether a pattern is indeed clustered, or dispersed, or whether it is, from a statistical perspective, random. To shed light on how spatial patterns are understood from a cognitive perspective, this article reports results from studies of spatial pattern interpretation represented in maps. For the purpose of experimental validation we used a two-color map. We systematically varied the ratio of the colors as well as the level of significance of clustering and dispersion; we targeted two groups: experts and non-experts. The task for both experts and non-experts was to sort patterns according to five specified categories of spatial autocorrelation structures. The results show clearly that patterns are understood on the basis of the dominant color, by both experts and non-experts. A third experiment, using a free classification paradigm, confirmed the dominance of the color effect. These results are important as they point to critical aspects of pattern perception and understanding that need to be addressed from the perspective of spatial thinking, especially how people relate concepts of randomness with spatial patterns (represented in maps).

Keywords: Tobler's First Law; Spatial Autocorrelation; Interpretation of Spatial Patterns

The Perception and Understanding of Spatial Patterns

Everything that has a location in geographic space inevitably creates or contributes to a spatial pattern. It does not come as a surprise that geography devotes a large amount of attention to analyzing, identifying, explaining spatial patterns, both in qualitative and quantitative terms (e.g., Abler, Adams, and Gould 1971; Dacey 1971; Blok et al. 1999; Griffin et al. 2006; Kwan and Ding 2008). Take, for example, Bailey and Gatrell's (1995) division of spatial analytic methods: point pattern analysis; analysis of spatially continuous data (Cressie 1990), analysis of area data (Moran 1948; Anselin 1995), and spatial interaction data (for example, gravity models Openshaw 1979 and network problems). The first three explicitly privilege physical space as the metric space that the methods operate in. In the last category, only network problems potentially could be evaluated without reference to physical space.

It also does not come as a surprise that geography education teaches students to understand and analyze patterns and to establish connections between spatial patterns and processes potentially responsible for causing patterns (e.g., O'Sullivan and Unwin 2003; National research council 2006). It is, indeed, worthwhile to discuss the role of spatial patterns in research and education from the perspective of spatially primitive concepts (Golledge 1995; Marsh, Golledge, and Battersby 2007). Spatial patterns can be derived from the primitive concept of location but are themselves not considered primitive in Golledge's framework. In contrast, work by the French geographer Brunet (1980) considers a limited number of spatial patterns themselves (or, to be more precise, the processes that lead to specific spatial patterns) as being primitive to a language of space.

In the light of the latter, the analysis of spatial patterns becomes particularly intricate if we look not only at spatial patterns created by the location that entities have in space, but at the

combination of spatial and attribute values. From a quantitative perspective, this theoretical construct is most famously summarized by Tobler's (1970) first law of geography (TFL): *Everything is related to everything else, but near things are more related than distant things*.

As Miller (2004) notes, the concept of TFL is implicit in the practice of spatial analysis. His article was part of a symposium that was held in honor of TFL which produced several research papers. To give further examples of the centrality that geographers give this aspect of understanding space and spatial relations, the following quote from Sui's (2004) article stresses the importance of TFL. In quoting Gould (1979) he writes "an innocent ignorance of Tobler's work now constitutes a constraint on the geographic imagination. That is to say, if a graduate student is not aware of certain pieces of Tobler's research, [his]/her own research abilities are jeopardized because [he]/she cannot gain a new and crucial perspective'' (p. 147).

In similarly strong terms O'Sullivan and Unwin (2003) write in their textbook on geographic information analysis that "If spatial autocorrelation [as a way to formalize TFL] were not commonplace, geographic analysis would be of little interest, and geography would be irrelevant." (p. 28) and they continue "If geography is worth studying at all, it must be because phenomena do not vary randomly through space." (O'Sullivan and Unwin 2003, p. 28)

The relationship between distance and similarity has additionally been extended into the realm of using space (distance) to represent semantic similarity (Montello et al. 2003), a concept also referred to as *spatialization* (Skupin and Fabrikant 2007). Montello and colleagues named their studies "The first law of cognitive geography", stressing the fact that distance ubiquitously is associated with similarity (or dissimilarity). It has to be noted though that several factors, such as connectedness contribute to distance estimations in map-like displays (e.g., Klippel et al. 2005; Fabrikant, Montello, and Mark 2006).

One interesting perspective on TFL is that it can be used as a hypothesis that can be tested (and be accepted or rejected). In other words, we can put the assumption that TFL is true (positive spatial autocorrelation) to the test and scrutiny of statistical inference. Of course, the broad application of the concept of statistical inference is a contested practice in geography (e.g., Gould 1970). If a spatial pattern is statistically significant it does not automatically follow that it is meaningful and likewise, a spatial pattern that does not turn out to be statistically significant can be meaningful in other ways. We will discuss various aspects of statistical significance and the importance associated with the concept of randomness in the light of the results of this paper.

However, for the purpose of this paper it is important to engage the concept of statistical significance for analyzing patterns; specifically, for revealing the autocorrelational structure that potentially exists in a study area. To simplify experimental design, this research focuses on one autocorrelation measure, join count analysis. This measure, while a simple one, is still actively used for research (Bell, Schuurman, and Hameed 2008). The results of our experiments, however, can be discussed in the light of a broader framework on a) spatial autocorrelation measures, b) aspects of spatial thinking, particularly the challenge of connecting the concepts of randomness to spatial patterns, c) factors that influence map interpretation.

Spatial autocorrelation analysis – join count

There is a distinction that is made in all spatial autocorrelation measures (as well as in, for example, point pattern analysis) which can be differentiated using various spatial analysis methods. This distinction is between three kinds of patterns:

- Clustered
- Dispersed

- Random

Figure 1 shows the US election results for 2004 and 2008 as a classic example of join count analysis (Cliff and Ord 1973; Goodchild 1986). The data characteristics are that we have a categorical distinction between two types of areas. Examples include voting behavior (Republican or Democratic), simple land use classifications (e.g., developed or undeveloped) and/or accidents occurring in a certain area or not (Bell, Schuurman, and Hameed 2008). A higher level distinction is whether we are interested in clustering (whether a specific autocorrelational structure exists in the dataset as a whole) or in clusters (identifying the presence of autocorrelation within sub-areas of a dataset). Clustering and clusters correspond to the distinction of global and local autocorrelation, respectively (Anselin 1995)[1].

To decide, from a statistical perspective, which of the three autocorrelational structures exist in a study area, a z-score can be calculated on the basis of the number a specific type of join, the expected number of joins, and the expected standard deviation (Cliff and Ord 1973). The three types of joins in a two color map are color 1-color 1, color 2-color 2, and color 1-color2. Additionally, there are different ways to define adjacencies of areas, such as queen, rook, or bishop contiguities. The last important aspect to keep in mind is that it is possible to find a (statistically significant) clustering for one color but not the other.

[1] There are several studies in geography and cartography that are concerned with the perception and identification of individual clusters in maps Slocum (1983); Lewandowsky et al. (1993); Sadahiro (1997) the measure of map complexity Olson (1975); Bregt and Wopereis (1990), similarity comparisons Steinke and Lloyd (1981), or influences of different visualization methods on cluster detection Walter (1993). In contrast to these studies our research focuses on clustering (as identified by spatial autocorrelation methods) as a global phenomenon in spatial patterns and does not aim for identifying the best visualization method to represent clusters.

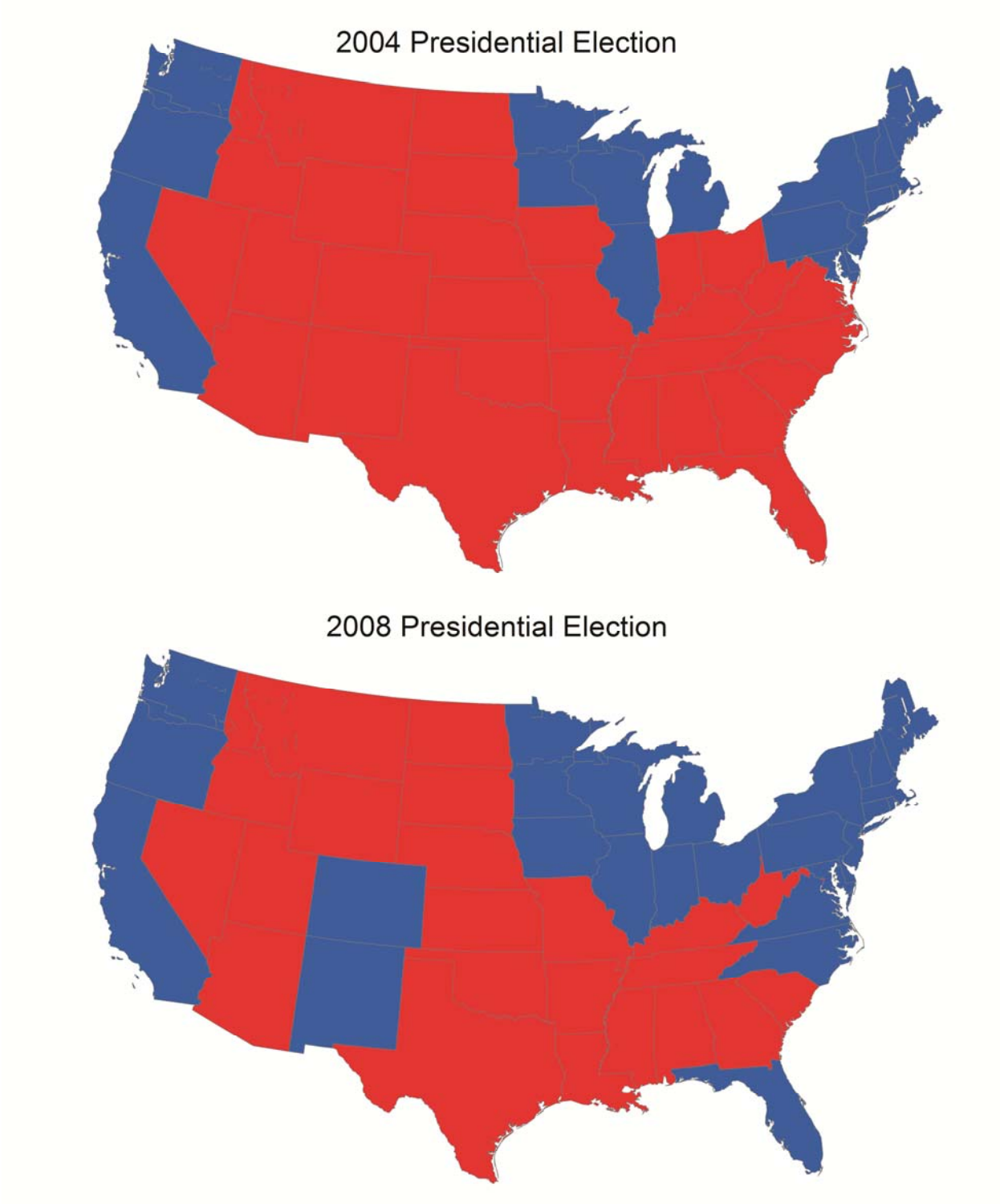

Figure 1. Two-color map: Election results (2004 – top, 2008 – bottom).

Hypothesis / Research question

The open question we are addressing in this paper is: when and how does a spatial pattern (statistically significant clustering or dispersion) represented on a map become perceptually and conceptually salient to someone interpreting the map? In other words, we find that the most commonly used statistical significance level is set to a p value equal or smaller than 0.05 (or equal or larger than a z score of 1.96); but, is this level identical to the significance level where a human map reader would set the cut-off point at which a pattern changes from being perceived as random to being identified as clustered (or dispersed)?

Methods

To shed light on these research questions, we performed two experiments using a grouping (classification) paradigm that is widely used in psychology in research on categorization (Medin, Wattenmaker, and Hampson 1987; Pothos and Chater 2002) and also showed its usefulness in spatial and cartographic sciences (Mark and Egenhofer 1994; Knauff, Rauh, and Renz 1997; Klippel and Li 2009). The first experiment we ran compared the classification behavior of experts versus non-experts to reveal the influence of geographic education (geographic expertise) on the ability to understand spatial patterns from a statistical perspective. A second experiment deepened our insights into spatial pattern understanding by using non-experts and tasked them with freely classifying (categorizing) patterns, that is, we addressed the question how humans naturally categorize (make sense of) spatial patterns.

Participants

Participants for the three conditions (Expert, Non-experts, and Free-Classifiers) came from two distinct pools. Their characteristics were as follows: Experts were students of an advanced 400 level spatial analysis class. The experiment was conducted at a time such that spatial autocorrelation had been discussed in class and that the underlying principles of join count statistics were known. Additionally, a prerequisite for this spatial analysis class is an introductory spatial analysis class that introduces join count statistics through lectures, homework, quizzes, and exams. We had 20 participants, nine female, average age 22.3.

For the Non-experts and Free-classifiers, we recruited participants from an Information Science and Technology (IST) class. Spatial autocorrelation and spatial analysis is not part of the curriculum in IST. We had 20 participants in the Non-expert group, 4 female, average age 19.8. Each student received extra credits in the class where s/he was recruited from as compensation of his/her participation (details on the free-classification experiments are provided in a later section).

Design

Our experiment has a 2 x 3 x 3 factorial design:

1. The first factor distinguishes the effect of expert knowledge on identifying statistically significant patterns. It is reflected in the two groups of participants, experts and non-experts, and is a between subjects factor. The abbreviation we use is *ex_nex* for expert and non-expert.
2. The second factor captures the effect of the statistical significance level on identifying statistically significant patterns (referred to as *sig_level*). Within this factor three levels are distinguished (for details see Section on materials).

3. The third factor addresses the potential influence of the ratio of areas of a certain type. In our case, the color ratio of a two-color map (blue and green, abbreviated as *ratio*). Within this factor we distinguish again three levels (for details see Section on materials).

The dependent variable we collected was the number of correctly classified patterns in five statistical pattern categories. These pattern categories (abbreviated as *pat_cat*) are the result of statistically measuring the spatial autocorrelation structure present in a region. The following cases can be distinguished:

- Only blue areas are significantly clustered (abbreviated as *B_Clust*)
- Only green areas are significantly clustered (abbreviated as *G_Clust*)
- Blue and green areas are significantly clustered at the same time (abbreviated as *BG_Clust*)
- The pattern of blue and green areas is dispersed (abbreviated as *Disp*)
- The pattern of blue and green areas is random (abbreviated as *Rand*).

While it would be possible to analyze the data using MANOVA (Multiple Analysis of Variance) by treating each pattern category individually, we settled on using repeated measures ANOVA as all participants in the two groups (experts and non-experts) created measures for all five pattern categories. This latter approach best reflects the nature of the data (see, Tabachnick and Fidell 2007). To complete the classification analysis, we also analyzedmisclassification behavior by creating confusion matrices.

A last question that we addressed looked into the natural classification (behavior) of non-experts to shed light on the question of how many patterns participants would naturally distinguish. For this question, a separate free-classification experiment was designed.

Materials

To realize the design detailed above, we created different spatial patterns visualized as two-color maps. All patterns consisted of a 10x10 grid structure, that is, a total of 100 grid cells that were either blue or green. This type of material is used frequently in geography textbooks (O'Sullivan and Unwin 2003; Longley et al. 2006; Burt, Barber, and Rigby 2009) to explain the concept of spatial autocorrelation (join count analysis) for phenomena that fall sensibly into two classes (e.g., Republicans and Democrats O'Sullivan and Unwin 2003 or the presence and absence of accidents Bell, Schuurman, and Hameed 2008). This general design allowed us to control for the factors (the significance level, the ratio of blue and green cells, and the dependent variable of five different types of patterns) detailed above that determine the spatial autocorrelation structure and potentially how the patterns are perceived.

For the realization of our patterns we follow the advice of Cliff and Ord (1981), who recommend Monte Carlo simulation to compare a given distribution against an appropriate reference distribution. The patterns were generated in a two-pass manner for each blue/green ratio: First 100,000 patterns were generated with the required proportion of green (and blue, respectively) cells. The required proportion is arrived at by starting with all cells being blue, and then randomly picking cells to turn green until the appropriate number (as specified by the different ratios, see below) is reached. We developed a pluggable system which allowed for using various random number generation processes in join count statistics, and we compared the

Blum Blum Shub (Blum, Blum, and Shub 1986), Mersenne Twister (Makoto and Takuji 1998), Cellular Automaton Rule 30 (Hortensius et al. 1989), linear congruential (Borosh and Niederreiter 1983), and cryptographically strong random number generators. The Mersenne Twister was selected for its speed and high-quality randomness. The observed standard deviation is noted from the counts of the different types of joins in the 100,000 runs. From the different options to determine contiguity for the join count analysis, we chose rook contiguity, that is, only cells that actually share a boundary were considered to be adjacent to each other.

In the second pass, lists were created of z-score ranges for the different join types required for the various test stimuli (i.e., the icon sets). Then, a random pattern was generated and the z-scores calculated by subtracting the expected number of joins from the observed number and then dividing by the observed standard deviation. The pattern was then tested to see if its z-scores are in the appropriate ranges. If so, it is turned into an icon and its statistical properties were noted. This process is repeated until all necessary icons are generated (see more details below).

After defining this way of determining the significance level, and to create patterns we used the following rationale to realize the design of our experiments (as detailed above). All the details are additionally summarized in Table 1. For the factor sig_level, we chose three common significance levels (ranges): from 0.1 to 0.05 (a z score from 1.645 to 1.96); from 0.05 to 0.01 (a z score from 1.96 to 2.576); from 0.01 to 0.001 (a z score between 2.576 and 3.291). In the following we will refer to these ranges as P10, P05, and P01, respectively. These three levels of significance can be and were distinguished within each of the four autocorrelational structures mentioned above (B_Clust, G_Clust, BG_Clust, and Disp). Any pattern below a z score of 1.645 (or a p value < 0.1) was considered random (Rand).

The second factor that we were interested in was the potential influence of the ratio of the two colors, that is, how many of the 100 cells were blue and how many of the cells were green. We selected the following three ratios for the experiments, explicitly avoiding extreme ratios:

- G40: 60 blue cells and 40 green cells
- G50: 50 blue cells and 50 green cells
- G60: 40 blue cells and 60 green cells

Furthermore, for each of the potential combinations of four different categories (B_Clust, G_Clust, BG_Clust, Disp), the three levels of significance (P10, P05, P01), and the three ratios (G40, G50, G60) we randomly generated two patterns (see above on how the patterns were generated). That means that we have, so far, 2 x 3 x 3 x 4 = 72 different patterns. For the number of random patterns we chose the following experimental logic: In each category (B_Clust, G_Clust, BG_Clust, Disp) we have 18 spatial patterns, hence, the number of random patterns should be the same, that is, 18 (6 from each ratio). We ended up with a total of 90 different, systematically varied spatial patterns that we used as stimuli in this experiment. Table 1 summarizes this design and Figure 2 shows examples of the resulting patterns.

A note on the color selection: The colors were chosen from the Color Brewer website (Harrower and Brewer 2003) using a qualitative color scheme. We additionally decided on colors suitable for color blind people. The RGB values for green were: 178, 223, 138; and the RGB values for blue were: 31, 120, 180.

Table 1.*The distribution of spatial patterns across different levels of significance and different ratios (blue-green). For reasons of readability the pattern categories are abbreviated.*

Significance Range	40 / 60	50 /50	60 / 40	**Total**
1.645 – 1.96	B: 2, G: 2 BG: 2, D: 2	B: 2, G: 2 BG: 2, D: 2	B: 2, G: 2 BG: 2, D: 2	**B: 6, G: 6** **BG: 6, D: 6**
1.96 – 2.576	B: 2, G: 2 BG: 2, D: 2	B: 2, G: 2 BG: 2, D: 2	B: 2, G: 2 BG: 2, D: 2	**B: 6, G: 6** **BG: 6, D: 6**
2.576 – 3.291	B: 2, G: 2 BG: 2, D: 2	B: 2, G: 2 BG: 2, D: 2	B: 2, G: 2 BG: 2, D: 2	**B: 6, G: 6** **BG: 6, D: 6**
Total	**B: 6, G: 6** **BG: 6, D: 6**	**B: 6, G: 6** **BG: 6, D: 6**	**B: 6, G: 6** **BG: 6, D: 6**	**B: 18, G: 18** **BG: 18, D: 18**
Random pattern	R: 6	R: 6	R: 6	**R: 18**
Total				**90**

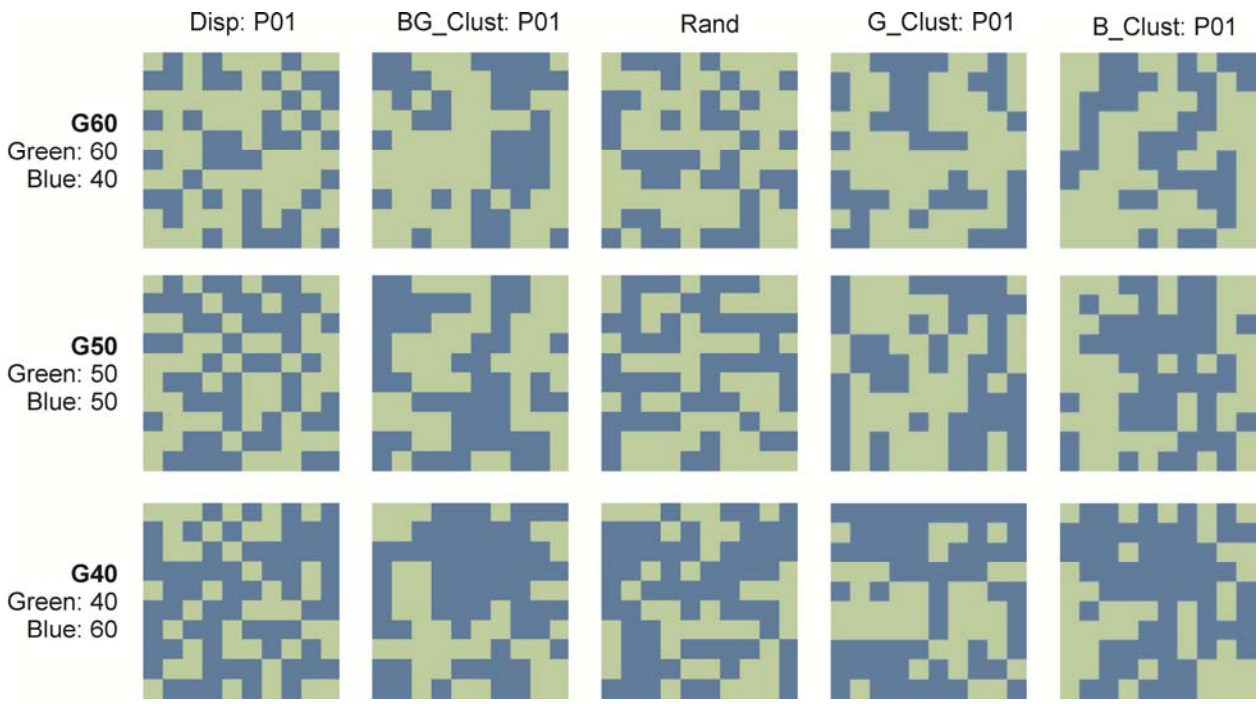

Figure 2. Examples of patterns used in the experiment. Examples for the three different ratios (G60, G50, and G40) for all pattern categories (Disp, BG_Clust, Rand, G_Clust, and B_Clust) at the highest significant level (P.01 corresponding to a z score range of 2.576 – 3.291) are shown.

Procedure

The experiment was organized as a group experiment and took place in a GIScience laboratory at the Pennsylvania State University, Department of Geography. The lab is equipped with 16computers that all have 24'' wide screen displays and are well suited for grouping tasks with large sets of icons. The lab was prepared with view blocks such that participants could only see their own screens.

Participants entered the room together and were randomly assigned to the computers. After providing consent, participants entered their personal data into our custom made software. We collected basic data about their age, gender, and other personal information. Following this, they received instructions in how to classify the patterns. They were reminded of the concept of

statistical significance and that they had to create five distinct groups according to the categories specified above, that is, B_Clust, G_Clust, BG_Clust, Disp, and Rand (see Table 1). They also were presented with one example of each of the five groups.

Afterwards participants received a small warm up task (sorting animal icons) to acquaint them with the interface before they started the main experiment. Figure 3 shows two screenshots: the top part shows the initial screen of the main experiment, the bottom part shows an ongoing mimicked experiment. It is important to note that in the beginning all icons are randomly placed on the left side of the screen. Because we explicitly asked participants to create 5 specific groups, the right side showed 5 groups upon starting the experiment. Icons could be placed into groups by simply using drag and drop. Icons could also be placed from one group into another or back onto the left side of the screen. The experiment was completed once all icons from the left side had been placed into the five groups on the right side.

After the participants finished the main task, they also were presented with a questionnaire to assess their self evaluated knowledge about statistics in general, spatial autocorrelation, and general geographic/cartographic interests.

Figure 3. Screenshot of the grouping tool. The top part shows the initial screen that participants saw with all icons on the left side and predefined groups on the right side. The bottom part shows

a mimicked ongoing experiment. In the third experiment, free classification, the participants did not receive a predefined number of groups but had to create groups themselves. Otherwise the procedures and materials were identical.

Results

We will first descriptively look into patterns that are clustered, that is, either blue areas are clustered (B_Clust), green areas are clustered (G_Clust), or both, blue and green areas are clustered (BG_Clust). We compare this data across all three significance levels, all three ratios, and between experts and non experts.

Figure 4 shows the average number of spatial patterns that were classified correctly (for B_Clust, i.e., only blue is clustered) in the nine combinations that pertain to the two central questions: the influence of different levels of significance (P10, P05, P01) and three different ratios (G40, G50, G60). Additionally, the three charts are separated for experts and non-experts. There are a couple of important observations to make: Independent of the level of significance (P10, P05, P01) we find that spatial patterns in which blue is the dominant color (G40) and blue is significantly clustered a much higher number of spatial patterns are classified correctly. In other words, in the case where blue is significantly clustered and it is the dominant color the number of correctly classified patterns is high. In contrast, if blue is significantly clustered but green is the dominant color, the number of correctly classified patterns is low. However, we do not see a tendency for patterns with higher levels of significance to be more often correctly classified. While the overall pattern is the same for both experts and non-experts, the total numbers of correctly classified spatial patterns seem to be slightly lower for non-experts.

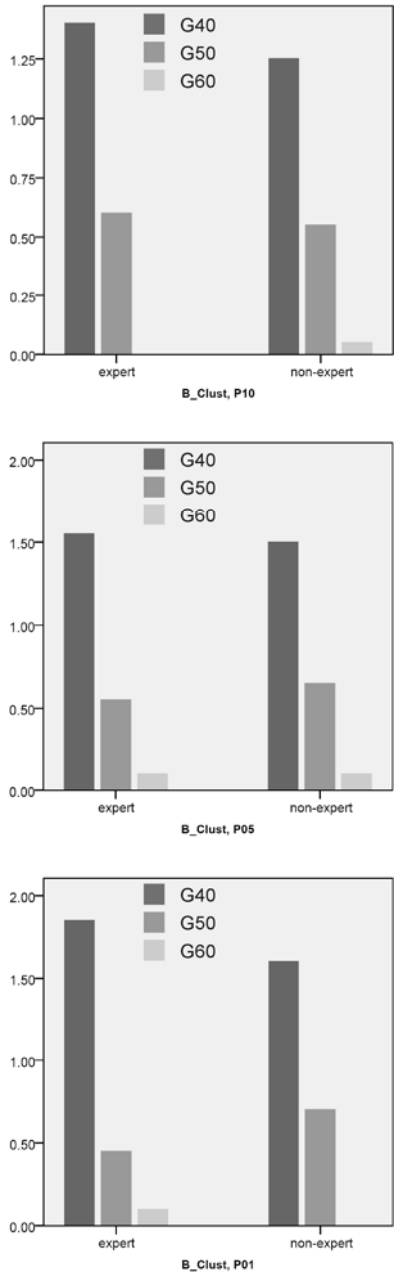

Figure 4. Shown are the mean values of correctly classified patterns where blue areas are significantly clustered across different levels of significance (P10 – top; P05 – middle; P01 – bottom) and different blue-green ratios for both experts and non-experts. For both groups (experts and non experts) the same pattern emerges: For all three significant ranges we find the

same trend, that is, the number of correctly classified patterns degreases with an increase of number of areas of the non-significant (here: green) color.

Figure 5 shows the same distinctions (three significant ranges and the three ratios for both experts and non experts) for significant clusters of only green areas (G_Clust). While we do not find an identical pattern for increasing correct classifications with higher significance, we do find the same general pattern with respect to the effect that the different blue-green ratios have on the number of correctly classified spatial patterns: The higher the number of green areas, the better the classification result for significant green clustering (that is, correctly classified G_Clust patterns) for both experts and non-experts. Other factors such as the significance level do not reveal a clear picture.

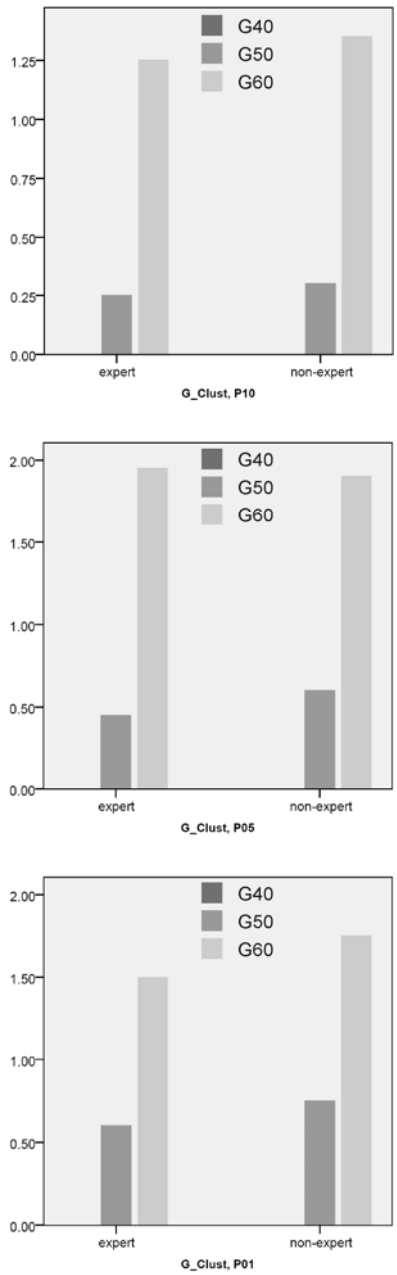

Figure 5. For mean values of significantly clustered and correctly classified green areas we find a very similar pattern (compare Figure 4). In both groups (experts and non-experts) the blue-green ratio is the dominant factor while the different significance levels (separated into three charts) do not provide a significant influence. Values for patterns with the lowest (40) number of green areas frequently are not correctly classified at all.

With these two very clear results regarding the influence of ratio on the number of correctly classified patterns, we now look into the results for the case where both blue and green areas are significantly clustered (BG_Clust). Figure 6 provides the details. The different significance levels (again) do not offer a clear picture of correctly classified or miss-classified spatial patterns, but the blue-green ratios do. This is the case for both the experts and non-experts in the same way: we find that in the case where the task for the participants was to find spatial patterns in which both blue and green are significantly clustered at the same time participants highly favored a blue-green ratio of 50/50 (G50). In case of the lowest significance level, P10, we also observe a tendency that G40 is more often classified correctly than G60.

To summarize, while the different ratios (G40, G50, G60) draw a very clear picture, the different significance levels do not.

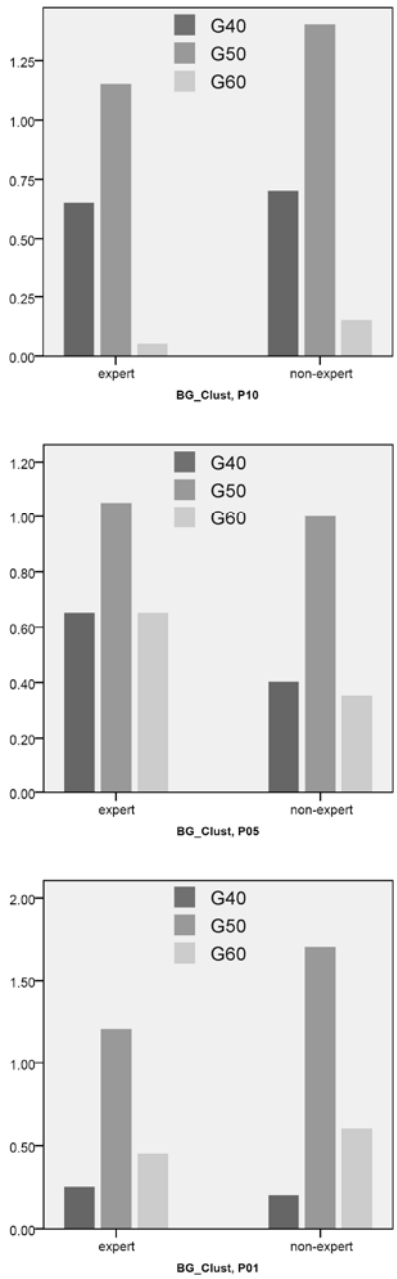

Figure 6. The correctly classified patterns where blue and green are significantly clustered at the same time are shown. The three charts show the three significance levels and the three blue-green ratios for both experts and non-experts. While the different significance levels do not deliver a clear picture, the different blue-green ratios do: In case both, blue and green areas are significantly clustered, we find that the 50/50 ratio yields significantly higher correctly classified patterns than either of the other two ratios.

To reveal the statistical significance of these results, we report the results of two mixed repeated measures ANOVAs (and several contrasts). For the first ANOVA we summed over all ratios and over all significance levels. The purpose of this analysis is to reveal differences between different types of patterns (pat_cat: B_Clust, G_Clust, BG_Clust, Disp, Rand) and experts and non-experts (ex_nex) in general. Figure 7 visually represents the following results. First, Maulchy's test of sphericity shows that the assumption of sphericity is violated. We therefore decided to use the Huynh-Feldt correction as suggested by Cardinal and Aitken (2006). We find that the main effect of pattern category (pat_cat) was highly significant ($F(3.648, 138.644) = 16.704$, $p < 0.001$, $\eta^2 = 0.305$). However, we also find a significant interaction effect that indicates that experts and non experts performed differently across different pattern categories ($F(3.648, 138.644) = 2.514$, $p = 0.05$, $\eta^2 = 0.062$), while the main effect of ex_nex was not significant ($F(1, 38) = 1.66$, $p = 0.205$, $\eta^2 = 0.042$). We defined several contrasts to shed light on relevant aspects of the results. The first addressed the question whether there is a difference between correctly classified blue and correctly classified green patterns. Contrast revealed that there is no significant difference ($F(1,38) = 0.121$, $p = 0.73$, $\eta^2 = 0.003$). Random patterns (from our experience in the classroom and from the facts that humans tend to see patterns everywhere) are harder to correctly classify than significantly clustered or dispersed patterns, but expert should have more correct patterns than non-experts. Contrast revealed that indeed random patterns are significantly less often correctly classified than all other patterns (B_Clust, G_Clust, BG_Clust, Disp) ($F(1,38) = 46.567$, $p < 0.001$, $\eta^2 = 0.551$), and that this task was difficult for both, experts and non-experts, that is, there is no statistical difference between them ($t = 1.061$, $p = 0.295$, $\eta^2 = 0.029$). On the other hand, we do find that the regularity and

probably salience of a dispersed pattern has a positive influence on classifying these patterns correctly ($F(1,38) = 24.889$, $p < 0.001$, $\eta^2 = 0.396$) and that experts were significantly better (overall) at detecting them ($t = 2.12$, $p = 0.041$, $\eta^2 = 0.106$).

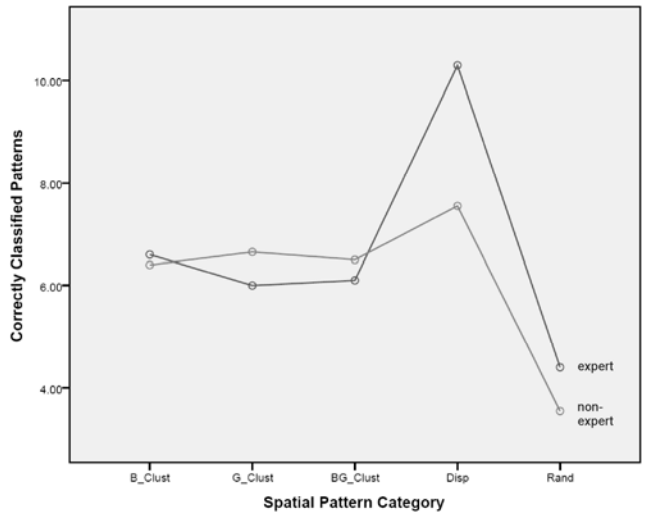

Figure 7. Summed results are shown, that is, all ratios and all p-levels are combined to compare the performances of experts and non-experts across all five pattern categories (pat_cat).

The second ANOVA addresses a finer granularity of analysis: How did the different ratios (ratios) influence the classification results across the five pattern categories (pat_cat) distinguished for experts / non-experts (ex_nex)? Figure 8 shows two rather complex graphs one for experts (top) and one for non-experts (bottom). We are going to look at these graphs in more detail in the following and reveal their message by several individual comparisons. Part of this information has already been shown in Figures 4, 5, and 6 and has been descriptively discussed above.

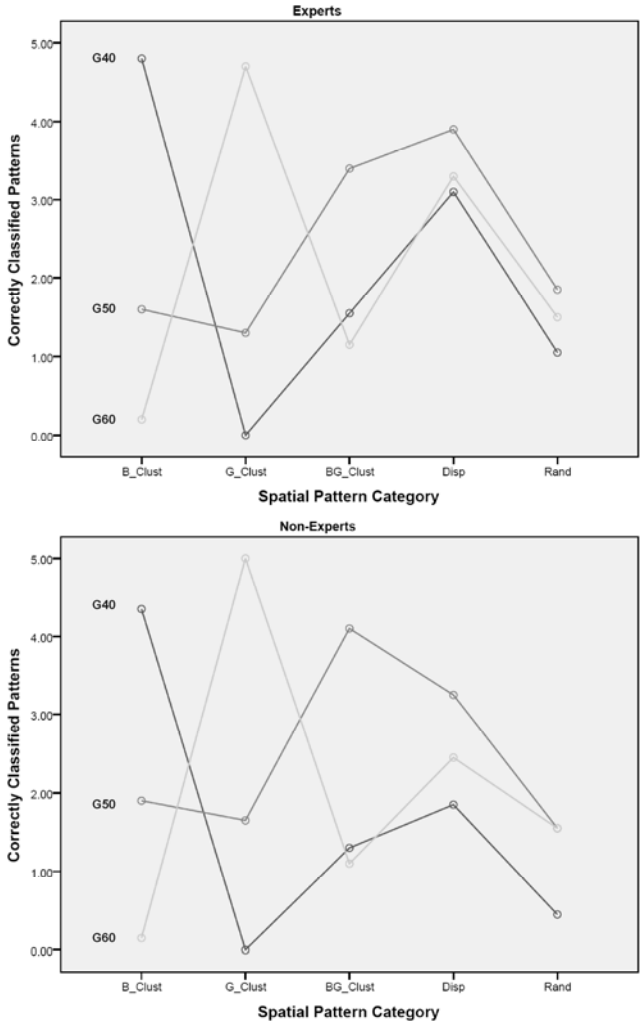

Figure 8. Correctly classified patterns for experts and non-experts by pat_cat and ratios.

We first turn our attention to category one and two, clustering of blue and green, respectively, and discuss how the different ratios influence the success of classification. We have already observed in Figures 4 and 5 that different ratios have a strong influence on the classification behavior, and that the identification of significant clustering of one color is strongly supported by the dominance of that color. While we know already that the main effect pat_cat is significant, it does not come as a surprise that the main effect ratio is highly

significant, too ($F(2,76) = 17.06$, $p < 0.001$, $\eta^2 = 0.309$)[2]. However, the interaction between ratio and pat_cat turns out to be significant as well ($F(8,304) = 94.106$, $p < 0.001$, $\eta^2 = 0.717$) and is therefore the more important result to report. In other words, ratio had a different effect on different pattern categories. To shed light on these significant effects we specified several contrasts. The first contrast looked at the difference between B_Clust (only blue areas are significantly clustered) and G_Clust (only green areas are significantly clustered). While the general difference between these two is not significant, we have a highly significant interaction effect for comparing the pair B_Clust/G40, B_Clust/G60 and G_Clust/G40, G_Clust/G60 ($F(1,38) = 1422.921$, $p < 0.001$, $\eta^2 = 0.974$)[3].

For both, B_Clust and G_Clust, the ratio G50 takes a middle position (see Figure 8). More interesting though is the fact that in case of the third group (BG_Clust) the pattern changes and now ratio G50 yields the highest number of correctly classified pattern. We specified a contrast that compared B_Clust and G_Clust together against BG_Clust, which turned out to be not significant ($F(1,38) = 0.061$, $p = 0.806$, $\eta^2 = 0.002$). This indicates that the overall number of correctly classified patterns is statistically indistinguishable. However, the specification of a contrast for the interaction effect between B_Clust and G_Clust against BG_Clust on the one hand, and G40 and G60 against G50 on the other hand reveals a highly significant effect ($F(1,38) = 138.739$, $p < 0.001$, $\eta^2 = 0.785$). This indicates once more that the factor ratio had significantly different effects on the number of correctly classified patterns in different pattern categories.

To look deeper into the classification results we created confusion matrices that not only show the number of correctly classified spatial patterns (that we used in the ANOVAs) but additionally reveal in detail how a misclassified pattern was actually classified. For example, if a

---

[2] We used the Huynh-Feldt correction again as Mauchly's test is significant here, too.

[3] All three way interaction effect including the expert non-expert (ex_nex) distinction were not significant. This indicates that in this comparison experts and non experts performed comparably.

spatial pattern that shows significant clustering of only blue areas (B_Clust) was misclassified, it could have been classified as any of the four remaining categories (G_Clust, BG_Clust, Disp, Rand). Figure 8 shows eight confusion matrices separated for experts and non-experts. Besides a summary across all spatial patterns, we also provide information separated by the color ratio as this factor turned out to be a dominant criterion for the classification. The descriptive analysis of these matrices confirms and extends the results we obtained thus far and can be summarized as follows: We do not find large differences between the experts and non-experts. If one color is dominating (blue or green) participants perceive this pattern as significantly clustered with respect to the dominating color. Hence, in addition to correctly classified B_Clust and G_Clust spatial patterns, we also find that in the case of blue being the dominant color that B_Clust received the largest number of misclassifications (first column) and in the case of green being the dominant color that G_Clust receives the largest number of misclassifications (second column). This is, again, true for both experts and non-experts. We also find that the classification behavior is nearly identical (reversed regarding to the dominating color) for B_Clust and G_Clust, indicating that no one directional figure ground effects were present. In other words, neither blue nor green was perceived as only as figure or only as ground. Random patterns are basically randomly classified with an interesting bias in the expert group toward being interpreted as dispersed (Disp).

Experts Green 40 / Blue 60

	B_Clust	G_Clust	BG_Clust	Disp	Rand
B_Clust	96	0	16	5	3
G_Clust	66	0	41	4	9
BG_Clust	77	1	31	4	7
Disp	18	1	4	62	35
Rand	70	1	11	17	21

Non-Experts Green 40 / Blue 60

	B_Clust	G_Clust	BG_Clust	Disp	Rand
B_Clust	87	0	22	8	3
G_Clust	70	0	40	5	5
BG_Clust	88	0	26	5	1
Disp	36	0	12	37	35
Rand	83	0	18	10	9

Experts Green 50 / Blue 50

	B_Clust	G_Clust	BG_Clust	Disp	Rand
B_Clust	32	5	50	12	21
G_Clust	3	26	54	11	26
BG_Clust	5	15	68	10	22
Disp	0	2	6	78	34
Rand	0	1	14	68	37

Non-Experts Green 50 / Blue 50

	B_Clust	G_Clust	BG_Clust	Disp	Rand
B_Clust	38	2	55	11	14
G_Clust	1	33	57	15	14
BG_Clust	2	14	82	8	14
Disp	1	0	11	65	43
Rand	1	5	27	56	31

Experts Green 60 / Blue 50

	B_Clust	G_Clust	BG_Clust	Disp	Rand
B_Clust	4	68	25	8	15
G_Clust	1	94	11	5	9
BG_Clust	0	88	23	1	8
Disp	0	23	2	66	29
Rand	0	54	1	35	30

Non-Experts Green 60 / Blue 40

	B_Clust	G_Clust	BG_Clust	Disp	Rand
B_Clust	3	82	22	5	8
G_Clust	1	100	10	3	6
BG_Clust	1	92	22	2	3
Disp	0	35	5	49	31
Rand	1	60	6	22	31

Experts total

	B_Clust	G_Clust	BG_Clust	Disp	Rand
B_Clust	132	73	91	25	39
G_Clust	70	120	106	20	44
BG_Clust	82	104	122	15	37
Disp	18	12	26	206	98
Rand	70	56	26	120	88

Non-Experts total

	B_Clust	G_Clust	BG_Clust	Disp	Rand
B_Clust	128	84	99	24	25
G_Clust	72	133	107	23	25
BG_Clust	91	106	130	15	18
Disp	37	35	28	151	109
Rand	85	65	51	88	71

Figure 9. Confusion matrices. The (mis-) classification results for all pattern categories (pat_cat) for experts and non-experts differentiated by color ratio are shown. Results for each pattern category are provided in one row providing the number of correct classifications (gray cells) and the misclassifications (white cells) for the remaining four pattern categories.

We performed a couple of additional analysis we briefly summarize here. The classification behavior was relatively straight forward. Participants in the expert group reclassified patterns (placing an icon from one group into a different group) 7.35 times on average; non-experts reclassified patterns 5.25 times on average.

We also looked into the difference of defining adjacency either as rook or as queen. While the experimental design required to select one over the other we calculated spatial autocorrelation measures for queen contiguity, too. The values are not identical but a correlation

analysis confirmed that the significance assessment has the same tendency. All correlations (performed on the z scores of each pattern for three different join-types: G-G, B-B, G-B) are highly significant ($p < 0.001$) with correlation coefficients larger than 0.872.

We looked into the log files of the experiments that contain information which icons were chosen first to be placed into the different categories. While a detailed analysis of this sequential data is beyond the scope of this article we did find that the icons that were chosen first seem to have a higher level of significance. Hence, while the significance level did not turn out to be a criterion for the classification behavior, it may be a criterion for chosen prototypical or salient members of a category. This was the case for both experts and non experts.

Participants filled out a questionnaire at the end of the experiment asking them about their expertise on the topic of spatial analysis and particularly join count analysis. The results confirm that knowledge about spatial autocorrelation and particularly join count analysis does not lead to a better performance, that is, more correctly classified patterns. The four questions that addressed participants spatial analysis knowledge and that they rated on a 1 (I completely agree) to 5 (I do not agree) Likert scale were: 1) I know about Tobler's first law of geography; 2) I am familiar with spatial autocorrelation; 3) I am familiar with spatial analysis; 4) I am familiar with join count statistics. We performed individual t-test on these questions comparing experts and non-experts. All results are highly significant ($p < 0.001$), meaning that our expert assessed themselves as experts and the non-experts did not. An additional correlation analysis comparing the four questions against the five pattern categories confirmed this picture: Two significant correlations were obtained for the dispersed pattern (the one that experts performed significantly better than non-experts). All other correlations did not yield significant values and

were smaller than 0.23. As the Likert scale provides ranking information, we confirmed the results with both parametric and non-parametric correlation coefficients.

Discussion

The results shed light on factors that influence the interpretation of spatial patterns, on the difficulty that students (and most likely humans at large) have on understanding the concept of statistical significance when it comes to interpreting spatial patterns, and that this difficulty exists even if students had several geography courses including spatial analysis in which they learned about spatial autocorrelation measures.

Factors influencing the perception and interpretation of spatial patterns

While the factor ratio was explicitly included in the design of the spatial patterns used as stimuli, we were surprised by the clarity with which it surfaced. Both blue and green worked in the same way, that is, when they were the dominant color and are significantly clustered they were fostering, one might say *inducing*, correct classification. Additionally, as the confusion matrices reveal, a dominant color also induced misclassification. In contrast, if blue or green were significantly clustered (B_Clust / G_Clust) but were the color was not dominant, the opposite effect is the result: the non-dominant color, even if it is the significantly clustered one, is significantly less often correctly classified (or not correctly classified at all). This effect of ratio is complemented by the fact that in case both, blue and green are significantly clustered, the ratio 50/50 (G50) is the one that is perceptually in line with this fact and hence, is the one that yields the highest number of correctly classified (and also misclassified) patterns. Additionally,

that this effect surfaces for both green and blue shows that the Color Brewer has provided good advice on color selection for nominal data, that is, both colors are equally salient perceptually.

The effect of significance levels did not surface consistently in the analysis of the classification behavior of the participants. While we find sometimes tendencies to correctly classify patterns if they are more significant, the general pattern is not consistent. The main reason for this might be the design of the experiment. While it is in general desirable to randomize aspects in the design of experiments, fixed significance values might have served the purpose of the experiment better (compared to randomly choosing from significant ranges). However, we felt it to be important to create the patterns randomly and not artificially design patterns. One solution to this problem would be to perform a different analysis that would build on the fact that we could use the z-scores of each spatial pattern directly (rather than the three significant ranges, that is, factors). These techniques are discussed under different names such as *multilevel analysis* or *linear mixed models* (Field 2009). We did not follow this path in this article as the other analyses (the descriptive analysis shown in Figures 4 to 6, the ANOVAs, and the confusion matrices) all led us to identify the color ratio as the dominating effect. The only aspect where the significance levels seem to have played a role was in the order in which spatial patterns were selected. This may be related to the complexity that correlates to spatial autocorrelation measures (Olson 1975).

Differences with respect to pattern categories

Once we account for the effect of the different ratios, the three significantly-clustered patterns (B_Clust, G_Clust, and BG_Clust) are not significantly different from each other. From our experience in the classroom we know, however, that students have difficulties with the

concept of one color being clustered (significantly) while the other is not. This result is therefore somewhat unexpected as we would have hypothesized that spatial patterns with only one color being statistically significantly clustered would be harder to classify correctly compared to patterns where both green and blue are clustered at the same time. The one pattern that stood out in being correctly classified is the dispersed pattern. The experts performed especially well in this category and the classical chessboard pattern must have stuck in their memory. In hindsight, it is probably the easiest pattern to identify as it is clearly different from the others. Unsurprisingly, the random patterns were the hardest to classify. Somewhat disappointingly, the experts did not outperform the non-experts. Participants' performance for correctly identifying random patterns is just above chance.

Free classification experiment

We previously mentioned some details about the free classification experiment. This experiment was conducted to complement the analysis and to shed light on the natural grouping (categorization / classification) behavior that participants would show when given the task to create as many groups as they see fit for the spatial pattern icons. This methodological approach is well documented in the psychological literature and is essential to elicit conceptual knowledge (Ahn and Medin 1992; Cooke 1999). In other words, how are the spatial patterns meaningfully placed into groups? This experiment was identical to the previous one, with the exception that participants were allowed to sort the icons into any number of groups, rather than the five groups that we specified in the first experiment.

Participants

Participants were recruited from the same IST course as the non-expert group. We had 20 participants, 8 female, average age 19.95. Students received extra credit for their participation.

Results

The results show a surprisingly similar pattern compared to the other two experiments in the sense that the color ratio, that is, the ratio of green and blue areas, is the dominating categorization criterion. To show this result we depict a dendrogram as the results of a cluster analysis using Ward's method in Figure 10. After an initial analysis it became obvious that ratio was again the main criterion so we changed the labels to highlight this aspect. The three ratios are abbreviated as G60, G50, and G40. Choosing a three cluster solution a clustering structure surfaces that is based on the ratio (with only minimal exceptions), independent on the level of significance and independent even on the kind of spatial pattern (only blue, only green, blue and green, dispersed, or random). We also visualized these results by applying multidimensional scaling (MDS) to the grouping data that the participants created. The results are shown in Figure 11. Ratio as the dominant grouping criterion surfaces here, too. In the upper right part of the figure we find patterns dominated by green regions, furthest away (lower left part), we find spatial patterns dominated by blue regions, and in the middle we find the spatial patterns with the 50/50 ratio.

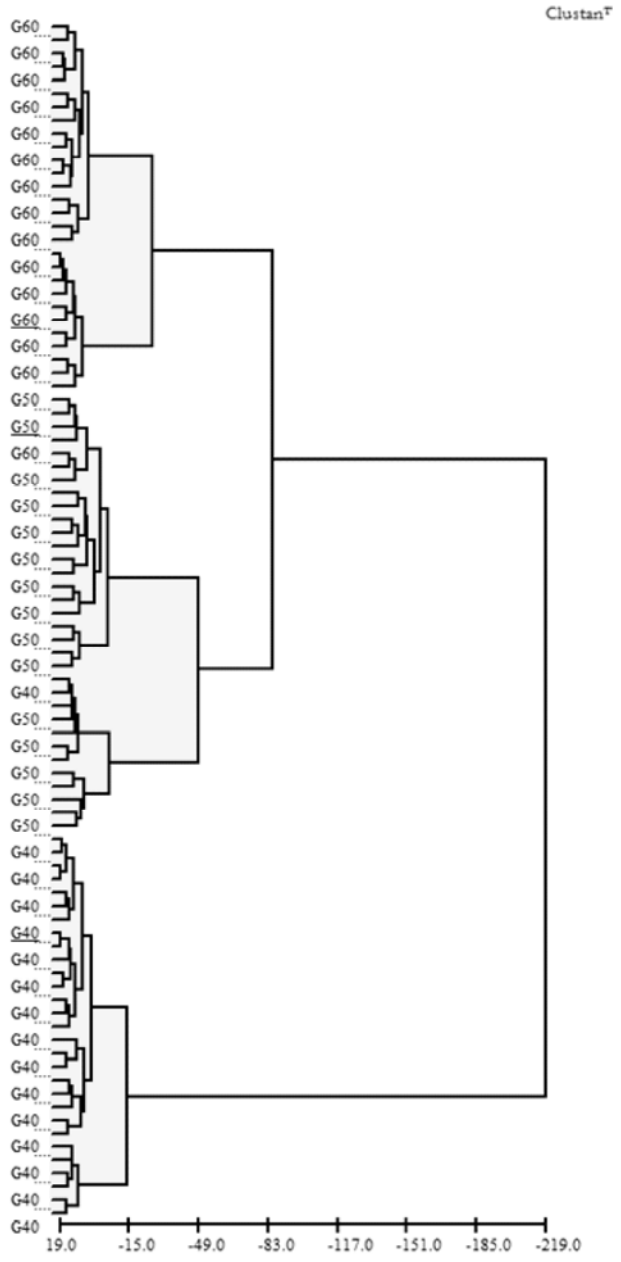

Figure 10. Cluster analysis (Ward's method, CLUSTAN™) showing a three cluster solution. For readability purposes and after analyzing the clustering structure, we replaced file names with new file names that only indicate the ratio (G40: 60/40; G50: 50/50; G60: 40/60). It is obvious that in the free classification task the ratio of the colors has been used as the overall criterion to categorize the spatial patterns.

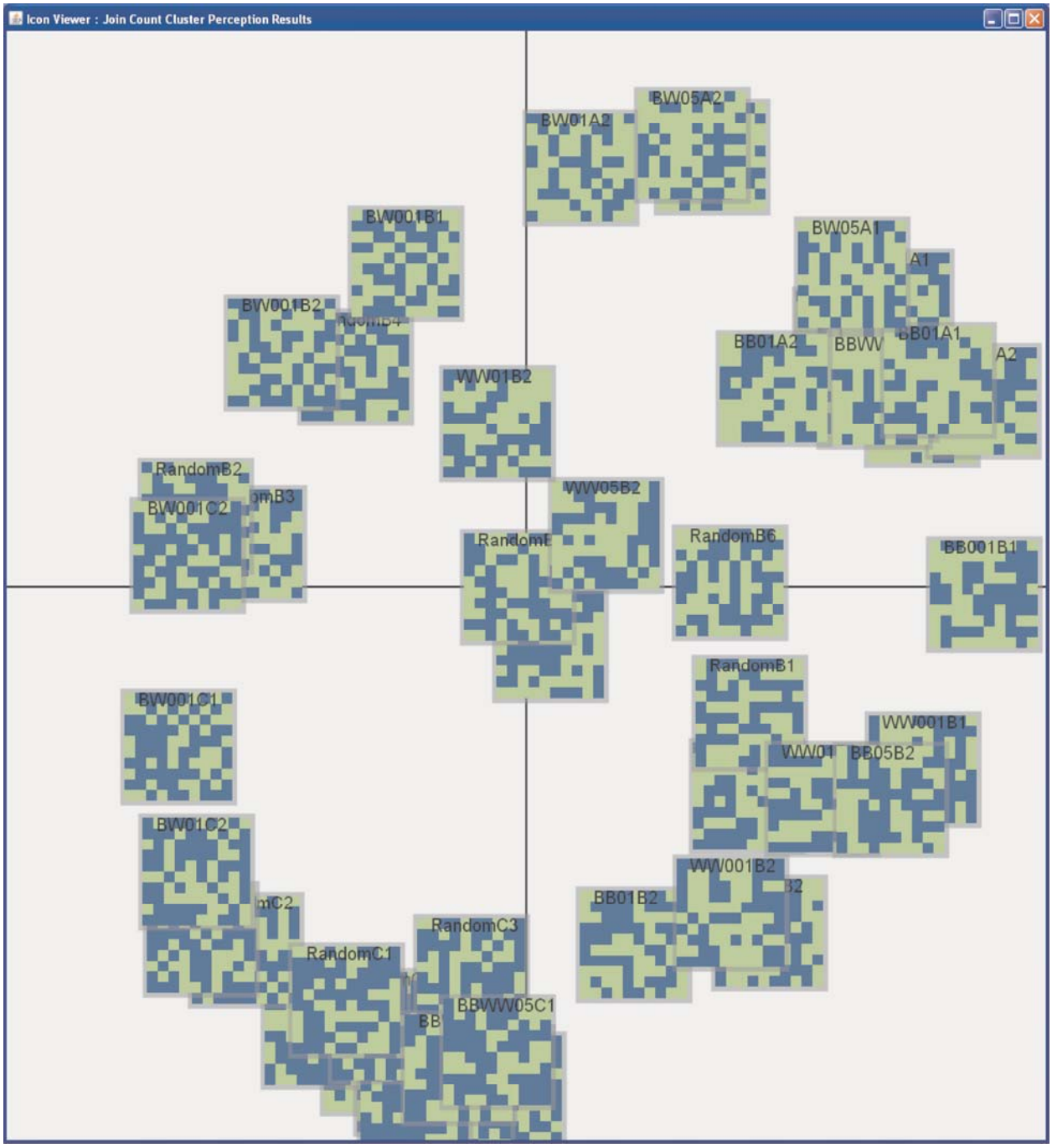

Figure 11. MDS plot of the grouping behavior of participants in the free classification task. The figure shows clearly that the participants in the free classification task grouped primarily on the basis of the ratios, that is, the dominant color.

Conclusions

The analysis and understanding of spatial patterns is essential to all subfields of geography. Spatial patterns come into existence as the manifestation of spatial processes and can be visualized using maps. To approach spatial patterns from the perspective of quantitative analysis, spatial analysts have established a framework that allows for grounding the characterization of spatial patterns on the mathematically established principle of randomness. Randomness is a pre-requisite to establish statistical significance, that is, comparing observed patterns to a set of patterns that could be the result of a random process (complete spatial randomness / independent random process, O'Sullivan and Unwin 2003). Unsurprisingly, randomness is a key concept in many sciences and plays a vital role in areas as diverse as quantum mechanics as well as social systems analysis (Nickerson 2002). Differentiating random and non-random events is important for both scientific and non-scientific affairs as several matters of consequence depend on it (Bar-Hillel and Wagenaar 1991). While the problems of understanding randomness (for students as well as for humans at large) is a research topic in many disciplines, there are barely any geographically tailored assessments from an educational and / or cognitive perspective. This is astonishing, as the difficulty of understanding random spatial processes is pronounced in case visual representations are involved: humans tend to see patterns in maps and readily interpret patterns ignoring the possibility that they might be the outcome of a chance process.

Toward an explanation

The results of our study can potentially be explained from two perspectives: from the standpoint of perception, we can relate our findings to theories of how the information represented in our maps is perceived. In Treisman's (Treisman and Gelade 1980) terminology aspects that can be differentiated are referred to as dimensions (here: color, texture, and shape). Color as well as shape are preattentively processed (see Ware 2004 for an introduction) and are critical in theories of object recognition as well as information visualization. The features (particular values on a dimension according to Treisman and Gelade 1980) are easy to identify for the dimension color, that is, blue and green; they are more difficult to assess for the dimension shape. A general assessment could actually come from the level of statistical significance that did not surface in the classification results but might have played a role in choosing the most salient spatial patterns as first patterns to be placed into a category. While our work did not ask for the assessment of complexity of spatial patterns, this interpretation would be in line with findings by Olsen (1975) who found a partial positive correlation between spatial autocorrelation measures and the self assessed complexity of a map (she used the same ten by ten abstract grid cell maps we employed in our experiments).

However, our results are not only perceptual and completely in the realm of preattentive or perceptual processing. Our participants performed an analysis task and made conscious choices. The perceptual characteristics of our spatial patterns are not comparable to hard-wired perceptual effects that are most clearly demonstrated by visual illusions. Figure 12 shows two illusions (Muller-Lyer and Ponzo) to demonstrate this aspect: Both parts of the Figure demonstrate effects that our perceptual system creates and that cannot be eliminated by conscious, cognitive processes. We cannot 'see' that the two lines (Figure 12A – Muller-Lyer

illusion) have the same length and we cannot 'see' that both lines in Figure 12B (Ponzo illusion) are the same size.

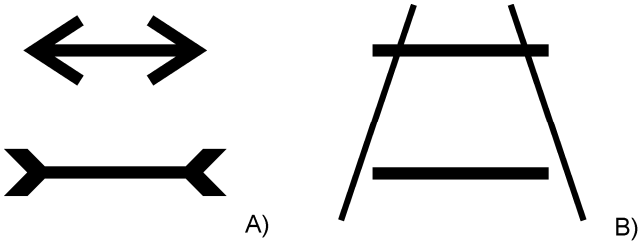

Figure 12. Two visual illusions: A) Muller-Lyer illusion; B) Ponzo illusion.

In contrast, it is possible, to interpret the spatial patterns we used in our experiments in the correct way. Maybe not all of them as there are some borderline cases but after designing these experiments and teaching spatial analysis courses for several years, the authors feel competent (and tested themselves) that it is possible to correctly identify most pattern. The reasons for misinterpreting and misclassifying the spatial patterns therefore cannot completely lie in the realm of perception, we must look in the realm of conscious cognitive processing as well. The misinterpretation then has to come from assumptions that are made about spatial patterns that are incorrect. These wrongly held assumptions are common and are being approached in a series of research experiments, addressing both statistics and probabilities as well as visualizations. For example, recent research has demonstrated that people favor 3D representations over 2D representations but that performance using 3D representations is actually worse (Hegarty, Smallman, and Stull 2009). There are also long held beliefs about probabilities such as the goat problem (Randow 2001), which, as one example, demonstrates the difficulties to overcome common (folk) believes when it comes to decision making and probabilities. In our case one of the most likely explanations is that the significance of a spatial pattern is intimately linked with the number of cells of a particular color. This aspect could be associated with an

assumption about non-independence, that is, a cell has to be either blue or green and if there are more blue cells than green cells it follows that blue (as it could potentially have more joins) has to be the significantly clustered pattern. The same logic would apply to green being the dominant color or both colors being at equilibrium and therefore perceived as both being significantly clustered. What is overlooked in this scenario is that it is possible to change (to some extent) the number of joins independently of the number of blue and green areas.

Hence, while there are certainly perceptual aspects that play a role in identifying spatial patterns, our results led us to the conclusion that it they are more likely the results of a false belief, that is, that the number of areas of a particular color is directly linked to the statistical significance of a spatial pattern. This interpretation is also consistent with current cognitive and neuro-science approaches that see the brain as an organ that not simply awaits to be activated by external stimuli but that processes sensory information based on the memory of past experiences to make decisions (Kveraga, Ghuman, and Bar 2007). In Egenhofer and Mark's often cited article on *Naïve Geography* (Egenhofer and Mark 1995), the topic of randomness and of the tendency to see patterns is not discussed. We consider it essential though to increase the research on the difficulties that students have with the concept of spatial randomness and its manifestation in map patterns.

Implications for education

The results of our study show the difficulties that participants, whether our group of trained experts or non-experts, have relating a spatial pattern that they observe to the concept of randomness and statistical significance. This fact is not new and many textbooks provide hands on examples that use this effect for classroom exercises. For example, Rogerson (Rogerson

2001) uses a drawing exercise which demonstrates to students that if asked to create a random point pattern, they will fail in most cases: Instead of a random pattern that shows both dispersion and clustering students tend to create dispersed patterns by placing dots homogenously on the sheet of paper (they fill the empty spots). This example is a valuable exercise and from the feedback of students it is an exercise with educational value. The results of our experiments, however, stress two important aspects in how to educate students: First, it is essential to supplement map interpretation with a statistical analysis. This is necessary, because human perception is prone to 'detect' patterns that are not necessarily there (i.e., are potentially results of chance processes). Recent approaches in visual analytics have noticed this (e.g., Thomas and Cook 2005; Kraak 2008) and provide a complex set of tools that allows for the detection of spatial clusters and clustering. Second, the inability to detect significantly clustered patterns and the factors that influence pattern detection can be used to educate students. It would be a valuable exercise for students to deepen their understanding of randomization and statistical significance through interactive tools. These tools should focus on teaching students the basics of random processes and statistical significance and only later on focus on real world examples. This will be a difficult task for geography students who seem to prefer real world examples. It is, however, questionable, how much of the underlying statistical concept students learn if they apply them only in software solutions such as ArcMap or GeoDA. It is a challenge, but we think it is a necessary one, to tackle to design educational software for spatial analysis that focuses on the theoretical concepts rather than on real world applications.

Statistical significance

In the age of computers it has become possible to approach statistical significance from different perspectives. The classical approach is to find a mathematically determined distribution related to the problem at hand and to derive, on theoretical grounds, values for the statistical significance of, for example, the number of joins in a study area. The alternative is to use computer simulations that allow for estimating the probability density function and to determine the location of the given values within this distribution. The classical approach has the advantages of being better grounded in theory and computational efficiency. The simulation approach has the advantage of being more resistant to the violation of statistical assumptions and simplicity. Clearer guidance on these choices and their application would be advantageous especially from an educational perspective.

Design sapient interfaces that warn against over-interpretation.

A picture is (sometimes) worth a thousand words (Larkin and Simon 1987; Mayer and Gallini 1990). It has been long noted, however, that visual representations have also severe drawbacks as they are facing challenges with different kinds of information. For example, to visually represent that something is not there (Habel 1998), that something is underspecified or vague such as a location that is not known (MacEachren 1991), or that a phenomenon is multidimensional in nature (Miyake and Shah 2005; Nelson 2007; Chen, Härdle, and Unwin 2008), are aspects that are difficult to represent in graphic displays.

From the results of our research, we can add another insight into the interpretation of map-like representations. In our case, it is the dominance of the number of areas that have a specific color that clearly influence the interpretation of the pattern at hand. This leads to some

interesting considerations. The first one addresses the question of how we can improve the veridicality of interpreting visual stimuli. For example, when people draw a sketch map to indicate how to get from A to B the characteristics of the drawing "warn" the sketch interpreter to not expect an exact one-to-one mapping. However, if a computer would draw a sketch map using straight lines and proper symbols, human sketch interpreters have a tendency to be overly confident in the interpretation of the spatial information(Agrawala and Stolte 2001; Zanola, Fabrikant, and Çöltekin 2009). It is therefore necessary, to a) more fully understand the human perceptual-conceptual systems and b) to develop means in graphic interfaces that compensate (warn) for human "errors". In the case of spatial patterns we could, based on our results, easily program an indicator that provides advice on how much to trust your own visual analysis: the higher the percentage of areas of a specific kind, the more likely this pattern is perceived as being significantly clustered (even if it is not). It might additionally be possible to offer a visual representation that standardizes areal extend. This would be comparable to the standard cartographic advice to not map raw counts as a way to avoid misleading interpretations (e.g., Slocum 2009).

One question that needs to be answered to make such an approach applicable to real world scenarios that has not been covered in our experiments is whether the effect of ratio we observe is related to the number of areas that bear a certain color or whether it is the overall amount of color (in terms of areal extent) that is present in the map. Most real world maps will have a wide variety of differently sized areas. While the area itself will not influence the statistical analysis (if we keep the number of joins identical, as in the case here) it certainly will influence the interpretation (Battersby and Montello 2009). Hence, there is a need for follow up studies to clarify the question of different ratios, that is, whether the number of areas is the

dominant effect or the areal extent. Our results show clearly, though, that a deeper understanding of factors that influence the interpretation of map patterns and the associated processes requires more studies on how visual representations, formal and statistical models, and the human sense-making process of both interact.

Acknowledgements

We would like to thank the reviewers for wonderful comments that helped to shape the line of argument in this paper. This material is based upon work supported by the National Science Foundation under Grant No. 0924534.

References

Abler, R., J. S. Adams, and P. Gould. 1971. *Spatial organization. The geographer's view of the world.* Englewood Cliffs, NJ: Prentice-Hall.

Agrawala, M., and C. Stolte. 2001. Rendering effective route maps: Improving usability through generalization.

Ahn, W.-K., and D. L. Medin. 1992. A two-stage model of category construction. *Cognitive Science* 16:81–121.

Anselin, L. 1995. Local Indicators of Spatial Association - Lisa. *Geographical Analysis* 27 (2):93–115.

Bailey, T. C., and A. C. Gatrell. 1995. *Interactive spatial data analysis.* Harlow, Essex: Longman Scientific & Technical.

Bar-Hillel, M., and W. A. Wagenaar. 1991. The perception of randomness. *Advances in Applied Mathematics* 12:428–54.

Battersby, S., and D. R. Montello. 2009. Area estimation of world regions and the projection of the global-sclae cognitive map. *Annals of the Association of American Geographers* 99 (2):273–91.

Bell, N., N. Schuurman, and M. Hameed. 2008. Are injuries spatially related? Join-count spatial autocorrelation for small-area injury analysis. *Injury Prevention* 14 (6):346–53.

Blok, C., B. Köbben, T. Cheng, and A. A. Kuterema. 1999. Visualization of relationships between spatial patterns in time by cartographic animation. *Cartography and Geographic Information Science* 26 (2):139–51.

Blum, L., M. Blum, and M. Shub. 1986. A Simple Unpredictable Pseudo-Random Number Generator. *SIAM Journal on Computing* 15 (2):364–83.

Borosh, I., and H. Niederreiter. 1983. Optimal multipliers for pseudo-random number generation by the linear congruential method. *Bit* 23 (1):65–74.

Bregt, A. K., and M. C. S. Wopereis. 1990. Comparison of complexity measures for choropleth maps. *The Cartographic Journal* 27:85–91.

Brunet, R. 1980. La composition des modéles dan l'analyse spatiale. (engl. tranlsation (1993). Building models for spatial analysis.). (engl. tranlsation (1993). Building models for spatial analysis.). *L'Espace Géographique* 4:253–65.

Burt, J. E., G. M. Barber, and D. L. Rigby. 2009. *Elementary statistics for geographers*. 3rd Ed. New York: Guilford Press.

Cardinal, R. N., and M. R. F. Aitken. 2006. *ANOVA for the behavioural sciences researcher.* Mahwah, N.J: L. Erlbaum.

Chen, C., W. Härdle, and A. Unwin, ed. 2008. *Handbook of data visualization.* Berlin: Springer.

Cliff, A. D., and J. K. Ord. 1973. *Spatial autocorrelation.* London: Pion.

Cliff, A. D., and J. K. Ord. 1981. *Spatial processes. Models [and] applications.* London: Pion.

Cooke, N. 1999. Knowledge elicitation. Applied Cognition. In *Handbook of Applied Cognition*, ed. F. T. Durso. Chichester: John Wiley & Sons.

Cressie, N. 1990. The Origins of Kriging. *Mathematical Geology* 22 (3):239–52.

Dacey, M. F. 1971. Poly: A two dimensional language for a class of polygons. *Pattern Recognition* 3 (2):197–208. http://www.sciencedirect.com/science/article/B6V14-48MXMMH-8/2/13b14e363e020b95890c19b34b85d966.

Egenhofer, M. J., and D. M. Mark. 1995. Naive geography. In *Spatial Information Theory. A Theoretical Basis for GIS. International Conference, COSIT 95, Semmering, Austria, September 21-23, 1995, Proceedings*, ed. A. U. Frank and W. Kuhn, 1–15. Berlin: Springer.

Fabrikant, S. I., D. R. Montello, and D. M. Mark. 2006. The distance-similarity metaphor in region-display spatializations. *IEEE Computer Graphics and Applications*:34–44.

Field, A. P. 2009. *Discovering statistics using SPSS. and sex, drugs and rock'n'roll (3rd Ed).* London: Sage Publ.

Golledge, R. G. 1995. Primitives of spatial knowledge. In *Cognitive aspects of human - computer interaction for geographic information systems*, ed. T. L. Nyerges, D. M. Mark, R. Laurini, and M. J. Egenhofer, 29–44. Dordrecht, Boston, London: Kluwer.

Goodchild, M. F. 1986. *Spatial Autocorrelation.* Norwich: Geo Books.

Gould, P. 1970. Is statistix inferens the geographical name for a wild goose? *Economic Geography* 46:439–48.

———. 1979. Geography 1957-1977. The Augean period. *Annals of the Association of American Geographers* 69:141–51.

Griffin, A. L., A. M. MacEachren, F. Hardisty, E. Steiner, and B. Li. 2006. A comparison of animated maps with static small-multiple maps for visually iIdentifying space-time clusters. *Annals of the Association of American Geographers* 96 (4):740–53.

Habel, C. 1998. Piktorielle Repräsentation als unterbestimmte räumliche Modelle. *Kognitionswissenschaft* 7 (2):58–67.

Harrower, M., and C. A. Brewer. 2003. ColorBrewer.org: An online tool for selecting color schemes for maps. *The Cartographic Journal* 40 (1):27–37.

Hegarty, M., H. S. Smallman, and A. T. C. M. Stull. 2009. Naive cartography. How intuitions about dispaly configuration can hurt performance. *Cartographica* 44:171–86.

Hortensius, P. D., R. D. McLeod, W. Pries, D. M. Miller, and H. C. Card. 1989. Cellular automata-based pseudorandom number generators for built-in self-test. *Ieee Transactions on Computer-Aided Design of Integrated Circuits and Systems* 8 (8):842–59.

Klippel, A., L. Knuf, B. Hommel, and C. Freksa. 2005. Perceptually induced distortions in cognitive maps. In *Spatial Cognition IV. Reasoning, Action, Interaction. International Spatial Cognition 2004, Frauenchiemsee, Germany, October 11-13, 2004, Revised Selected Papers*, ed. C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel, and T. Barkowsky, 204–13. Berlin: Springer.

Klippel, A., and R. Li. 2009. The endpoint hypothesis: A topological-cognitive assessment of geographic scale movement patterns. In *Spatial Information Theory, 9th International Conference, COSIT 2009, Aber Wrac'h, France, September 21-25, 2009 Proceedings*, ed. K. Stewart Hornsby, C. Claramunt, M. Denis, and G. Ligozat, 177–94. Berlin: Springer.

Knauff, M., R. Rauh, and J. Renz. 1997. A cognitive assessment of topological spatial relations: Results from an empirical investigation. In *Spatial information theory: A theoretical basis for GIS*, ed. S. C. Hirtle and A. U. Frank, 193–206. Berlin: Springer.

Kraak, M. J. 2008. From geovisualisation toward geovisual analytics. *The Cartographic Journal* 45 (3):163–64.

Kveraga, K., A. S. Ghuman, and M. Bar. 2007. Top-down predictions in the cognitive brain. *Brain and Cognition* 65 (2):145–68. http://www.sciencedirect.com/science/article/B6WBY-4PW9C0J-2/2/98975e7acd7b3d22ed1fc6b0a1fd9640.

Kwan, M.-P., and G. Ding. 2008. Geo-narrative: Extending geographic information systems for narrative analysis in qualitative and mixed-method reserach. *Professional Geographer* 60 (4):443–65.

Larkin, J., and H. Simon. 1987. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science* 82 (4):207–25.

Lewandowsky, S., D. J. Herrmann, J. T. Behrens, S.-C. Li, L. Pickle, and J. B. Jobe. 1993. Perception of clusters in statistical maps. *Applied Cognitive Psychology* 7:533–51.

Longley, P. A., M. F. Goodchild, D. J. Maguire, and D. W. Rhind. 2006. *Geographi Information Systems and Science*. *2nd Edition*: John Wiley & Sons.

MacEachren, A. M. 1991. Visualizing uncertain information. *Cartographic Perspectives* 13:10–19.

Makoto, M., and N. Takuji. 1998. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.* 8 (1):3–30.

Mark, D. M., and M. J. Egenhofer. 1994. Modeling spatial relations between lines and regions: Combining formal mathematical models and human subject testing. *Cartography and Geographic Information Systems* 21 (3):195–212.

Marsh, M., R. G. Golledge, and S. Battersby. 2007. Geospatial concept understanding and recognition in G6-College Students: A preliminary argument for minimal GIS. *Annals of the Association of American Geographers* 97 (4):696–712.

Mayer, R. E., and J. K. Gallini. 1990. When is an illustration worth ten thousand words? 82 (4):715–26.

Medin, D. L., W. D. Wattenmaker, and S. E. Hampson. 1987. Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology* 19 (2):242–79.

Miller, H. J. 2004. Tobler's first law and spatial analysis. *Annals of the Association of American Geographers* 94 (2):284–89.

Miyake, A., and P. Shah, ed. 2005. *Handbook of Visuospatial Thinking*: Cambridge University Press.

Montello, D. R., S. I. Fabrikant, M. Ruocco, and R. S. Middleton. 2003. Testing the first law of geography on point-display spatializations. In *Spatial Information Theory: Foundations of Geographic Information Science. International Conference, COSIT 2003, Ittingen, Switzerland, September 24-28, 2003, Proceedings*, ed. W. Kuhn, M. Worboys, and S. Timpf, 335–51. Berlin: Springer.

Moran, P. A. P. 1948. The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 10 (2):243–51.

National research council. 2006. *Learning to think spatially.* Washington, DC: National Academies Press.

Nelson, E. S. 2007. The face symbol: Research issued and cartographic potential. *Cartographica* 42 (1):53–64.

Nickerson, R. S. 2002. The production and perception of randomness. *Psychological Review* 109 (2):330–57.

Olson, J. M. 1975. Autocorrelation and visual map complexity. *Annals of the Association of American Geographers* 65 (2):189–204.

Openshaw, S. 1979. Alternative methods of estimating spatial interaction models and their performance in short-term forcasting. In *Exploratory and Explainatory Statistical Analysis of Spatial Data*, ed. C. P. A. Bartels and R. H. Ketellapper. Nijhoff: Leiden.

O'Sullivan, D., and D. Unwin. 2003. *Geographic information analysis.* Hoboken, NJ: Wiley.

Pothos, E. M., and N. Chater. 2002. A simplicity principle in unsupervised human categorization. *Cognitive Science* 26 (3):303–43.

Randow, G. v. 2001. *Das Ziegenproblem. Denken in Wahrscheinlichkeiten.* Reinbek bei Hamburg: Rowohlt.

Rogerson, P. A. 2001. *Statistical methods for geography.* London, Thousand Oaks, New Dehli: SAGE Publications.

Sadahiro, Y. 1997. Cluster perception in the distribution of point objects. *Cartographica* 34 (1):49–61.

Skupin, A., and S. I. Fabrikant. 2007. Spatialization. In *The handbook of geographic information science*, ed. J. Wilson and A. S. Fotheringham, 61–79. Malden, MA: Blackwell.

Slocum, T. A. 2009. *Thematic cartography and geovisualization.* Upper Saddle River, NJ: Pearson Prentice Hall.

Slocum, T. A. 1983. Predicting visual clusters on graduated circle maps. *The American Cartographer* 10 (1):59–72.

Steinke, T. R., and R. E. Lloyd. 1981. Cognitive integration of objective choropleth map attribute information. *Cartographica* 18 (1):13–23.

Sui, D. Z. 2004. Tobler's first law of geography. A big idea for a small world? *Annals of the Association of American Geographers* 94 (2):269–77.

Tabachnick, B. G., and L. S. Fidell. 2007. *Using multivariate statistics.* Boston Mass.: Pearson/Allyn & Bacon.

Thomas, J. J., and K. A. Cook, ed. 2005. *Illuminating the path: Research and Development Agenda for Visual Analytics.* Richland, WA: IEEE Press.

Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46:234–40.

Treisman, A. M., and G. Gelade. 1980. A feature-integration theory of attention. *Cognitive Psychology* 12 (1):97–136. http://www.sciencedirect.com/science/article/B6WCR-4D6RJM2-46/2/b34aa05384b2a7702189c22840489174.

Walter, S. D. 1993. Visual and statistical assessment of spatial clustering in mapped data. *Statistics in Medicine* 12 (12):1275–91. http://dx.doi.org/10.1002/sim.4780121202.

Ware, C. 2004. *Information visualization*. *Perception for design /*. San Francisco, Calif.: Morgan Kaufmann.

Zanola, S., S. I. Fabrikant, and A. Çöltekin. 2009. The effect of realism on the confidence in spatial data quality. In *Proceedings, 24th International Cartography Conference (ICC 2009), Santiago, Chile, Nov. 15-21, 2009.*

Correspondence: Department of Geography, GeoVISTA Center, The Pennsylvania State University, University Park, PA 16802, e-mails: klippel@psu.edu (Klippel), hardisty@psu.edu (Hardisty), rui.li@psu.edu (Li).