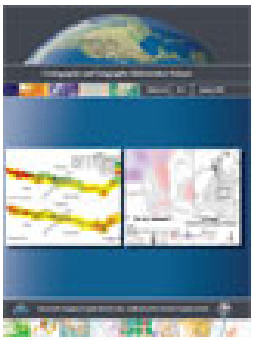

Cartography and Geographic Information Science

ISSN: 1523-0406 (Print) 1545-0465 (Online) Journal homepage: http://www.tandfonline.com/loi/tcag20

Measured and perceived visual complexity: a comparative study among three online map providers

Susan Schnur , Kenan Bektaş & Arzu Çöltekin

To cite this article: Susan Schnur , Kenan Bektaş & Arzu Çöltekin (2017): Measured and perceived visual complexity: a comparative study among three online map providers, Cartography and Geographic Information Science, DOI: 10.1080/15230406.2017.1323676

To link to this article: http://dx.doi.org/10.1080/15230406.2017.1323676

Published online: 06 Jun 2017.

CARTOGRAPHY AND GEOGRAPHIC INFORMATION SCIENCE, 2017
https://doi.org/10.1080/15230406.2017.1323676

ARTICLE

Measured and perceived visual complexity: a comparative study among three online map providers

Susan Schnur [a], Kenan Bektaş [b] and Arzu Çöltekin [b]

[a]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, USA; [b]Department of Geography, University of Zürich, Switzerland

ABSTRACT
We present a study on human perception of map complexity, with the objective of better understanding design decisions that may lead to undesirable levels of complexity in web maps. We compare three complexity metrics to human ratings of complexity obtained through a user survey. Specifically, we use two algorithmic approaches published by others, which measure feature congestion (FC) and subband entropy (SE), as well as our own approach of counting object types rather than individual objects. We compare these metrics with each other as well as with human complexity ratings for three maps of the same area from map providers Google Maps, Bing Maps, and OpenStreetMap. Each map design is assessed at three different scales (levels of detail). We find that (1) the FC and SE metrics appear to be adequate predictors of what humans consider complex; (2) object-type counts are slightly less successful at predicting human-rated complexity, implying that clutter is more important in perceived complexity than diversity of symbology; and (3) generalization choices do impact human complexity ratings. These findings contribute to our understanding of what makes a map complex, with implications for designing maps that are easy to use.

ARTICLE HISTORY
Received 21 October 2016
Accepted 24 April 2017

KEYWORDS
Map complexity; web maps; visual complexity; user perception; map generalization

1. Introduction

As dynamic digital maps begin to dominate our map use experience, we are also seeing an increase in efforts to apply cartographic methods that were originally designed for static maps to these new media (Roth, Donohue, Wallace, Sack, & Buckingham, 2014). For example, the ability to zoom in and out requires changing map scale and level of detail each time the operation is requested. Along with other dynamic features, the expectation of a smooth change during zoom operations introduces challenges, such as automating map generalization (Jones & Ware, 2005). The generalization process has geometric as well as semantic aspects, and it involves many qualitative decisions made by the cartographer that are not easily automated. There are a number of well-established as well as newly proposed approaches, methods, and algorithms for cartographic generalization (e.g. Aliakbarian & Weibel, 2016; Brassel & Weibel, 1988; Cecconi, 2003; Guilbert, Gaffuri, & Jenny, 2014; Harrie & Weibel, 2007; Jiang, Liu, & Jia, 2013; Mackaness & Gould, 2014; Stanislawski, Buttenfield, Bereuter, Savino, & Brewer, 2014; Stoter, Post, Van Altena, Nijhuis, & Bruns, 2014). However, reducing visual complexity while retaining sufficient detail is an ongoing challenge to this day.

As in other areas of design, geovisualization typically operates with the goal of making map products usable. Among the many components of usability, the so-called SEE metrics (Satisfaction, Effectiveness, Efficiency) are widely considered in research concerning human factors (ISO 9241-11) (Abran, Khelifi, Suryn, & Seffah, 2003). The SEE metrics posit that a product, such as a map, should allow the user to effectively complete the task ("accuracy"), to be efficient ("speed"), and should present a satisfying user experience (Duchowski, 2007; Nivala, Brewster, & Sarjakoski, 2008). Based on this line of thinking, selection of map symbols and generalization of map objects are conducted with the end goal of maximizing usability for a given scale (Slocum et al., 2001). Whether a map is usable (based on the SEE metrics) typically depends on three components: its *visualization design*, i.e. visual properties of the map such as the level of detail, visual complexity, organization of visual elements, *characteristics of its user*, i.e. the intellectual qualities of the viewer such as exposure to map use, expertise in reading and interpreting graphics, spatial ability, and *the task at hand*, e.g. visual search, information decoding, and object cognition (Çöltekin, Lokka & Zahner 2016). Arguably, the design of a map

CONTACT Arzu Çöltekin arzu@geo.uzh.ch

can be more easily controlled than the intellectual qualities of users or the tasks they choose to complete.

We therefore focus on determining the overall *visual complexity* of a map design. It is important to note that something that is visually "simple" may still be considered complex because complexity has many dimensions, some of which are much more difficult to define and measure, such as intellectual complexity. However, if visual complexity can be successfully assessed, it may be a first step in determining whether and how much design decisions contribute to the usability of a map. For example, measuring the visual complexity of a map before and after generalization procedures may give cartographers a tool to understand the potential impact of their design decisions on the user. However, only a few studies have attempted to quantify visual map complexity (e.g. Harrie & Stigmar, 2010; Harrie, Stigmar, & Djordjevic, 2015; MacEachren, 1982) or measured map complexity as perceived by map users (e.g. Castner & Eastman, 1984, 1985; Monmonier, 1974; Olson, 1975). To this day, little is known about whether and how well human-perceived visual complexity matches algorithmically measured visual complexity.

Even though most metrics for measuring the visual complexity of a display use principles derived from perception and cognition (e.g. Itti, Koch, & Niebur, 1998; Rosenholtz & Nakano, 2007), and are therefore intended to replicate principles of human vision, these metrics are rarely validated through formal user studies (e.g. Dumont, Touya, & Duchêne, 2016), especially with maps (e.g. Ciolkosz-Styk & Styk, 2013; Fairbairn, 2006). In addition to this, very few of these studies use realistic (and published) cartographic output but rather use simplified maps generated specifically for the purposes of analysis (e.g. Fairbairn, 2006). We have taken a step toward filling these gaps. Specifically, we have developed a new method for measuring the visual complexity of maps and paired this with two existing algorithmic approaches in order to obtain values for three types of *measured complexity* on three different map designs, each at three different scales. We then deployed an online survey to establish user perception of map complexity (*perceived complexity*) on all studied map displays. Realistic published maps were selected from three popular web map providers: Google Maps (GM), Microsoft Bing Maps (BM), and OpenStreetMap (OSM). Once all measures were obtained, we compared measured and perceived complexity in order to understand how well the different measures validate each other. Our results provide insights into how some aspects of map design, including factors that can and cannot be controlled by the designer, may contribute to increased map complexity and therefore hurt user performance with maps.

2. Map complexity

The term complexity appears to be understood in a myriad of ways by different researchers working in different domains (Schnur, Bektaş, Salahi, & Çöltekin, 2010). In fact, MacEachren (1982) suggests that the study of map complexity has been hindered by the lack of a consensus on the definition of the term itself. Despite this difficulty, Cartography and Geographic Information Science (GIScience) researchers have studied map complexity from various perspectives. Among these studies, we observe two categories of map complexity: visual (sometimes also called graphic) and intellectual (Castner & Eastman, 1985; Fairbairn, 2006; MacEachren, 1982). The notion of *visual complexity* often refers to the content of the display and builds upon what we (visually) perceive, and how we process visualized information. Perception and processing of the visual information are often directly related to properties of the graphic content of the geographic visualization, such as the spatial distribution of the objects, their relative positioning, and other visual variables (Bertin, 1983). On the other hand, *intellectual complexity* is linked to cognition and semantics, i.e. the way we comprehend, understand, and attach meaning to what we see (Brophy, 1980; Fairbairn, 2006; MacEachren, 1982). A conceptually similar distinction has been proposed for visual attention in general. Both "bottom-up" and "top-down" processes take place in the human visual system, where the first is driven by perceptual mechanisms, and the latter by cognitive mechanisms (Ware, 2010).

Arguably, measuring intellectual complexity is much more difficult than measuring visual complexity as the former varies radically for every user (e.g. Brophy, 1980). Such individual and group differences have been confirmed by various user studies (e.g. Çöltekin et al., 2016; Çöltekin, Fabrikant, & Lacayo, 2010; Deeb, Ooms, & Maeyer, 2012). However, distinguishing visual and intellectual complexity is often not straightforward. For example, Brophy (1980) states that "a stable picture of a map is built up only when the map percipient is fully aware of both the visual and intellectual components of the map" (Brophy (1980), as cited in Fairbairn (2006, p. 225). Similarly, Robinson (1952) disregards the distinction between these two types of complexities, claiming that the mind (especially the untrained eye) does not ordinarily distinguish between the responses to intellectual and visual stimuli. Tufte (1989) also recognizes these individual differences and notes that every person has a "preferred" level of complexity in visualizations. This preference may depend on the viewer's artistic taste, their prior knowledge on the topic, or their visual processing abilities and cognitive resources at a given moment. All of these factors make it difficult to isolate the effects of visual complexity on human perception of geovisualizations, especially within the limits of user studies where we can reasonably measure only a few factors at a time.

In many early studies of map complexity, the focus was on graphical complexity through measures of vertices, edges, and faces based on mathematical analyses of object geometries (e.g. Dietzel, 1983; MacEachren, 1982; Muller, 1976). In more modern studies, attention-based entropy measures (e.g. Rosenholtz & Nakano, 2007), object counts (e.g. Fairbairn, 2006; Harrie & Stigmar, 2010), and sophisticated image processing methods (e.g. Ciolkosz-Styk & Styk, 2013) have been proposed, which are potentially appropriate also for dynamic maps and other geographic visualizations such as aerial or satellite images or three-dimensional (3D) models.

While various theoretical approaches to measure map complexity have been proposed in cartography, empirical approaches appear to be rare (Schnur et al., 2010). In an early study, Phillips and Noyes (1982) investigated the effects of symbology clutter on map reading performance. Harrie and Stigmar (2010) have shown that the quantitative measures *number of objects, number of points,* and *object line length* had better correspondence with human judgment than *object area*. However, in this study only one object type (buildings) was evaluated, and only two people assessed the perceived map complexity (the authors themselves). The authors expanded on this later in a follow-up study of map readability in which single and composite measures of readability were contrasted with results from a user study involving 214 people (Harrie et al., 2015). These initial results, focusing specifically on cartographic complexity, encourage further work. Outside cartography, visual complexity has been extensively studied in vision research and perceptual psychology, especially in relation to visual search (e.g. Neider & Zelinsky, 2011; Rosenholtz, Yuanzhen, & Nakano, 2007; Wolfe, 1998). Importantly, Alvarez and Cavanagh (2004) linked visual complexity with information load, measuring the efficiency of search for a target among distracters from the same category, and observed a trade-off between the complexity of objects and the total number of objects that can be stored in (human) memory.

The definition and measurement of the complexity of visual forms and visual arrays has also been considered in Gestalt psychology (e.g. Brady, Konkle, & Alvarez, 2011; Donderi, 2006; Forsythe, 2009). Among these, Forsythe's (2009) study bears conceptual similarities to ours; in which the author assessed various image complexity measures for natural scenes. Forsythe's (2009) findings have shown a strong link between visual complexity and *familiarity* (Schnur et al., 2010). Familiarity is one of many dimensions involved in understanding visual complexity. Another dimension is design-related factors, such as *quantity of objects, clutter,* and *variety of colors* (Oliva, Mack, Shrestha, & Peeper, 2004). Several other (interdisciplinary) studies focused on human factors such as cognitive and emotional aspects involved in processing websites of varying visual complexity (e.g. Harper, Michailidou, & Stevens, 2009; Tuch, Bargas-Avila, Opwis, & Wilhelm, 2009). However, evaluations conducted on other types of media might not necessarily apply to cartographic displays. Cartographic displays are highly abstract in comparison to natural images and require a different (spatial and visual) cognitive capacity. There are only a few previous studies that link cartography and the cited measures of visual complexity. For example, Schnur et al. (2010) offer a small-scale preliminary user study, Brychtová, Çöltekin, and Pászto (2016) provide an early benchmarking effort with different levels of generalization, Dumont et al. (2016) investigate the effects of map zooming and therefore clutter changes in measures of visual complexity, and Touya, Decherf, Lalanne, and Dumont (2015) assess the clutter in generalized maps using various clutter measures. Similarly, Ciolkosz-Styk and Styk (2013) have tested image processing methods to assess graphical complexity of maps.

In this study, we expand on previous work by empirically examining quantitative as well as perceptual aspects of visual complexity for cartographic displays. More specifically, we measured the visual complexity of three digital web maps (each designed by different map providers) at three levels of detail (LODs) using three quantitative approaches and asked a large number of participants (n = 130) to subjectively rate the complexity of the studied maps through an online experiment. We then compared and contrasted the findings of these two types of complexity measures.

3. Methods

An overview flowchart of our procedure (described below) is shown in Figure 1.

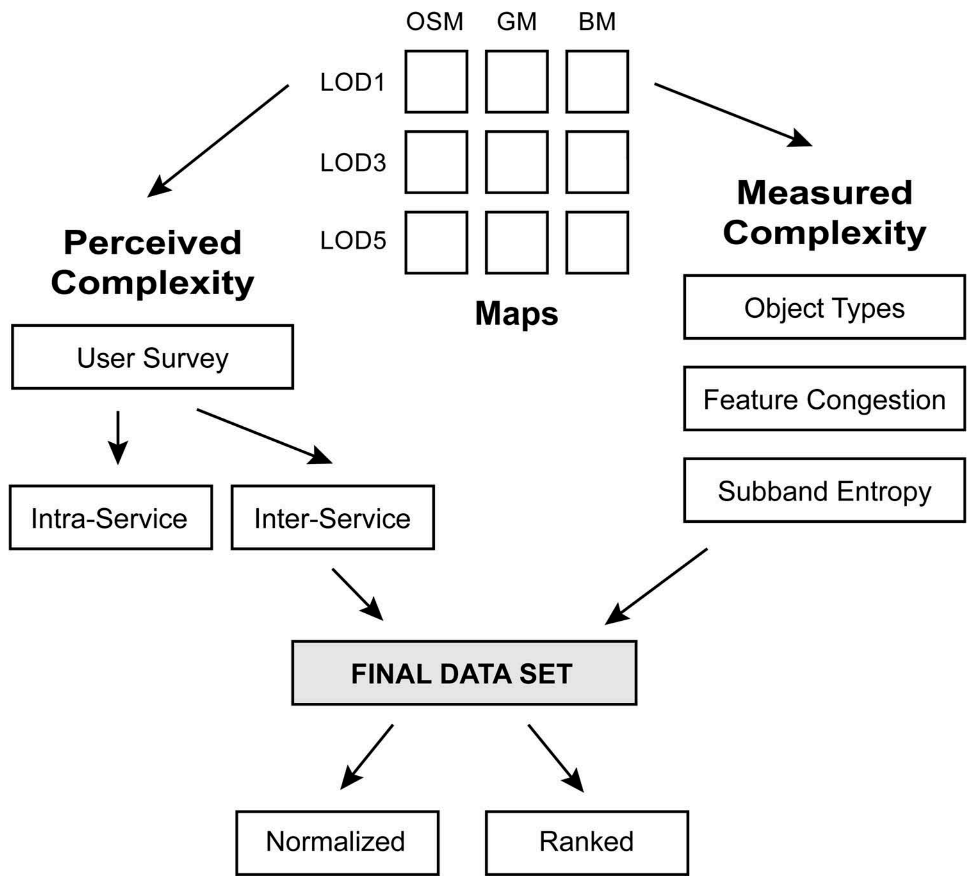

Figure 1. Flowchart showing different parts of our procedure to compare measured and perceived complexity. Nine maps are used as input stimuli, from which perceived complexity measured via a user survey as well as three other quantitative complexity measures are derived. These values are then combined into a final data set and analyzed by means of normalization and ranking.

3.1. Stimuli/materials

For our online experiment, we used actual web maps with a realistic level of heterogeneity in symbolization and object types, increasing our study's applicability to real-life situations. Our work therefore enriches previous studies that used only simplified and idealized stimuli. Specifically, we used a total of *nine maps* from three popular online map providers: GM, BM, and OSM. Because these are dynamic maps, we selected maps at three scales from each provider, ranging from approximately 1:190,000 to about 1:6,500 (Figure 2). Multiple LODs were necessary in order to study the effect of changing map content and symbolization on measured and perceived complexity (Cecconi & Weibel, 2000). These specific scales were selected in order to provide three different views of the city of Zürich: a focused neighborhood view, an intermediate city view, and a regional view. During initial pilot testing, five LODs were selected for each provider but later levels 2 and 4 were discarded so that participants would more easily recognize the differences between consecutive scales. This is why you will see LODs marked as 1, 3, and 5 throughout the paper (see Figure 2). All maps were centered on the main train station of the city of Zürich, Switzerland, and the screenshots were cropped to a size of 500 × 350 pixels to ensure that people who use laptops with smaller screens or tablets could also participate in the study. The maps were presented in a static manner – viewers were not able to dynamically switch between maps at different scales.

3.2. Methods for quantifying measured complexity

We quantified map complexity in two ways: (1) counting *object types* and (2) running two algorithms for measuring visual *clutter*. The first approach is based on the assumption that perceived complexity is potentially affected not by the total number of individual objects viewed, but by the number of categories of objects (i.e. different map symbols, or object types). This assumption regarding object types is based on the findings of others; for example, even if an object has multiple features, the cognitive system is reported to treat that object as a single, combined entity (e.g. Luck & Vogel, 1997; see Scholl, 2001, for a review).

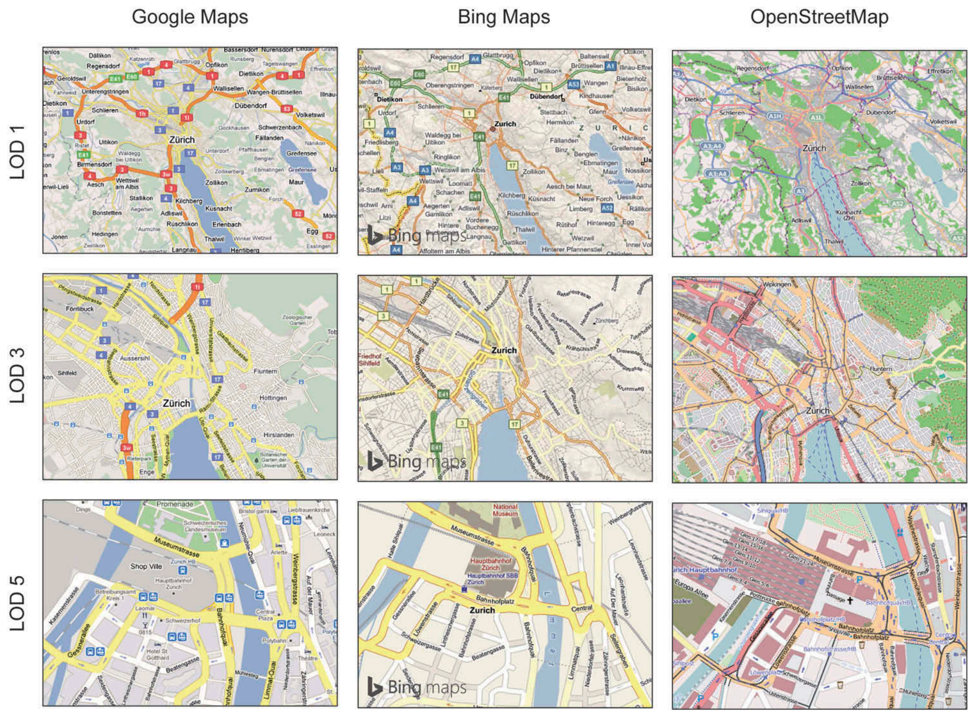

Figure 2. The nine maps used in this study, representing three LODs and three web map providers. Images shown here are scaled down from the full 500 × 350 pixel size used in the online survey. LOD5 corresponds to the most zoomed-in map, whereas LOD1 refers to the most zoomed-out map. Actual map scales are approximately 1:190,000, 1:50,000, and 1:6,500 for LODs 1, 3, and 5, respectively. These maps date from 2010 and no longer resemble the maps currently displayed online. Although this makes re-examining the map designs difficult, it does not influence the validity of this study. Our results focus on user perception of each particular design and are not intended to judge the quality of the design choices made by each map provider. Bing maps reproduced with permission of Microsoft corporation. Google map data © Google. OpenStreetMap (openstreetmap.org) maps © OpenStreetMap contributors; data available under the Open Database Licence.

Furthermore, the capacity of the visual working memory is limited to seven plus or minus two (Miller, 1956), or in more conservative measures, three objects at a time (e.g. Cowan, 2001; Ware, 2010). Therefore, we expect that increasing the number of distinct map symbols, rather than the number of instances the symbol is used, will present an increased load on human working memory, increasing the perceived complexity of the map.

To obtain a measure of the number of distinct object types, we manually counted the symbols present on each of the nine maps (Table 1). In this study, we consider a *distinct object type* (similar to Schnur et al., 2010) as unique based on its defining visual variables (specifically, *shape, size*, and *color* were considered in this study). In order to validate the reproducibility of object counting, an individual who was not directly involved in the project was asked to redo the object counts, given only the information available in this paper. The individual was able to successfully reproduce our initial counts. Images of distinct object types for each map at scale 5 are shown in Figure 3 to further clarify what is meant by *distinct object types*. The figure shows that all symbols that are unique in terms of shape, size, and color are counted separately. Therefore, labels with the same font size but different colors are counted separately, but multiple instances of the same label style, differing only in terms of the actual text, are counted as a single distinct object type. Similarly, roads with different thicknesses and colors are counted separately, as are different water colors and different land use types, such as parks as opposed to buildings. Note that we do not consider orientation as part of a distinct object type. That is, a

Table 1. Distinct object counts for GM, BM, and OSM maps at three LODs.

	LOD1			LOD3			LOD5		
	GM	BM	OSM	GM	BM	OSM	GM	BM	OSM
Labels	6	10	6	8	8	5	14	7	12
Hydrology	2	2	1	1	1	1	1	2	1
Roads	4	6	4	4	6	8	4	5	8
Buildings	4	2	3	6	3	8	4	3	4
Transport	0	0	2	1	1	3	1	1	4
Other	1	1	2	0	0	0	1	2	1
Total	17	21	18	20	19	25	25	20	30

Note: Counts are broken down into separate categories, although further analyses are conducted only on the sum total.

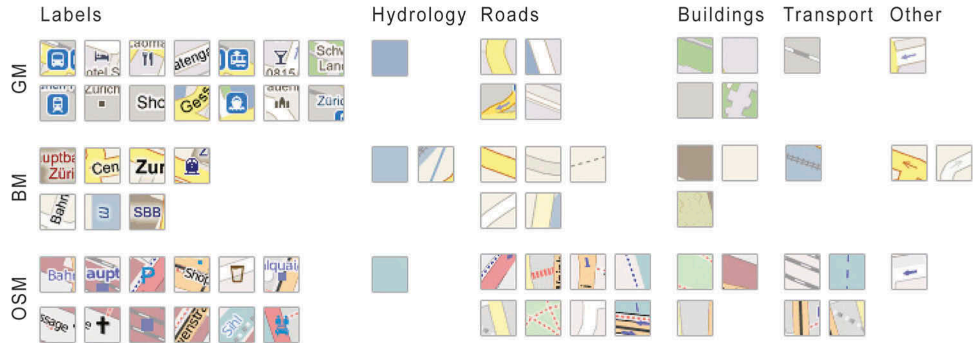

Figure 3. Distinct object types for maps at LOD5, which generally represents the scale with the greatest number of object types. Map snippets from images in Figure 2.

curved label and a straight label are counted as one object type given that they are otherwise identical. This leads to a count of the total number of distinct object types the eye must process when viewing the map. This approach has rarely been used in the literature, and we believe the implementation we describe here is unique.

For validation and further testing, we employed an additional method of measuring visual complexity. In this study, we chose to focus on *clutter*, which is defined by Rosenholtz et al. (2007) as "the state in which excess items, or their representation or organization, lead to a degradation of performance at some task" (Rosenholtz et al., 2007, p. 3). In order to measure clutter for complex images that do not necessarily have distinct objects, Rosenholtz et al. (2007) introduced a statistical measure of object saliency, which they report correlates well with performance in visual search tasks (the more salient the object, the faster the user performs).

In this paper, we measure clutter for all nine of our maps using two algorithms developed by Rosenholtz et al. (2007). Rosenholtz et al.'s (2007) feature congestion (FC) algorithm maps all the objects in an image into 3D attribute space using the variables *color*, *orientation*, and *luminance contrast*, where the term luminance refers to the brightness of an image. The output of the algorithm is the volume of the covariance ellipsoid encompassing the mapped points, relative to the maximum possible feature space volume. The FC therefore quantifies the spread of colors, orientations, and luminance variations present in the image, which represents the visual congestion of the image. This gives a sense of how easy it would be to place a new object into the image and still have it stand out to the viewer from amongst the other objects already present (Rosenholtz et al., 2007). Rosenholtz et al. (2007) also suggest a subband entropy (SE) algorithm, which seeks to represent the organization of objects in a scene using image encoding efficiency as a proxy. A standard RGB image is decoded into luminance and chrominance subbands (CIELab), with luminance representing the brightness component of the image and chrominance representing the color component of the image. The subbands represent wavelet decompositions of varying spatial frequencies in these brightness and color image components. The number of bits required to encode each subband is measured and summed to calculate the SE, based on the assumption that more bits are required to represent a higher-fidelity wavelet transform, reflecting the inherent redundancy in the image. This measure therefore attempts to capture whether similar objects are spatially grouped together or repeated in an image. The MATLAB code for these algorithms is freely available[1] and was not altered in any way for this study.

3.3. Methods for quantifying perceived complexity

To assess perceived complexity, we conducted an online *user survey*. We consider an online survey most appropriate for this study as it allows us to reach a relatively large and diverse group of participants. Also, since we are testing online maps, an online experiment may be most similar to how these maps are actually used. We designed the survey to focus on purely visual complexity and minimize the contribution of intellectual complexity. Therefore, we avoided assigning specific tasks and asked participants to simply consider the visual complexity of what they see.

3.3.1. Participants

A total of 130 participants responded to the survey (49 females, 81 males, average age 32.7). Five participants (3.8%) reported having imperfect color vision and were therefore excluded from the analysis. Participants were from 22 different countries and about half (48.4%) of them reported being familiar with Zürich. 43.0% of them had academic training or were professionally familiar with at least one of seven related disciplines (Geography, Geomatics, Cartography, GIS, Remote Sensing, Computer Graphics, Graphics Design, Fine Arts). 98.4% of the participants had experience in using graphics or spatial data of some kind, or online maps and virtual globes. 26.9% of the participants were familiar with OSM, 12.3% with BM, and 93.8% with GM. 82.3% of the participants reported using web maps frequently (i.e. daily or weekly). They used web maps to find an address (93%), to plan trips (81.5%), to measure distances or areas (63.8%), and for fun (33.8%).

3.3.2. Survey procedure

The survey was advertised in our personal as well as professional circles. Following standard experimental protocols, we first introduced the study, briefly explained what was expected of the participants, assured confidentiality, and stated the estimated total time to complete the tasks. Following this, the four sections of the experiment were presented. In the *first section*, we documented the participants' demographic characteristics (e.g. age, gender), their level of field-specific expertise and familiarity with relevant components in the experiment (e.g. with the study area, digital maps). The *second and third sections* comprised the main part of the survey. For each question, three maps were presented and the participants were asked to rate the visual complexity of the maps using a five-point Likert scale (Figure 4). Participants were instructed to use their own understanding of visual complexity when rating, independent of any specific task. The *fourth section* consisted of a set of black and white and blurred maps for a preliminary investigation of the contribution of color and object shape in perceived complexity.

In total, each participant viewed 24 maps, of which some were shown twice in order to confirm whether participants were being consistent with their complexity ratings. That is, the nine main maps were presented as part of *inter-service* (three different map designs, same scale) and *intra-service* (three different scales, same map design) comparison groups. The inter-service questions allowed the participants to compare differences in complexity *between* the map providers, while the intra-service comparison allowed them to assess the difference in complexity between different scales *within* the same map design. Both the order of map providers and LODs were rotated to avoid possible bias from order effects (e.g. learning, assuming importance, judging the first/last ones differently, fatigue). Participants were not told which map provider produced which map. If the participant gave the same map a different complexity rating the second time it was presented, we interpreted that their perceived complexity was *relative*, rather than *objective*. That is, their rating would likely change depending on what other maps they were viewing at the time.

The fourth section of the survey consisted of two tasks that were designed to explore the relative effects of color, labels, and underlying shapes on perceived complexity. This is important as these factors appear to have a large contribution to the overall visual complexity of a visualization (Moacdieh & Sarter, 2015; Purchase, Freeman, & Hamer, 2012). We prepared grayscale versions of the LOD5 maps used in section two (inter-service) and asked the participants to rate their visual complexity in order to compare their ratings with the color versions of the same maps. For the second task in this section, we prepared blurred versions of the LOD5 maps used in section two (also inter-service). We reasoned that introducing blur reduces the dominance of labels while preserving the perception of overall spatial relationships, thereby reducing detail. This could potentially allow the participants to distinguish the map objects or understand structural information based only on shape and color.

After the main sections were completed, participants were presented with two optional questions to further investigate their understanding of the word "complex" (this was necessary because we did not give them a definition, nor a specific task to constrain the meaning of "complex"). The questions were designed to include a list of options (multiple answers were allowed) and a comment box where participants could provide their own responses. Following this, participants rated how important they considered color, spacing between objects and labels, and number of labels for

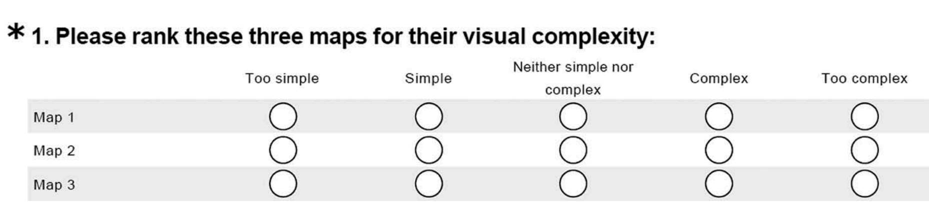

Figure 4. The Likert scale used for complexity ratings by survey participants. *Too simple* is quantified as (1) and *too complex* as (5).

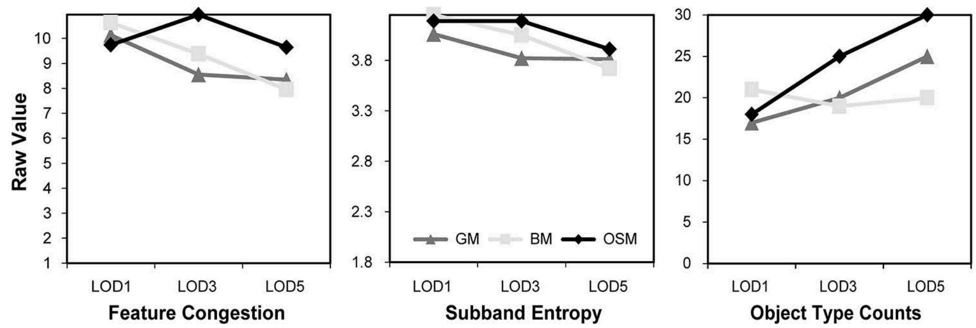

Figure 5. Three measures of complexity, feature congestion (FC), subband entropy (SE), and object-type counts, for three different map designs, assessed at three LODs. Vertical axes show raw values measured for these three complexity metrics. Y-axis intervals are selected based on assumed maximum and minimum values. The selection of these normalization boundaries is discussed in the text.

understanding the content of the maps they had just viewed. Each of these features has been proposed as a source of complexity in the visual psychology and cartography literature (Brychtová & Çöltekin, 2016a, 2016b; Fairbairn, 2006; Neider & Zelinsky, 2011). The experiment ended with a box for open comments.

4. Results

4.1. Measured complexity

Table 2 and Figure 5 show the object type, FC, and SE values determined from each map. OSM presents the most complex maps in most cases (six times out of nine combinations of LOD and metric). GM and BM tie for least complex (both four times out of nine), although at LOD5 BM is consistently the least complex for all three complexity measures. In terms of changes between LODs, three patterns stand out. First, object-type counts tend to increase with LOD, with the exception of BM LOD1. Second, Both FC and SE tend to decrease with LOD, with the exception of OSM LOD3. Thirdly, comparing the object counts with the two clutter measures we see that object type counts tend to agree with either FC or SE (seven out of nine times) when considering complexity rankings (i.e. which provider is least and most complex) across map providers at the same LOD. In fact, object-type rankings exactly matched both FC and SE rankings for LOD5.

Table 2. Quantitative measures of map complexity reported as *raw* values.

	LOD1			LOD3			LOD5		
	Object types	FC	SE	Object types	FC	SE	Object types	FC	SE
OSM	18	9.99	4.15	25	11.20	4.15	30	9.89	3.87
GM	17	10.38	4.02	20	8.80	3.78	25	8.60	3.77
BM	21	10.88	4.21	19	9.63	4.01	20	8.20	3.68

Note: Higher numbers indicate greater complexity. LOD1 represents a zoomed-out state and LOD5 represents a zoomed-in state.

4.2. Perceived complexity

The most critical initial step was to confirm that participants were rating complexity in an objective rather than a relative manner. To accomplish this, we compared ratings between the two times a participant had seen the same map design but in a different comparison group (inter- vs. intra-service comparisons). We used ±1 rating point as a measure of similarity (Figure 6). That is, where the rating difference was ≤ 1, the scores were considered "similar." Agreement

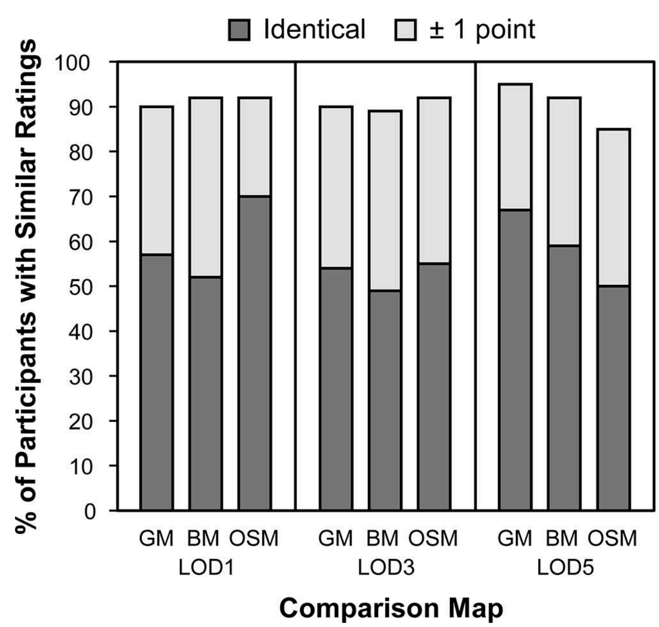

Figure 6. Consistency of complexity ratings when comparing inter- and intra-service ratings of the same map viewed in a different comparison group. Bars show percentage of participants who gave a similar or identical rating to the same map.

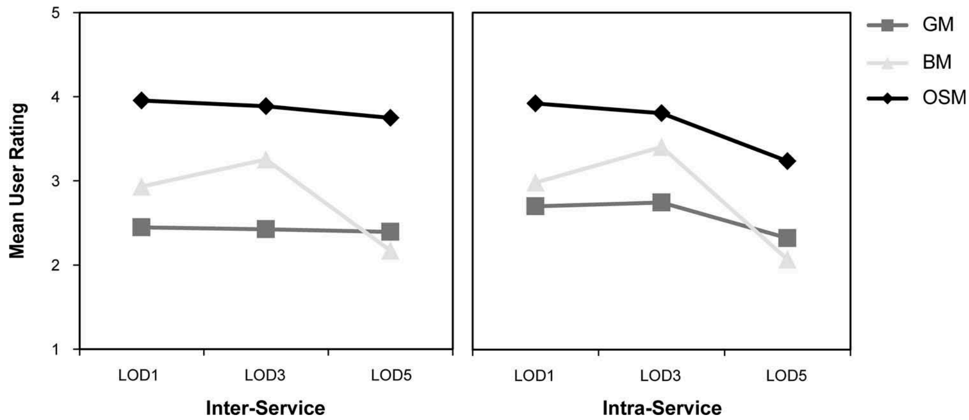

Figure 7. Survey participant complexity ratings, in both an inter-service (left) and intra-service (right) comparison. Values represent the mean of all survey responses. Ratings range from 1 (least complex) to 5 (most complex). Error bars (±1σ SEM) are in most cases smaller than the symbol size (≤0.1) and are therefore omitted. Please see the relevant text for the values.

between the ratings is fairly high, ranging from 85% to 95%, and is fairly consistent between LODs and map providers, except for a slightly lower (−5%) concordance at LOD5 for OSM. There is no clear reason for this minor discrepancy other than generally lower complexity scores for OSM5 in the intra-service comparison, indicating that participants found this map less complex when compared with OSM maps at other scales. Overall only 5–11% of participants gave a significantly different rating to the same map the second time they saw it. These results suggest that users were indeed mostly rating the *objective* complexity of each map. Consequently, the average user score for each map was very similar between intra- and inter-service comparison (differences ranged from −0.32 to 0.51). In order to clarify the results, we therefore chose to consider only the *inter-service* perceived complexity ratings for further analysis. We did so by taking the mean of all survey responses for each map viewed in the survey and present error in the form of ±1σ SEM.

The mean complexity ratings show distinct patterns when compared between LOD and map provider (Figure 7). OSM is consistently ranked as the most complex for all LODs. GM is usually ranked as the least complex, and BM as intermediate, except for LOD5, where GM and BM have a similar complexity rating. In terms of change between LODs, OSM and GM follow a similar gently decreasing pattern, whereas BM exhibits an unusual increase at LOD3 followed by a steep drop off to LOD5. BM therefore has the greatest percent of absolute change between LODs, with OSM second and GM third, except at LOD5, where GM decreases just slightly more than OSM.

We assessed these observed differences statistically using the Kruskal–Wallis test (Kruskal & Wallis, 1952) for independent samples, which determines if an ordinal variable differs significantly between two or more groups. Figure 8 summarizes these results and the following text reports the statistical parameters H (degrees of freedom), p (probability of significance), and r (effect size). At LOD1, complexity was significantly different among the services ($H(2) = 131.72$, $p < .001$). Pairwise comparisons with adjusted p-values showed that there was no significant difference between GM and BM ($p = 1.0$, $r = 0.24$). However, there was a significant difference between OSM and GM ($p < .001$, $r = -0.53$), and OSM and BM ($p < .001$,

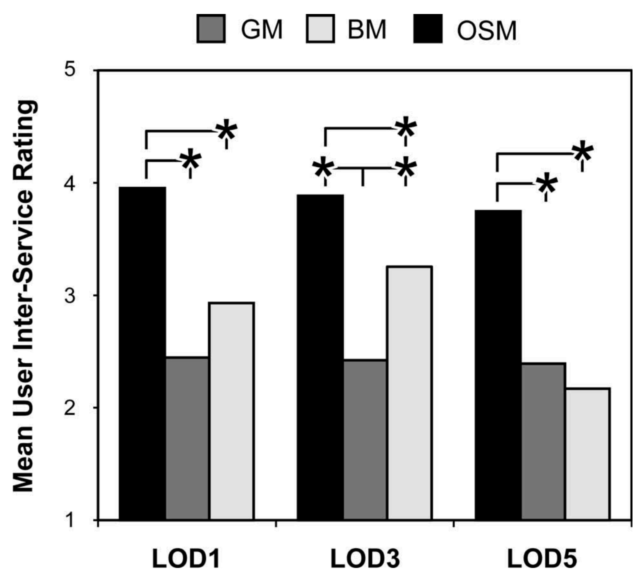

Figure 8. Statistical analysis of inter-service comparison of normalized perceived complexity results for three different LODs. * denotes $p < .001$, indicating that the mean complexity ratings for two map providers are significantly different. Error bars (±1σ SEM) are in most cases smaller than the symbol size (≤0.1) and are therefore omitted. Please see Table 3 for these values.

$r = -0.41$). At LOD3, complexity was significantly different among the web maps ($H(2) = 127.236$, $p < .001$). Pairwise comparisons with adjusted p-values showed that there was a significant difference between GM and BM ($p < .001$, $r = 0.33$), between GM and OSM ($p < .001$, $r = 0.57$), and between BM and OSM ($p < .001$, $r = -0.23$). At LOD5, complexity was significantly different among the web maps ($H(2) = 154.098$, $p < 0.001$). Pairwise comparisons with adjusted p-values showed that there was no significant difference between BM and GM ($p = .227$, $r = 0.09$), but there was a significant difference between BM and OSM ($p < .001$, $r = -0.6$), and between GM and OSM ($p < .001$, $r = -0.5$). These results indicate that most of the rating differences, and therefore rankings, shown in Figure 7 are statistically significant.

The fourth section of the experiment was designed to measure the effects of color and object/symbol distinctness on complexity at LOD5. This was motivated by the results of our initial pilot study (15 participants) where complexity rankings for LOD5 appeared to vary more than for other LODs. The results of this portion of the survey suggest that complexity increases slightly for every service if we show grayscale or blurred maps instead of color maps, while preserving complexity rankings (i.e. OSM remains the most complex). Blur causes a further increase in the complexity ratings for all maps, again preserving complexity rankings. These findings suggest that although decreasing color and distinctness increases map complexity rankings, these factors affect all map providers and LODs equally. They also highlight the fact that color likely plays only a minor role in controlling perceived complexity. We acknowledge that since most map users are unlikely to view blurred maps, participant familiarity may have affected these results, with those viewers familiar with the city of Zürich better able to mentally fill in details obscured by the blurring. In fact, the average complexity ratings of those familiar with the city of Zürich were 0.22 points higher than those unfamiliar with Zürich for the black and white maps, and about 0.30 points lower for the blurred map. This suggests some caution should be used in interpretation of these results.

4.2.1. Participants' opinions on reasons for visual complexity

At the end of the experiment, we asked multiple choice questions to determine if the participants understood more or less the same thing with the words "visually complex," and the responses indicated that this was indeed the case. When prompted about factors that contribute to complexity, 79% of participants responded that high map complexity was a result of an *excessive number of objects and patterns*, 65% indicated *too much text and/or too many labels,* and 48% selected *diversity of colors* as factors in increasing complexity. These responses support the results of the black and white and blurred map rankings. According to these findings, the number of map objects and patterns has a greater influence on map complexity than the number of colors used in the map. Participants also responded to a question about the relative importance of visual elements (color, labels, and spacing of objects and labels) in understanding the map content with nearly equal ratings: 80–82% rated each component as important or very important.

4.3. Comparison of measured and perceived complexity

In order to visually compare the measured and perceived complexity values, it was necessary to normalize each measure to a 1–5 scale, where a value of 1 indicates low complexity and 5 indicates high complexity (Table 3). Distinct object-type counts were normalized to a minimum of zero, and the maximum object count for all of the maps (which is 30 at OSM LOD5). The FC and SE measures were normalized based on the FC and SE measures for a blank sheet as a minimum (1.2483 and 1.7567, respectively) and the maximum measured FC and SE (11.20 for OSM LOD3 and 4.21 for BM LOD1, respectively). The final normalized values for measured complexity are shown in Table 3.

Visually comparing normalized values directly (Figure 9), we see that object counts are most similar to user perception. This metric is the closest to user ratings six out of nine times, of which four cases could

Table 3. Quantitative measures of map complexity *normalized* to a 1–5 scale for easier comparison with mean user ratings of complexity.

	LOD1				LOD3				LOD5			
	OT	FC	SE	User	OT	FC	SE	User	OT	FC	SE	User
GM	2.83	4.59	4.61	2.45 ± 0.06	3.33	3.79	4.11	2.42 ± 0.07	4.17	3.69	4.09	2.39 ± 0.06
BM	3.50	4.84	4.99	2.93 ± 0.08	3.17	4.21	4.60	3.25 ± 0.08	3.33	3.49	3.92	2.17 ± 0.07
OSM	3.00	4.39	4.87	3.95 ± 0.10	4.17	5.00	4.88	3.88 ± 0.09	5.00	4.34	4.32	3.75 ± 0.08

Note: Normalization is based on an assumed maximum and minimum for each method. LOD1 represents a zoomed-out state, whereas LOD5 represents a zoomed-in state. Error in user ratings is reported as ±1σ SEM (n = 130).

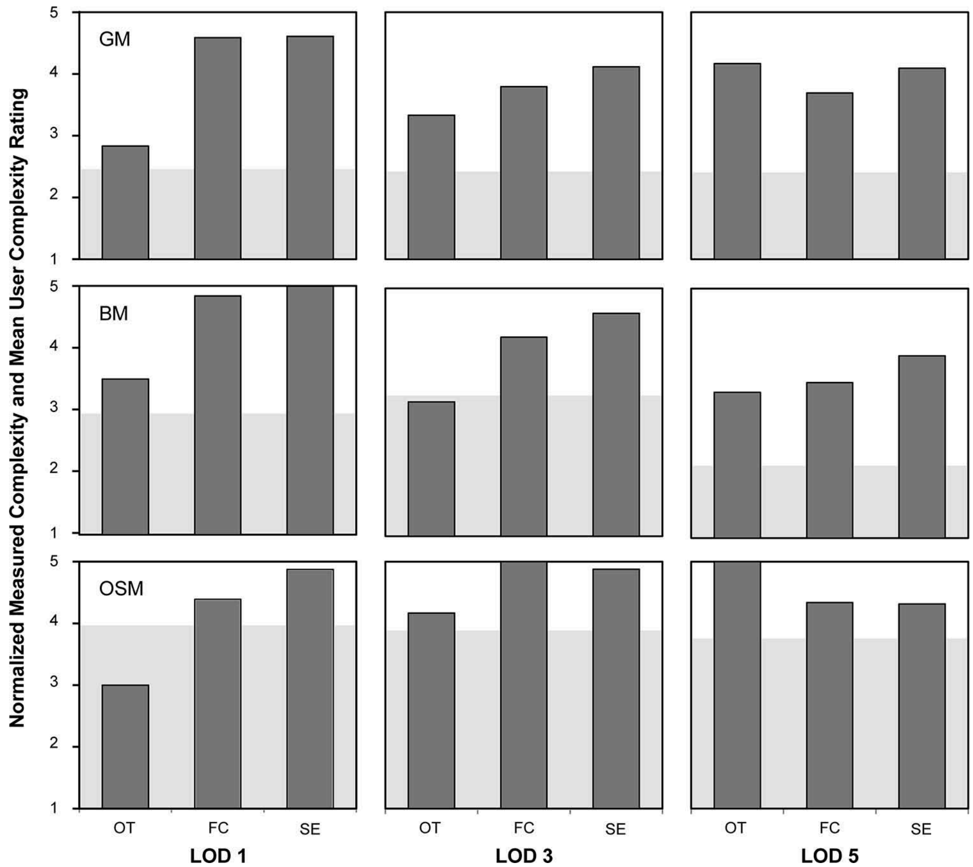

Figure 9. A comparison of normalized measured and perceived complexity values for all nine maps used in this study. On the horizontal axis, OT, FC, and SE stand for object-type counts, feature congestion, and subband entropy, respectively. The gray background in each graph indicates the mean user rating for that map. Error in the user ratings is minimal and is therefore not shown.

be considered a very close match. In three out of the nine maps, it could be argued that none of the metrics closely approximate user ratings. For the two remaining maps (OSM LOD1 and OSM LOD5), FC and SE provide a relatively good approximation of user ratings.

Note however that the two clutter measures provide much higher values of complexity than the object counts. This is likely due to normalization relative to a low minimum clutter. In this sense, comparison of absolute values may be misleading. More information can be gained from looking at *relative* differences between LODs and map providers for each complexity measure. In other words, if one metric reports a map design as most complex and the next metric ranks the same map as most complex, we can reason that the two metrics are in agreement.

Table 4. Relative rankings of different complexity metrics and user perception, where 1 is least complex and 3 is most complex compared to other map providers for a given map.

	LOD1				LOD3				LOD5			
	OT	FC	SE	User	OT	FC	SE	User	OT	FC	SE	User
GM	1	2	1	1	2	1	1	1	2	2	2	2
BM	3	3	3	2	1	2	2	2	1	1	1	1
OSM	2	1	2	3	3	3	3	3	3	3	3	3

When considering rankings within each LOD, we see a clear pattern: object counts, FC, and SE are all in concordance with user ratings for LOD5, and for LOD3 FC and SE (Table 4). The main differences arise in LOD1, where none of the metrics are in complete agreement with each other. In general, the object-type count behaves most differently from perceived complexity, matching in only four of nine maps. In comparison, FC matches user rankings in six out of nine maps and SE in seven out of nine maps.

The clearest pattern observed here is that all three of the complexity metrics seem to struggle at LOD1, regardless of map provider. In fact, the object count metric correctly matches perceived complexity for GM LOD1, whereas FC does not. Considering this, the object count method is only unusual in that it fails at LOD3, whereas FC and SE correctly match perceived complexity at this LOD. However, object counts do not fail completely as they match user perception for OSM LOD3. Depending on the comparison method, none of the metrics are overwhelmingly successful, nor are they a complete failure. On the whole, however, FC and SE present a slightly better match than object-type counts. These results suggest that these measures require further study in order to more precisely identify situations in which they succeed and fail. We must also determine how we can more objectively compare measured and perceived complexity when these two values are measuring the same parameter in such different ways. We therefore conclude that the object count method has potential and warrants more dedicated study.

5. Discussion

We have tackled the issue of map complexity by investigating how it can be measured and whether measured complexity correlates with perceived complexity. We used two algorithmic approaches developed by Rosenholtz et al. (2007), which measure FC and SE, and our own approach, which is based on counting distinct object types. This method was motivated by cognitive design theories, which suggest that the number of symbols displayed on a map may be more representative of perceived complexity than the actual number of objects. We compared these three metrics with each other and with user ratings of complexity for three map designs from online map providers GM, BM, and OSM. Each map design was analyzed at three different scales (thus, three LODs) to control for the effect of changing map content.

The results showed a fairly consistent trend in all tested metrics, suggesting that (1) FC and SE are promising predictors of what humans consider complex; (2) object-type counts, while not demonstrating the same consistency as the FC and SE measures, also show promise in certain cases and should be tested further; (3) map design, as expected, does matter for measured and perceived complexity even when it is exactly the same region represented at exactly the same scale; and (4) scale variation, thus generalization choices, can change relative perceived complexity, even within the same map provider, confirming what cartographers have known for many decades, and thus also confirming the internal validity of the study setup.

5.1. Measured complexity

The distinct object counts do not vary a great deal across map providers and LODs. Object counts vary between 17 (for GM LOD1) and 30 (OSM LOD5), but are generally less than 23, meaning there is a less than six object difference between what can be considered low and high complexity. At this point, it is unclear what can be considered a significant difference in terms of number of object types. If one map has one or two more object types than another, is this enough to contribute to increased complexity? Perhaps there is indeed a tipping point beyond which humans will perceive an increase in complexity. Assuming that ±2 is a valid margin of error, GM and OSM have the same complexity at LOD1, whereas BM is only slightly more complex. At LOD3, GM and BM have the same complexity, whereas OSM is more complex. At LOD5, all three maps have different complexities. In general, distinct object counts are lower for LOD1 (17–21) and higher for LOD5 (20–30), and the variability between maps is also greater at LOD5. Labels, roads, and buildings are the main object types, with hydrology, transport, and other objects playing a minor role. BM is the most complex at LOD1 according to distinct object counts mainly because it includes many labels and roads. OSM is by far the most complex at LOD3 because it shows many roads and buildings, and at LOD5 because it shows many labels and roads, and displays more transport objects. Different map providers tend to show certain classes of objects more at a given scale, such as OSM showing transportation infrastructure when zoomed in, whereas GM and BM do not. Conversely, at LOD5, GM has more labels than any of the other map providers but overall has only intermediate complexity based on distinct object counts. This variability indicates that it is not the addition of a specific class of object (e.g. transport, labels) that contributes to overall complexity, but rather the total number of distinct object types.

Rosenholtz et al.'s (2007) FC and SE both reflect a sense of image clutter. FC and SE were highest for LOD1 and lowest for LOD5, with the greatest variability in values seen in LOD3. Both FC and SE were very similar for LOD1 and LOD5, save for OSM FC at LOD5. Considering the maps, LOD3 represents the zoom level most in need of generalization choices, and therefore a level whose symbolization and clutter varied most between map services. For example, all three maps show minor side streets, but BM symbolizes them with lines, whereas both GM and OSM use polygons. OSM, however, chooses to display many other types of routes, such as hiking paths, adding to both feature and symbolization congestion. This leads to the highest FC measurement of all the maps. LOD5 has generally low levels of congestion as road polygons cover most of the screen and there are large empty spaces representing city blocks. OSM once again generates higher FC and SE values due to the greater number of objects shown and therefore increased clutter in the image. On the whole, the FC and SE images appear closely correlated with generalization choices made by the map designers for each level of detail. A larger follow-up study that includes more maps with more scale variation would be useful to confirm these observations.

5.2. Perceived complexity

We observed distinct patterns in user-rated complexity for the three map providers. At most scales, OSM was consistently rated most complex, which may be explained by its data-driven (rather than design-driven) nature. In other words, it may be because the OSM community includes many open-content enthusiasts and does not typically attract trained designers. In most cases, GM was rated as least complex, which may be because of its successful design, but it may also be because most participants were familiar with GM and familiarity with designs (be it maps, software interfaces, hardware devices, or kitchen layouts) is known to improve performance with the product (Boer, Çöltekin, & Clarke, 2013; Çöltekin et al., 2016; Perrone, 2016; Schnürer, Sieber, & Çöltekin, 2015; Wakeling, Clough, Wyper, & Balmain, 2015). BM is most often at intermediate complexity except at LOD5, where it is ranked least complex. Based on the distinct object counts, this could be because GM shows a large number of different label types at this scale, whereas the BM map is almost devoid of symbols. Additionally, at this scale GM chooses to display bus stops, whereas BM does not, which affects the total number of symbols visible on the screen. We suspect these simple generalization choices may be the main reason for the switch at LOD5. Overall, the perceived complexity findings are not surprising. GM is known for smooth designs targeted at general users, whereas OSM focuses on providing as much information as possible to people who may not otherwise be able to access it, placing less emphasis on the viewing experience.

5.3. User survey

Several characteristics of the user survey may have affected our results. The survey was conducted online, thus we have not controlled for environmental factors such as the characteristics of the display (size, color, resolution) and whether participants have completed the tasks without interacting with others, or without breaks. The characteristics of the population (age, gender, education, etc.) were not counterbalanced. Such factors may have had a direct or indirect influence on our findings. However, since the number of participants in our survey is fairly high (n = 130), such influences are likely addressed by averaging the numbers, and since the user ratings match convincingly well with all three metrics, we are confident that our results are valid.

Another important factor in the survey was that nearly half the participants were familiar with the study area (48.4%). People who were familiar with Zürich rated GM LOD5 ($p < .05$) and OSM LOD5 ($p < .05$) as less complex than the other maps. Although this may have changed the overall rating of complexity (we find places we know less complex in general (Kettunen, Irvankoski, Krause, Sarjakoski, & Sarjakoski, 2012; Li & Klippel, 2016)), it should have had no bearing on the relative ratings of the designs. More importantly, however, considerably more of the participants were familiar with GM (93.8%) and this may be partly the reason why they rated GM as the least complex. While this reasoning should be kept in mind, it is also important to note that it does not apply linearly to the other two designs, where people reported higher familiarity with OSM (26.9%) than BM (12.3%), but consistently rated OSM as most complex. Thus, we believe that familiarity with the map design might have played only a partial role in participant responses.

A clear difference was observed in complexity ratings based on expertise (43% of the participants), which was judged based on participant exposure to GIS and geovisualization topics. Expert participants rated OSM LOD1 less complex than nonexperts. As mentioned earlier, such differences between experts and nonexperts have been documented in the literature (Çöltekin et al., 2010, 2016; Deeb et al., 2012). In the case of this study, we did not design the study to measure differences based on expertise, thus we cannot derive robust conclusions. Nonetheless, the fact that we observe a difference between the expert and nonexpert participants suggests that further investigation on how participant background interacts with perceived complexity is necessary.

Another factor that may have affected the results of this study is the design of the survey. The survey was designed to focus on *perceptual* complexity by asking participants to rate the visual complexity of various map designs directly. One can argue that the choice of scale and the perceived complexity of a map would depend on the task to be completed. However, giving a specific task likely increases the effect of *intellectual* complexity and thus bias introduced by location familiarity and other factors such as the participants' background and spatial abilities would play a much larger role. Since the user ratings agree sufficiently with the algorithmic measures and the questions at the end of the survey revealed that most participants generally have the same concept of what "visually complex" means, asking participants to rate visual complexity without giving them specific tasks did not harm the study. This is supported by the findings of Neider and Zelinsky (2011), who found a relatively robust correlation between clutter estimates based on subjective ratings and those based on manual search performance. Therefore, we believe this direct approach has merit for measuring perceptual complexity.

5.4. Comparing measured and perceived complexity

Our results show that Rosenholtz et al.'s (2007) two measures of image clutter, FC and SE, match perceived complexity slightly better than distinct object-type counts. This appears to indicate that clutter plays a greater role in human perception of web map complexity than symbol variability. To a certain degree, FC and SE also take into account the abundance of colors and luminance values in an image, which will be affected by the number of distinct object types. They also add consideration of object orientation and redundancy (entropy), which are not captured in the distinct object counts. However, we observe that the distinct object types also predict perceived complexity to a certain degree. This indicates that both clutter and symbology contribute to map complexity, but that clutter plays a larger role than symbology. Therefore, the arrangement of objects in the map must be given the greatest weight when measuring complexity.

One possible explanation for the variable quality of fit between measured and perceived complexity is the fact that the maps represent quite similar levels of complexity. Considered across scales, GM shows only a 0.06-point average complexity difference between LODs and OSM only a 0.20-point average complexity difference, whereas BM shows a slightly larger average difference of 1.08 points. When comparing between map providers for the same scale, we see a larger and more consistent difference of about 1.5 points regardless of map provider. The measured complexity metrics were also similar when compared across maps, despite the fact that the maps span a wide range of spatial scales. This may have made it more difficult to tease out differences in the applicability of the different metrics and increased the role of user uncertainty in controlling our results.

This finding can be assessed by considering how changes in scale affected user complexity rankings because changes in scale will affect the number and arrangement of objects visible on the map. As we zoom in to a map, we generally see fewer objects, but there is a greater variety of object types. As we zoom out, we see more objects (roads, labels), but there are fewer object types due to the reduced detail. We see that FC and SE decrease as we zoom in to the map, indicating reduced clutter, whereas object-type counts increase. As we zoom out, FC and SE increase because the image becomes cluttered with roads, not all of which are efficiently generalized or removed from the image (e.g. OSM). The object-type count measure follows the opposite pattern, but does not consider image clutter or object abundance, leading it to predict lower perceived complexity when zoomed out.

The results of the survey do not disqualify the importance of distinct object types in representing complexity, but they do indicate that other factors must be considered as well. Oliva et al. (2004) found that visual complexity results from a combination of factors such as quantity of objects, clutter, and variety of colors. Our findings support this result in that distinct object-type counts were able to match perceived complexity more than half the time in a normalized comparison, but were not as successful as clutter measures in relative ranking comparisons. In the end, however, neither accurately predicted user perception for maps at LOD1. This indicates that quantity of objects, although perhaps not a successful predictor of visual complexity on its own, does play a role in human perception of complexity. It may be that a combination of total object counts, clutter algorithms, and distinct object-type counts is needed to most accurately predict human perception of map complexity.

It may also be that there are other circumstances in which distinct object counts are more successful at predicting human-perceived complexity (and therefore, usability). In this survey, we focused on measuring visual complexity without introducing specific tasks. It may be that a task-based measure of complexity would match the distinct object type measure better. When assessing visual complexity, the viewer may quickly jump from object to object without attempting to store object types in visual memory. Having a task may require them to focus on each object type in order to distinguish it from other object types. This would indicate that instantaneous perception, reflecting visual complexity, focuses more on clutter, whereas more detailed exploration of a map begins to engage the intellectual component of perception and requires that the map viewer interact more directly with the symbology.

6. Conclusions and outlook

The objective of this study was to (1) develop a new method for measuring the complexity of maps and assess this method by comparing it with previously developed algorithmic methods and (2) compare measured complexity with perceived complexity to determine if such quantitative methods can help map designers maximize usability by anticipating end user perception. We observed that algorithmic approaches established by Rosenholtz et al. (2007) can generally predict user perception of map complexity, whereas the newly developed method of counting distinct object types only matches perceived complexity for certain maps. We also found that a direct user rating of visual complexity accurately represented measured complexity. In conclusion, we believe quantitative measures of image complexity can be of use to cartographers in making choices about automatic generalization and map symbolization, with the end goal of increasing the usability of maps in general and web maps in particular.

Note

1. http://dspace.mit.edu/handle/1721.1/37593

Acknowledgments

We are grateful to M. Salahi for his contribution to our original pilot study and to the early drafts of this paper. We would also like to extend our thanks to all of the people who took the time to participate in our survey. This study would not have been possible without them. Finally, we would like to thank D. Heaton for acting as an external tester of the repeatability of our distinct object count method.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Susan Schnur http://orcid.org/0000-0002-8144-9218
Kenan Bektaş http://orcid.org/0000-0003-2937-0542
Arzu Çöltekin http://orcid.org/0000-0002-3178-3509

References

Abran, A., Khelifi, A., Suryn, W., & Seffah, A. (2003). Usability meanings and interpretations in ISO standards. *Software Quality Journal*, 11, 323–336. doi:10.1023/A:1025869312943

Aliakbarian, M., & Weibel, R. (2016, June). *Integration of folksonomies into the process of map generalization*. Paper presented at the 19th AGILE conference on Geographic Information Science, Helsinki, Finland.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, 15(2), 106–111. doi:10.1111/j.0963-7214.2004.01502006.x

Bertin, J. (1983). *Semiology of graphics: Diagrams, networks, maps*. Madison: University of Wisconsin Press.

Boer, A., Çöltekin, A., & Clarke, K. C. (2013, August). Evaluating Web-based geovisualizations online: A case study with abstraction-realism spectrum in focus. In M. F. Buchroithner, (Ed.), *Proceedings of the International Cartographic Conference, ICC 2013*. Dresden, Germany. Retrieved from http://icaci.org/files/documents/ICC_proceedings/ICC2013/_extendedAbstract/284_proceeding.pdf.

Brady, T. F., Konkle, T., & Alvarez, G. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, 11(5), 1–34. doi:10.1167/11.5.4

Brassel, K. E., & Weibel, R. (1988). A review and conceptual framework of automated map generalization. *International Journal of Geographical Information Science*, 2(3), 229–244. doi:10.1080/02693798808927898

Brophy, D. M. (1980). Some reflections on the complexity of maps. In *Technical papers of ACSM 40th Annual Meeting* (pp. 343–352). Falls Church, VA: American Congress on Surveying and Mapping.

Brychtová, A., & Çöltekin, A. (2016a). An empirical user study for measuring the influence of colour distance and font size in map reading using eye tracking. *The Cartographic Journal*, 53(3), 202–212. doi:10.1179/1743277414Y.0000000103

Brychtová, A., & Çöltekin, A. (2016b). The effect of spatial distance on the discriminability of colors in maps. *Cartography and Geographic Information Science*, 44(3), 1–17. doi:10.1080/15230406.2016.1140074

Brychtová, A., Çöltekin, A., & Pászto, V. (2016). Do the visual complexity algorithms match the generalization process in geographical displays?. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (Vols. XLI-B2, pp. 375–378). Göttingen: Copernicus. 10.5194/isprsarchives-XLI-B2-375-2016

Castner, H. W., & Eastman, R. J. (1984). Eye-movement parameters and perceived map complexity—I. *The American Cartographer*, 11(2), 107–117. doi:10.1559/152304084783914768

Castner, H. W., & Eastman, R. J. (1985). Eye-movement parameters and perceived map complexity—II. *The American Cartographer*, 12(1), 29–40. doi:10.1559/152304085783914712

Cecconi, A. (2003). *Integration of cartographic generalization and multi-scale databases for enhanced web mapping*. Doctoral Dissertation, University of Zürich, Switzerland.

Retrieved from http://e-collection.library.ethz.ch/eserv/eth:26494/eth-26494-01.pdf.
Cecconi, A., & Weibel, R. (2000, October). *Map Generalization for On-demand Mapping*. Abstract presented at GIScience 2000. Savannah, GA.
Ciolkosz-Styk, A., & Styk, A. (2013, August). Advanced image processing for maps graphical complexity estimation. In M. F. Buchroithner (Ed.) *Proceedings of the International Cartographic Conference, ICC 2013*. Dresden, Germany.
Çöltekin, A., Brychtová, A., Griffin, A. L., Robinson, A. C., Imhof, M., & Pettit, C. (2016). Perceptual complexity of soil-landscape maps: A user evaluation of color organization in legend designs using eye tracking. *International Journal of Digital Earth*, 1–22. doi:10.1080/17538947.2016.1234007
Çöltekin, A., Fabrikant, S. I., & Lacayo, M. (2010). Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings. *International Journal of Geographical Information Science*, 24(10), 1559–1575. doi:10.1080/13658816.2010.511718
Çöltekin, A., Hempel, J., Brychtova, A., Giannopoulos, I., Stellmach, S., & Dachselt, R. (2016). Gaze and feet as additional input modalities for interacting with geospatial interfaces. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, III-2, 113–120. doi:10.5194/isprsannals-III-2-113-2016
Çöltekin, A, Lokka, I, & Zahner, M. (2016). On the usability and usefulness of 3D (geo)visualizations – a focus on virtual reality environments. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (vol. XLI-BD, pp. 387–392). Göttingen: Copernicus. doi: 10.5194/isprs-archives-XLI-B2-387-2016
Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, 24(1), 87–114. doi:10.1017/S0140525X01003922
Deeb, R., Ooms, K., & Maeyer, P. D. (2012). Typography in the eyes of Bertin, gender and expertise variation. *The Cartographic Journal*, 49(2), 176–185. doi:10.1179/1743277412Y.0000000011
Dietzel, P. P. (1983, September). Measuring complexity on topographical maps. *Proceedings of ACSM-ASP Fall Convention* (pp. 45–49). Salt Lake City.
Donderi, D. C. (2006). An information theory analysis of visual complexity and dissimilarity. *Perception*, 35(6), 823–835. doi:10.1068/p5249
Duchowski, A. (2007). Suggested empirical guidelines. In *Eye tracking methodology* (pp. 171–179). London: Springer.
Dumont, M., Touya, G., & Duchêne, C. (2016, June). *Assessing the variation of visual complexity in multi-scale maps with clutter measures*. Paper presented at the ICA Workshop on Automated generalisation for on-demand mapping, 19th AGILE Conference on Geographic Information Science, Helsinki, Finland.
Fairbairn, D. (2006). Measuring map complexity. *The Cartographic Journal*, 43(3), 224–238. doi:10.1179/000870406X169883
Forsythe, A. (2009). Visual complexity : Is that all there is?. In *Engineering psychology and cognitive ergonomics. EPCE 2009. Lecture Notes in Computer Science* (Vol. 5639, pp. 158–166). Berlin: Springer.
Guilbert, E., Gaffuri, J., & Jenny, B. (2014). Terrain Generalisation. In D. Burghardt, C. Duchêne, & W. Mackaness (Eds.), *Abstracting geographic information in a data rich world: methodologies and applications of map generalisation* (pp. 197–225). Cham, Switzerland: Springer International. doi:10.1007/978-3-319-00203-3_8
Harper, S., Michailidou, E., & Stevens, R. (2009). Toward a definition of visual complexity as an implicit measure of cognitive load. *ACM Transactions on Applied Perception*, 6 (2), 1–18. doi:10.1145/1498700.1498704
Harrie, L., & Stigmar, H. (2010). An evaluation of measures for quantifying map information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(3), 266–274. doi:10.1016/j.isprsjprs.2009.05.004
Harrie, L., Stigmar, H., & Djordjevic, M. (2015). Analytical estimation of map readability. *ISPRS, International Journal of Geo-Information*, 4, 418–446. doi:10.3390/ijgi4020418
Harrie, L., & Weibel, R. (2007). Modelling the overall process of generalisation. In W.A. Mackaness, A. Ruas, & L.T. Sarjakoski, (Eds.), *Cartographic Modelling and Applications*. Oxford: Elsevier. doi:10.1016/B978-008045374-3/50006-5
Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), 1254–1259. doi:10.1109/34.730558
Jiang, B., Liu, X., & Jia, T. (2013). Scaling of geographic space as a universal rule for map generalization. *Annals of the Association of American Geographers*, 103(4), 844–855. doi:10.1080/00045608.2013.765773
Jones, C. B., & Ware, J. M. (2005). Map generalization in the Web age. *International Journal of Geographical Information Science*, 19(8–9), 859–870. doi:10.1080/13658810500161104
Kettunen, P., Irvankoski, K., Krause, C. M., Sarjakoski, T., & Sarjakoski, L. T. (2012). Geospatial images in the acquisition of spatial knowledge for wayfinding. *Journal of Spatial Information Science*, 5(5), 75–106. doi:10.5311/JOSIS.2012.5.85
Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583–621. doi:10.1080/01621459.1952.10483441
Li, R., & Klippel, A. (2016). Wayfinding behaviors in complex buildings: the impact of environmental legibility and familiarity. *Environment and Behavior*, 48(3), 482–510. doi:10.1177/0013916514550243
Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279–281. doi:10.1038/36846
MacEachren, A. M. (1982). Map complexity: comparison and measurement. *The American Cartographer*, 9(1), 31–46. doi:10.1559/152304082783948286
Mackaness, W. A., & Gould, N. M. (2014, September). *The role of geography in automated generalisation*. Paper presented at the ICA workshop on generalisation and multiple representation, Vienna, Austria.
Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing

information. *Psychological Review*, 63(2), 81–97. doi:10.1037/h0043158

Moacdieh, N., & Sarter, N. (2015). Display clutter: A review of definitions and measurement techniques. *Human Factors*, 57(1), 61–100. doi:10.1177/0018720814541145

Monmonier, M. S. (1974). Measures of pattern complexity for choroplethic maps. *The American Cartographer*, 1(2), 159–169. doi:10.1559/152304074784107728

Muller, J. C. (1976). Objective and subjective comparison in choroplethic mapping. *The Cartographic Journal*, 13(2), 156–166. doi:10.1179/caj.1976.13.2.156

Neider, M. B., & Zelinsky, G. J. (2011). Cutting through the clutter : Searching for targets in evolving complex scenes. *Journal of Vision*, 11, 1–16. doi:10.1167/11.14.7

Nivala, A.-M., Brewster, S., & Sarjakoski, T. L. (2008). Usability evaluation of web mapping sites. *The Cartographic Journal*, 45(2), 129–138. doi:10.1179/174327708X305120

Oliva, A., Mack, M. L., Shrestha, M., & Peeper, A. (2004). Identifying the perceptual dimensions of visual complexity of scenes. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th annual cognitive science society* (pp. 1041–1046). Austin, TX: Cognitive Science Society.

Olson, J. M. (1975). Autocorrelation and visual map complexity. *Annals of the Association of American Geographers*, 65(2), 189–204. doi:10.1111/j.1467-8306.1975.tb01030.x

Perrone, P. (2016). *The Effects of Experience and Familiarity on Performance in Virtual Environments*. Senior Thesis, Claremont McKenna College, Claremont, CA. Retrieved from http://scholarship.claremont.edu/cmc_theses/1420/.

Phillips, R. J., & Noyes, L. (1982). An investigation of visual clutter in the topographic base of a geological map. *The Cartographic Journal*, 19(2), 122–132. doi:10.1179/caj.1982.19.2.122

Purchase, H. C., Freeman, E., & Hamer, J. (2012). An exploration of visual complexity. In P. Cox, B. Plimmer, & P. Rodgers (Eds.), *Diagrammatic representation and inference* (pp. 200–213). Berlin: Springer. doi:10.1007/978-3-642-31223-6_22

Robinson, A. H. (1952). *The look of maps: An examination of cartographic design*. Redlands, CA: ESRI Press.

Rosenholtz, R., Yuanzhen, L., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, 7, 1–22. doi:10.1167/7.2.17

Roth, R. E., Donohue, R. G., Wallace, T. R., Sack, C. M., & Buckingham, T. M. A. (2014). A process for keeping pace with evolving web mapping technologies. *Cartographic Perspectives*, 78, 25–52. doi:10.14714/CP78.1273

Schnur, S., Bektaş, K., Salahi, M., & Çöltekin, A. (2010, September). *A comparison of measured and perceived visual complexity for dynamic web maps*. Paper presented at GIScience 2010: sixth International Conference on Geographic Information Science, Zurich, Switzerland.

Schnürer, R., Sieber, R., & Çöltekin, A. (2015). The next generation of atlas user interfaces - a user study with "digital natives." In *Modern Trends in Cartography - Selected papers of CARTOCON 2014* (pp. 26–36). Cham, Switzerland: Springer International. doi:10.1007/978-3-319-07926-4_3

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, 80(1–2), 1–46. Retrieved from doi;10.1016/S0010-0277(00)00152-9

Slocum, T. A., Blok, C., Jiang, B., Koussoulakou, A., Montello, D. R., Fuhrmann, S., & Hedley, N. R. (2001). Cognitive and Usability Issues in Geovisualization. *Cartography and Geographic Information Science*, 28(1), 61–75. doi:10.1559/152304001782173998

Stanislawski, L. V., Buttenfield, B. P., Bereuter, P., Savino, S., & Brewer, C. A. (2014). Generalisation Operators. In D. Burghardt, C. Duchêne, & W. Mackaness (Eds.), *Abstracting geographic information in a data rich world: Methodologies and applications of map generalisation* (pp. 157–195). Cham, Switzerland: Springer International. doi:10.1007/978-3-319-00203-3_6

Stoter, J., Post, M., Van Altena, V., Nijhuis, R., & Bruns, B. (2014). Fully automated generalization of a 1:50k map from 1:10k data. *Cartography and Geographic Information Science*, 41(1), 1–13. doi:10.1080/15230406.2013.824637

Touya, G., Decherf, B., Lalanne, M., & Dumont, M. (2015). Comparing image-based methods for assessing visual clutter in generalized maps. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, II-3/W5*, 227–233. doi:10.5194/isprsannals-II-3-W5-227-2015

Tuch, A. N., Bargas-Avila, J. A., Opwis, K., & Wilhelm, F. H. (2009). Visual complexity of websites: Effects on users' experience, physiology, performance, and memory. *International Journal of Human-Computer Studies*, 67(9), 703–715. doi:10.1016/j.ijhcs.2009.04.002

Tufte, E. R. (1989). *Visual design of the user interface: Information resolution, interaction of design elements, color for the user interface, typography and icons, design quality*. Armonk, NY: IBM Corporation.

Wakeling, S., Clough, P. D., Wyper, J., & Balmain, A. (2015). Graph literacy and business intelligence: Investigating user understanding of dashboard data visualizations. *Business Intelligence Journal*, 20(4), 8–19.

Ware, C. (2010). *Visual thinking for design*. Burlington, MA: Morgan Kaufmann.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, 9(1), 33–39. doi:10.1111/1467-9280.00006