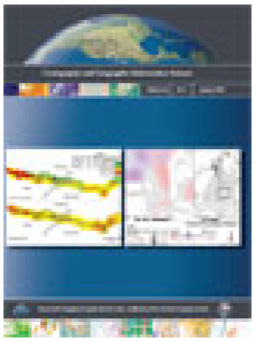

Taylor & Francis
Taylor & Francis Group

Cartography and Geographic Information Science

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

An empirical evaluation of three elevation change symbolization methods along routes in bicycle maps

Annina Brügger, Sara Irina Fabrikant & Arzu Çöltekin

To cite this article: Annina Brügger, Sara Irina Fabrikant & Arzu Çöltekin (2016): An empirical evaluation of three elevation change symbolization methods along routes in bicycle maps, Cartography and Geographic Information Science, DOI: 10.1080/15230406.2016.1193766

To link to this article: http://dx.doi.org/10.1080/15230406.2016.1193766

Published online: 16 Jun 2016.

CARTOGRAPHY AND GEOGRAPHIC INFORMATION SCIENCE, 2016
http://dx.doi.org/10.1080/15230406.2016.1193766

An empirical evaluation of three elevation change symbolization methods along routes in bicycle maps

Annina Brügger, Sara Irina Fabrikant and Arzu Çöltekin

Department of Geography, University of Zurich, Switzerland

ABSTRACT
Elevation change is critical for choosing a route when bicycling. We detail an empirical study in which we comparatively evaluated three linear symbolization types (color hue, color-coded arrows, and elevation profiles) to depict elevation change in bicycle maps for two common bicycle route planning tasks: relative height detection and slope identification. Participants performed most accurately with the color-coded arrows for relative height detection tasks, whereas symbolization did not significantly influence map-use performance for slope identification. Participants preferred the elevation profile, in spite of their lowest performance with this method overall. Our rare empirical findings offer much needed new insights into the function and appropriateness of common elevation symbolization methods, specifically to identify elevation change in bicycle route planning tasks in urban areas where map display real estate is already very limited.

ARTICLE HISTORY
Received 23 December 2015
Accepted 21 May 2016

KEYWORDS
Empirical study; bicycle maps; terrain mapping; map reading; eye tracking

Introduction

Whether it is for planning a route for daily commutes in congested urban spaces or for wayfinding in unknown territory, we often rely on maps and expect them to provide us with relevant environmental context information for solving such navigation tasks. The physical characteristics of the environment appear to be significant in shaping people's navigation choices (Lobben 2004). Freksa (1999) also postulates that people try to minimize both cognitive and physical effort when planning a route. Environmental context is especially relevant when navigating on foot or by bike, where, for example, significant elevation changes can be costly. In other words, the "ups and downs" along movement trajectories will have an impact in the way we make navigation decisions. Consequently, a perceptually salient way to depict elevation change along navigation paths is critical for effective and efficient navigation.

Typically 2D maps are used for everyday navigation and route planning tasks (Çöltekin, Lokka, and Boér 2015; Boér, Çöltekin, and Clarke 2013). Navigational charts or route planners often do not depict the elevation information, especially for motorized transportation modes. If elevation is depicted on road maps, then it is typically rendered over the entire terrain using contour lines or relief shading. This, however, can be problematic in dense urban environments as contour lines or relief shading might clutter the map with irrelevant information for route planning and navigation contexts. Fabrikant, Montello, and Mark (2010) and Hegarty, Canham, and Fabrikant (2010) have shown that perceptually salient, but thematically irrelevant information can hinder map reading performance, and lead to information overload (Cowan 2001). Urban environments tend to have a high density of geographic features within a small footprint. With such information-rich visual displays, the required cognitive effort to extract task-relevant information can be especially difficult. On the other hand, in the case of route planning and navigation, geographic features (e.g. landmarks, water bodies) alongside a route can be just as important as elevation change, while terrain information beyond a route is often considered irrelevant for a navigator on a network (Sutula 2007). One way to address this design challenge is to depict elevation information *only* along routes. We call such solutions "linear elevation symbolization" for the remainder of this paper.

We approach this challenge from a user-centric cartographic design perspective in our study, considering map usability and navigation utility. We aim at minimizing cognitive load for decision-making, which is critical for a cognitively supportive and perceptually salient design of linear elevation symbolizations. We

CONTACT Annina Brügger annina.bruegger@geo.uzh.ch

contend that only when thematically relevant elevation information is rendered in a perceptually salient manner along routes (i.e. in a bicycle map), by considering empirically validated cartographic design guidelines (i.e. using the appropriate visual variables), the map will support map-based decision-making processes effectively and efficiently (Fabrikant, Montello, and Mark 2010; Hegarty, Canham, and Fabrikant 2010).

Below we review approaches to depict thematically relevant elevation change information on a network as to motivate linear elevation symbolization design solutions to be systematically assessed for bicycle maps. Following that, we detail a controlled experiment to comparatively assess participants' map reading performance with typical tasks that are representative of real-life situations when planning a route for a bicycle trip: the assessments of relative height information (height detection) and the appraisal of slopes (slope identification) along bike path segments.

Additionally, we consider the so-called Naïve Cartography theory proposed by Hegarty et al. (2009), given that Naïve Cartography suggests a potential mismatch between users' map style and symbolization preferences compared to their actual map-based decision-making performance.

Related work

We structure the related work section according to the pillars of map design: map theme, map purpose, and map audience. First, we review research on elevation symbolization methods for navigation purposes along routes, followed by relevant human factors research.

Symbolization of elevation information along routes

The undulating terrain is typically shown on maps with either absolute quantitative methods (e.g. spot heights, contour lining), or relative depiction methods (e.g. relief shading, hypsometric tinting) (Muehrcke and Muehrcke 1992). Bertin proposed visual variables *size*, and *color value* (combined with *color hue*) to depict quantitative elevation data on maps for ordinal data (Bertin 1967). These have been tested empirically for thematic maps (Garlandini and Fabrikant 2009), and thus serve as prime candidates to depict elevation change on maps. Huffmann (2009) suggests the visual variable *size*, thus representing slope changes using varying line thickness. Color value and color hue seem to be popularly used on bike maps in recent years (Wessel and Widener 2015), and Su et al. (2010) demonstrate that steep route segments can be quickly identified with the visual variable color hue in a color-coded slope visualization. We contend that, for spatial efficiency reasons, the variable *color* (value|hue|saturation), rather than Bertin's first ranked *size* variable has significant advantages for elevation depiction along a network. Variations in line thickness require map display space that is often not available in already densely symbolized city maps (Wessel and Widener 2015). In city maps, relevant information for navigation are often shown directly on or next to route segments, such as street labels or landmarks. Moreover, the variation of line thickness may be associated with different road types or speed, and potentially mislead map readers.

Besides modifying the visual variables *size* (i.e. line thickness) and *color*, the combined use of the visual variables *shape* and *orientation* (such as with arrows) have also been proposed for coding elevation changes along routes in bike maps, but have not yet been tested empirically. Arrows can denote a spatial interaction between two locations (Tobler 1987), and can also express linearity and asymmetry (Tversky 2001). The arrow symbol is especially useful for networks, as it can indicate direction of movement, quantity of movement flows, order, transitions, and associations (Kurata and Egenhofer 2008). Arrows have been empirically tested for mobile maps in the context of pedestrian navigation (Crease and Reichenbacher 2011), and were suggested to be effective navigational aids to indicate direction (Chittaro and Burigat 2004). Arrows can be combined with other visual variables when additional information is needed; including arrow length, size, color, shape, and labeling. For example, a combination of arrows and labels was found to be an effective method for indicating indoor floor changes (Bigler et al. 2014); which, we hypothesize, is also applicable for representing outdoor elevation changes. Huffmann (2009) suggests changing the size or color saturation of arrows to visualize slope, but these have not yet been empirically tested. These prior studies suggest that, aside from direction, distance and elevation can also be depicted with arrows, while still leaving room for street labels alongside a route. However, it is important to note that arrow symbols can have varied semantic roles, for example, forward or backwards, and up or down (Horn 1998). The added flexibility of intuitively understood arrow symbols to depict a range of directional information on networks can also be its greatest limitation: Arrow symbols need to be understood within each specific map use context (Horn 1998), and thus the interpretation of arrows always needs to be supported by a clear and unambiguous map legend (Slocum et al. 2009). For our study, the question arises how the variety of meanings based on a diversity of visual variables might affect cognitive load for decision-making, and, therefore, influence map-use performance.

In addition to map-intrinsic network symbolization methods such as line thickness, color (value|hue|saturation), and arrows, also map-extrinsic methods are used to depict elevation information along routes, such as with transects or elevation profiles. Elevation profiles are commonly found in hiking maps and digital route planners to depict elevation changes along trails. Elevation profiles are in essence line graphs, where the sequences of spot heights along a route are depicted in chronological order. Profiles are typically shown next to a map display; therefore, a map reader has to integrate the elevation information displayed on the profile with the respective locations on the map. This visual integration is only possible if reference points such as markers, text labels, or lines of equal elevation are provided in the profile and on the map (Sutula 2007). Sutula (2007) contends that people seem to have a good understanding of a trail's terrain characteristics when profiles are provided with maps, but this contention is not based on empirical evidence. We expect that displaying reference points both in the map and on a profile will support elevation information integration in this study. However, it is still an open question whether matching extrinsic elevation profile information with respective locations on a map might lead to more cognitive effort compared to only map-intrinsic linear elevation symbolization using colors and/or arrows. We hypothesize that the integration of extrinsic and intrinsic map information will require more eye movement activity, and be potentially susceptible to split attention effects.

Human factors

People's effectiveness and efficiency in map reading and the related cognitive effort they exert in achieving such visuospatial tasks are not only influenced by map design and the context of use, but also by the individual and group characteristics of map users (Lobben 2004). Individual differences such as visual and spatial abilities, and group differences such as gender, cultural background and training etc. have been shown to predict map reading capabilities (Hegarty et al. 2002). For example, gender and spatial ability seem to interact with certain visuospatial tasks (Montello et al. 1999), and have been shown to influence the effectiveness and efficiency of route selection with map displays (Wilkening and Fabrikant 2011).

Independently from individual and group differences that can predict map-use performance, people can have strong preferences or attitudes regarding map types based on the aesthetics or the (perceived) functionality of a map. However, Naïve Cartography theory posits that such self-reports or intuitions about performance might not always match actual performance with respective map types or designs (Hegarty et al. 2009). In an empirical study relating to realism in map displays, Hegarty et al. (2009) demonstrate that while participants prefer more realistic 3D map displays, they generally perform better with the more abstract 2D maps. Interestingly, even domain experts do not necessarily perform better with their preferred displays (Hegarty et al. 2009). On the other hand, in a similar study, Smallman and Cook (2011) demonstrate that people with higher spatial abilities are able to adjust their preferences after they worked with the displays, while people with lower-spatial abilities do not. For this reason, it is always advisable to combine preference and performance tasks in empirical map use studies.

Below, we detail our study, in which we empirically assessed participants' preferences and performance (i.e. effectiveness and efficiency) in two typical route planning tasks with bicycle maps, based on three linear elevation symbolization methods.

Methods

Consider the following bike route planning scenario: A user consults an interactive online bike route planning application before heading out on a bike ride. Static bike maps with chosen routes and respective elevation profiles are produced after the user has entered a particular start and end location for the planned ride.

We designed a controlled laboratory experiment to comparatively assess three symbolization methods to show elevation changes along routes on such static bicycle maps in urban areas. We varied route segments in the tested maps to show two classes of ordered elevation change information using (1) color hue with a stoplight metaphor (green = downhill, magenta = uphill) on colored route segments (Color), (2) color-coded arrows using the same metaphor (Arrow), and (3) a separate elevation profile placed adjacent to the map display (Elevation Profile). An example stimulus for each symbolization condition is shown in Figure 1. Color and Arrow conditions depict elevation information intrinsically within the map, using different visual variables, while the Elevation

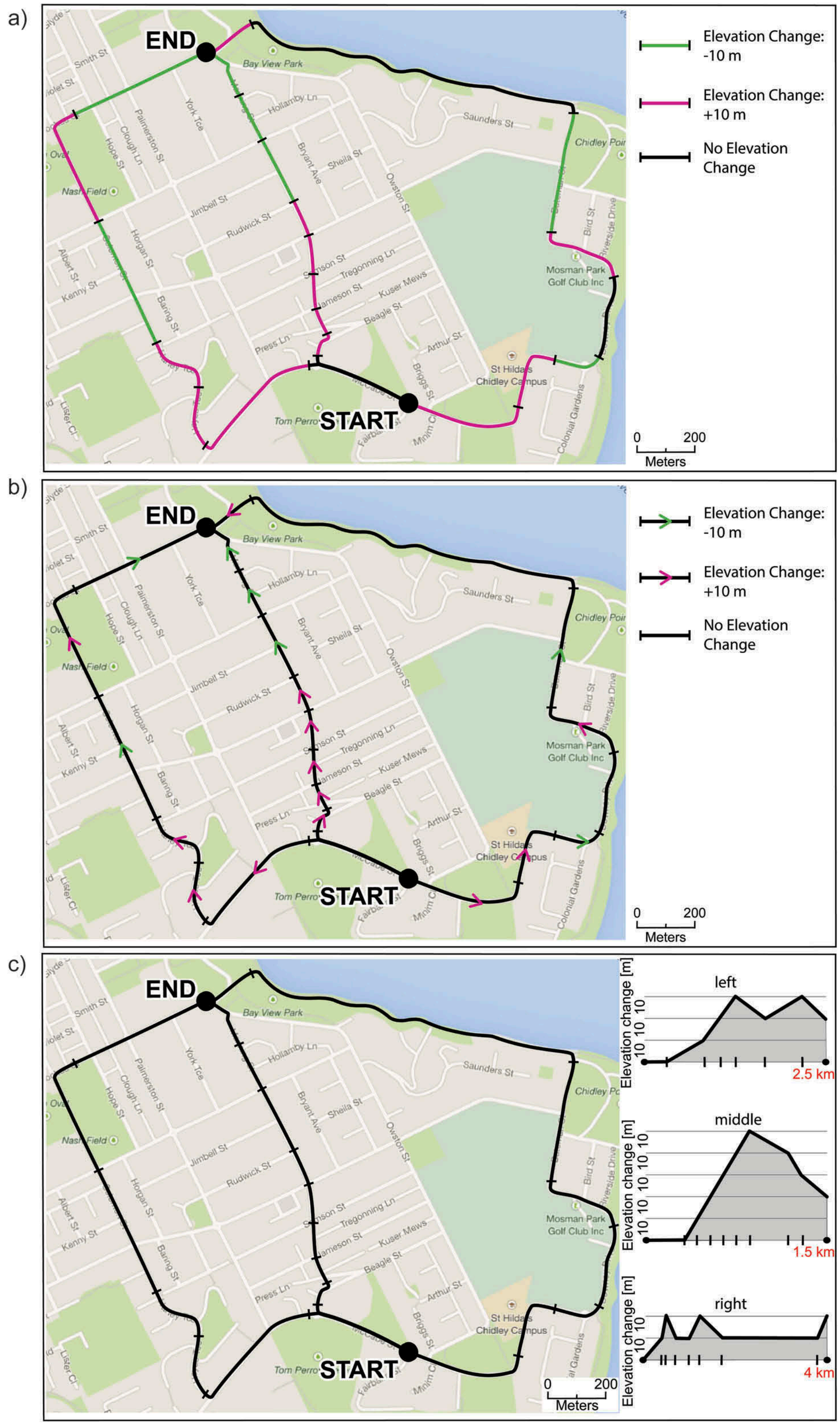

Figure 1. Example stimuli (e.g. Perth base map extract) including three tested symbolization methods: Color(a), Arrow(b), Elevation Profile(c). Due to readability issues, text sizes in the legends have been enlarged for this publication.

Profile is extrinsic to the map. Literature suggests that the Elevation Profile might look more realistic than the Color and Arrow depictions (Sutula 2007).

Inspired by prior work reviewed above, we hypothesized that participants would perform worse with the extrinsic Elevation Profile compared to the intrinsic

Color and Arrow, due to split attention effects potentially increasing cognitive load, as participants switch between map and profile to integrate the elevation information with the locational reference. We further hypothesized that participants would respond more efficiently (i.e. faster) with the Arrow than the Color, as the color-coded arrows explicitly show directional information in addition to elevation changes. Also inspired by the reviewed literature, we predicted that spatial ability would influence decision-making performance and symbolization type preferences. Specifically, we expected participants to intuitively choose what is most familiar to them, irrespective of their actual performance, while we did not expect gender to influence performance.

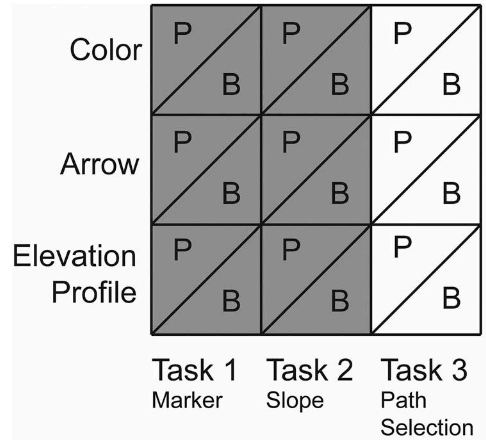

Figure 2. Experimental design including three symbolization methods, three tasks, and two map locations (P = Perth | B = Boston), thus 18 stimuli total. We doubled the size of the trials by inclusion of 18 rotated stimuli (i.e. by 180°). Only results of Task 1 + 2 are reported in this article (gray shading).

Experimental design

In a within-participant factorial design, we implemented the controlled factor symbolization type as our primary independent variable for elevation change along a route with three levels (Color, Arrow, and Elevation Profile). The second independent variable we assessed was task type. We asked participants to solve the following tasks in a bicycle related route planning scenario for all symbolization conditions:

(1) Identify the elevation of three given points along a route, and select the highest point to take a photograph (Marker task)
(2) Select the path that features the steepest slope segment (Slope task)
(3) Select the fastest path (Path selection task)

These tasks were selected for their representativeness in a typical bicycle route planning scenario, that is, accurate identification of elevation and steepness of slopes to minimize physical effort.

We developed six trials for each task by varying the map configuration using two suburban map locations. Each of the six stimuli was designed using the three symbolization types. We thus used 18 map stimuli in total (Figure 2). Each stimulus included a scenario, a specific question, and a map with a given set of response options. Participants received a randomized order of stimuli, and were given a 30-minute limit to complete all tasks. We found the time limit was necessary for the proper administration of the experiment, and we identified this limit based on a pilot test. All participants were tested under identical conditions. Should a time limit introduce stress, it would be the same for all tested conditions, and thus would not affect relative comparisons of results.

Below we only report results for the first two tasks (i.e. Marker and Slope, highlighted with gray shading in Figure 2), due to space limitations. We plan to report on the collected empirical data for Task 3 in a follow up publication. For the remainder of the paper, we thus only detail and discuss the Marker and Slope conditions. The employed Color stimuli are shown in Figure 3. In the first scenario (Marker task), we asked participants to select stop locations along a given route, which is something navigators frequently do when planning a trip (Hochmair 2004; Wilkening and Fabrikant 2013). Specifically, we asked participants to select the highest view point among a given set of yellow markers on the map to take a picture. Figure 3(b) and (d) show an example stimulus with labeled markers (G, P, M). Marker labels were randomly selected for each stimulus, as to avoid the effect of implicit alphabetical ordering in people's minds (Woolfolk, Castellan, and Brooks 1983). For the second scenario (Slope task), participants were asked to identify the path with the steepest slope segment amongst three choices. The start and end positions of the path segments were indicated using tick marks perpendicular to the path. Each question had four possible response options including one correct answer, two false answers, and an "I don't know" option to minimize guessing.

We quantitatively assessed participants' map reading performance (i.e. response accuracy and response time), and additionally collected eye movement data to further analyze participants' visuospatial decision-making processes. To contrast with performance data, we asked participants' map preferences before and after the study. To control for potential individual and group differences, we collected participants' demographic

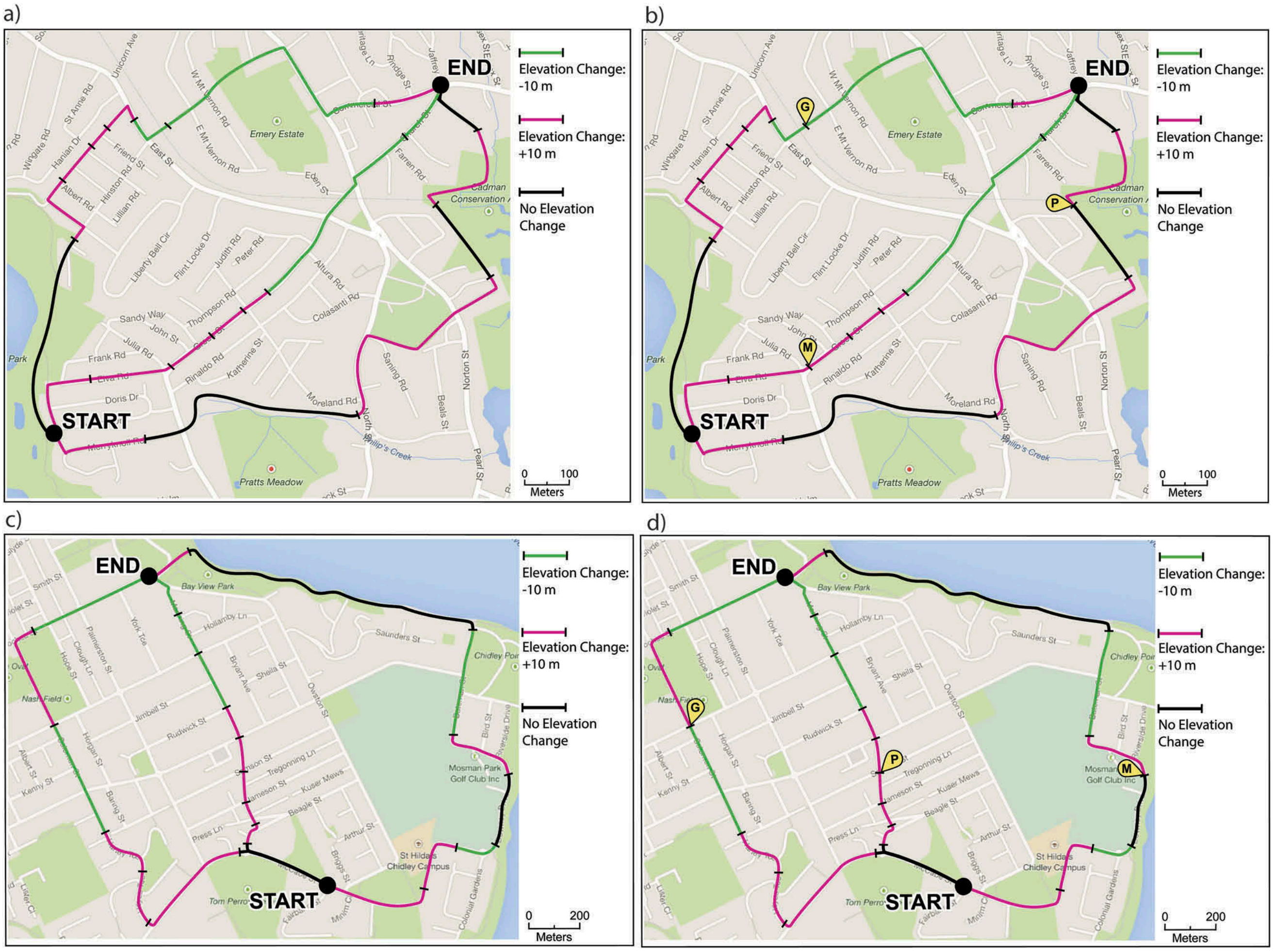

Figure 3. Color symbolization stimuli for Tasks 1–3. Map locations show Boston(a and b), and Perth(c and d). Tasks 2 and 3 included stimuli 3a and 3c. Task 1 included 3b and 3d with three potential viewpoints (i.e. yellow markers G, P, M). Due to readability issues, text sizes in the legends have been enlarged for this publication.

characteristics (i.e. age, gender, expertise, experience, interest in bike maps, etc.), and assessed participants' individual visuospatial abilities with a Hidden Patterns Test (French, Eksstrom, and Price 1963). For the Hidden Patterns Test, we asked participants to find a given (hidden) pattern embedded in 2D spatial configurations. We selected this particular spatial ability test, because for all symbolization types we test, this type of perceptual ability (i.e. the parsing of a linear pattern embedded in a network map or graph surrounded by other map features) is especially important. Following Wilkening and Fabrikant (2013), we split participants into two (high- and low-spatial) groups at the data analysis step, based on their Hidden Pattern Test scores.

Participants

A total of 43 people (23 females and 20 males; between 18 and 45 years) voluntarily participated in the study. They were all students or graduates of the University of Zurich (Switzerland). The majority of the participants (60%) stated that they were professionals or had daily exposure to spatial data, mobile maps, GIS, and cartography. In terms of familiarity, an overwhelming majority (98%) had seen an Elevation Profile before, while only 25% had seen Arrow and Color symbolization types on bicycle maps. In summary: most participants were familiar with Elevation Profiles, but did not have much experience in using them, and they were mostly unfamiliar with the Arrow and Color.

Materials

We extracted two random suburban areas Boston (MA, USA) and Perth (WA, Australia) from Google Maps as base maps for our stimuli. We assume that our participants would have little to no previous knowledge of these places, because none of them were natives of

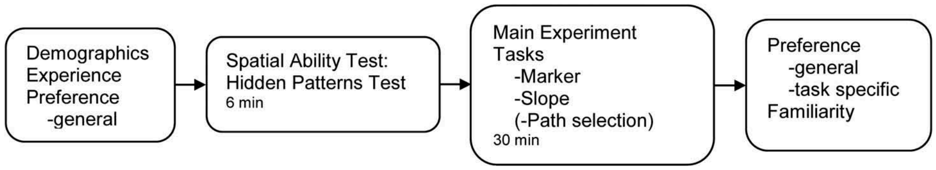

Figure 4. Experiment procedure. Eye movements were recorded only during the main experiment. The spatial ability test and the main experiment included set time limits.

these cities (or countries), and we carefully avoided touristic locations and landmarks. We specifically avoided vertical north-south paths because of the identified cognitive association to uphill/downhill when using navigation instructions with cardinal directions (Brunyé et al. 2012), and due to the well-known vertical–horizontal illusion effect (Gregory 1987). We added three randomly chosen routes and labeled start and end points. Participants could refer to individual routes using the words "left," "middle," and "right," which was specified in the introduction to the experiment.

We additionally rotated the two chosen footprints by 180° to generate additional stimuli without changing the scene content, and randomized their presentation to minimize potential learning effects. The stimuli were optimized to be "informationally equivalent" for comparability (Larkin and Simon 1987): We consistently used elevation *change* for all stimuli (rather than meters above sea-level or slope), kept the scale and the map extent the same, and used identical fonts and font sizes. We segmented and depicted the routes using three elevation classes including uphill (+10 m), no change, and downhill (−10 m). Each elevation change class included a clearly visible and consistent map symbol across the Color and Arrow conditions. For the static Elevation Profile stimuli, path segments and elevation changes were not only indicated with the Elevation Profile, but also with the corresponding start/end locations and tick marks in the maps (see Figure 1(c)). Each task clearly specified to navigate from the marked start to the end location for all three symbolization types.

Procedure

The experiment sessions were held at the eye movement lab at the Department of Geography of the University of Zurich. This lab provides the necessary controlled environment. The study was run on a dedicated Dalco Intel Core i5 760 workstation equipped with an Estecom 23″ color display at a screen resolution of 1920*1080 pixels and connected to a Tobii TX300 eye tracker. Eye movement data was recorded at 300 Hz temporal resolution, with a spatial accuracy of 0.4°. Participants were individually tested, and assigned an unambiguous code only known to the experimenter to ensure anonymity in the data analysis and the reporting of results. At arrival in the lab, participants were introduced to the study procedure, and a consent form was signed by both the participant and the experimenter. Following this, participants filled in a background questionnaire, rated their overall preference amongst the three tested elevation symbolizations, and took a digital version of the Hidden Patterns Test. Then, the participants were given a general introduction to the eye movement technology, and the eye tracker was calibrated for each participant. At this point, a training session that included a map use scenario, respective tasks, and definitions took place. The slope concept and the term route segment were explained with visual examples, as both terms were important throughout the study. Respective warm-up trials followed to ensure that participants had similar knowledge relevant for the main portion of the test. Following the training session, participants solved all tasks of the main experiment in a randomized order, and filled in a post-test questionnaire indicating task-specific preferences regarding the tested symbolization types. Following this, they again rated their overall symbolization type preferences (irrespective of tasks). In doing so, we could cross-check whether their preferences had changed through the exposure to the map stimuli used during the test session. After the completion of the experiment, we offered participants a bar of chocolate as a thank you gift for participation. An overview of the experiment procedure is schematically depicted in Figure 4.

The experimenter orally informed participants of the remaining time after 15 minutes had passed, and again 5 minutes before the allotted 30 minutes were up. Additionally, progress was also displayed at the top left corner of the screen. On average, participants needed about 50 minutes to complete the entire session.

Results

Below we first present performance results across symbolization type conditions, followed by results relating to user background and training, spatial ability, and gender. We complete this section with eye-movement analyses based on selected areas of interest (AOI), followed by users' symbolization type preferences.

Response accuracy

As stated above, participants were asked to find the location with the highest elevation on a route (Marker task) and the steepest slope segment (Slope task) on a map of an urban area with different elevation symbolizations. We coded the "I don't know" response as false, based on the assumption that if a participant was not able find the answer, then the respective map design did not provide effective and efficient decision-making support. The overall accuracy was calculated as the proportion of correct answers to the total number of possible correct answers (a total of 12) across both tasks. Figure 5 shows the overall average of correct responses (irrespective of task) across the three tested symbolization types.

As shown in Figure 5, participants are most accurate (M = 87%, SD = 4%) in their decision-making with the Arrow, closely followed by Color (M = 84%, SD = 5%), and as hypothesized, are least effective with the extrinsic Elevation Profile (M = 70%, SD = 4%). A repeated measures ANOVA reveals a statistically significant difference across symbolization types ($F(2,78) = 17.3$, $p < .001$, $\eta^2_p = .307$). A pairwise comparison of main effects further indicates a statistically significant difference ($p < .01$) between the Elevation Profile with both the Arrow, and the Color ($p < .01$), but not between the Arrow and the Color ($p > .05$).

We further analyzed the collected data considering potential effects of gender and spatial ability. As hypothesized, we do not observe any significant gender effect for accuracy. However, as has been shown in many prior

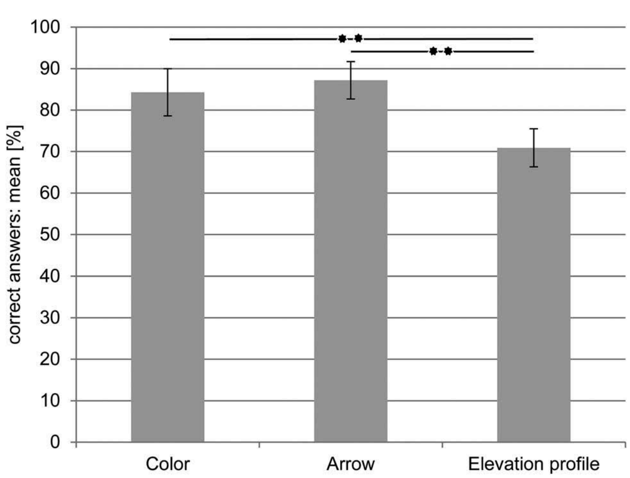

Figure 5. Overall mean accuracy for both map-reading tasks across the symbolization types. Error bars: ±2SEM, **p < .01.

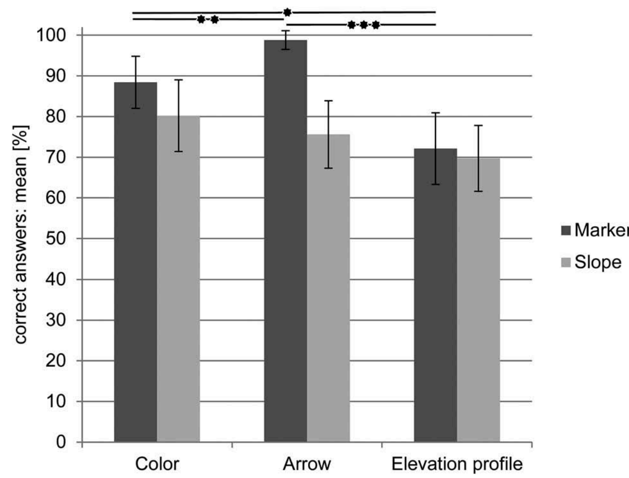

Figure 6. Mean accuracy for the Slope and the Marker tasks, across the symbolization types. Error bars: ±2SEM, *p < .05, **p < .01, ***p < .001.

studies involving spatio-temporal decision-making with maps, we do find a significant effect of spatial ability ($p < .05$). Overall, high-spatial participants are more accurate in their decision-making (M = 85%, SD = 8%) than low-spatial participants (M = 77%, SD = 12%). However, spatial ability does not interact with symbolization type, that is, high-spatial participants are more accurate in their decision-making irrespective of the symbolization type.

Next, we turn to the main factor task type. Figure 6 shows the average of correct responses in percent for the Slope and Marker tasks across all tested symbolization types.

As shown in Figure 6, for the Slope task, participants are most accurate with Color (M = 80%, SD = 8%), followed closely by Arrow (M = 75%, SD = 8%). We again find the lowest average response accuracy for the trials with the Elevation Profile (M = 69%, SD = 8%). However, these differences are not statistically significant ($F(2,78) = 1.808$, $p > .05$, $\eta^2_p = 0.044$). We also do not observe any effects of gender or spatial ability ($p > .05$).

However, the response pattern looks different for the Marker task. Overall, participants achieve higher accuracy scores in the Marker trials compared to the Slope trials, which could have been perceived to be more difficult (Wilkening and Fabrikant 2013). Participants reach near-perfect accuracy (M = 98%, SD = 2%) with the Arrow symbolization type, followed by Color (M = 88%, SD = 6%), and they perform worst (nearly 27% more errors) with the Elevation Profile (M = 72%, SD = 8%). Symbolization type significantly affects response accuracy ($F(2,78) = 15{,}464$, $p < .001$, $\eta^2_p = 0.284$) for the Marker trials. A pairwise comparison further reveals that response accuracy significantly differs between tested symbolization types in the Marker trials ($p < .05/.01/.001$). These differences are apparent in Figure 6. We do not find any effects of spatial ability for the Marker task ($p > .05$), but surprisingly, we do observe significant interaction effects between gender and response accuracy ($p < .05$). While men's response accuracy is significantly lower compared to that of women with the Elevation Profile (M_{Men} = 76%, M_{Women} = 76%), they perform better than women with the Arrow (M_{Men} = 100%, M_{Women} = 97%) and Color (M_{Men} = 0.97%, M_{Women} = 80%).

Response time

All participants were able to answer all questions within the given time limit. Figure 7 illustrates average response times for each stimulus across symbolization conditions (for correct responses only). As mentioned above, we only report results for the first two (Marker and Slope) tasks which were answered much more rapidly than the third (Path selection) task (thus our results yield relatively short average response times between 30–40 s).

Overall, as Figure 7 shows, the response time patterns across symbolization conditions are similar to the pattern visible in Figure 5 for response accuracy. While response times are almost identical for the Color (M = 34.0 s, SD = 4.3 s) and the Arrow (M = 34.7 s, SD = 3.5 s), participants are slower with the map-extrinsic Elevation Profile (M = 44.2 s, SD = 4.9 s), as expected. On average, participants need approximately 10 s longer to respond with the Elevation Profile than with the Arrow or Color. A repeated measures ANOVA indicates that this response time difference is statistically significant ($F(2,78) = 11.406$, $p < .001$, $n^2_p = 0.226$). While pairwise comparisons do not reveal any response time differences between the Arrow and Color ($p > .05$), participants respond significantly faster with both the Arrow ($p < .01$) and the Color ($p < .001$), compared to the Elevation Profiles. Interestingly, high-spatial participants spend significantly more time ($p < .05$) to respond with the Elevation Profiles than the low-spatial participants ($M_{high,ElevationProfile}$ = 46.2 s, $M_{low,ElevationProfile}$ = 42.1 s). As expected, the two intrinsic methods seem to support more efficient decision-making: Both the high- and low-spatial participants need less time to respond with the Arrow and Color on average ($M_{high,Arrow}$ = 30.5 s, $M_{high,Color}$ = 31.3 s, $M_{low,Arrow}$ = 38.9 s, $M_{low,Color}$ = 36.8 s) in comparison to the Elevation Profile. We do not find any interaction effects with respect to gender ($p > .05$).

Similarly to the response accuracy analysis procedure, we analyzed the response time patterns across map use tasks. Figure 8 illustrates the mean response times for the two tasks across elevation symbolization methods.

As can be seen in Figure 8, for the Slope task, participants need on average 29.6 s (SD = 5.2 s) to respond with the Color, compared to the Arrow (M = 32.8 s, SD = 5.4 s), and the Elevation Profile (M = 33.3 s, SD = 5.7 s). A repeated measures ANOVA reveals no significant differences in response times across the symbolization methods ($F(2,78) = 0.755$, $p > .05$, $n^2_p = 0.019$). We do not discover any effects of spatial ability or gender for the Slope task ($p > .05$).

Mirroring the results for response accuracy, Figure 8 again reveals a different pattern for response time for

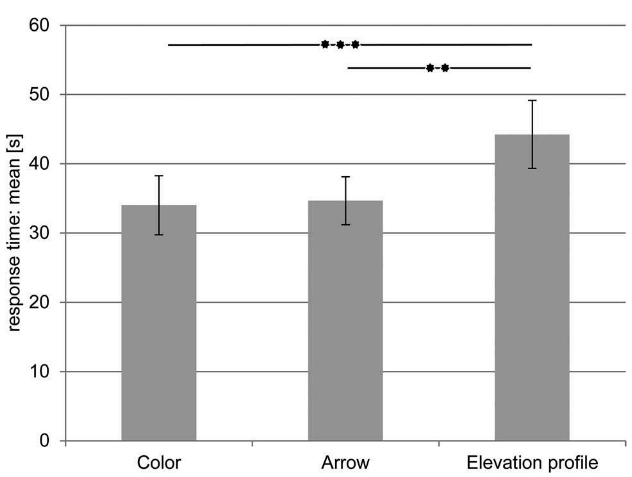

Figure 7. Mean response times in seconds for both map reading tasks across the symbolization types. Error bars: ±2SEM, **p < .01, ***p < .001.

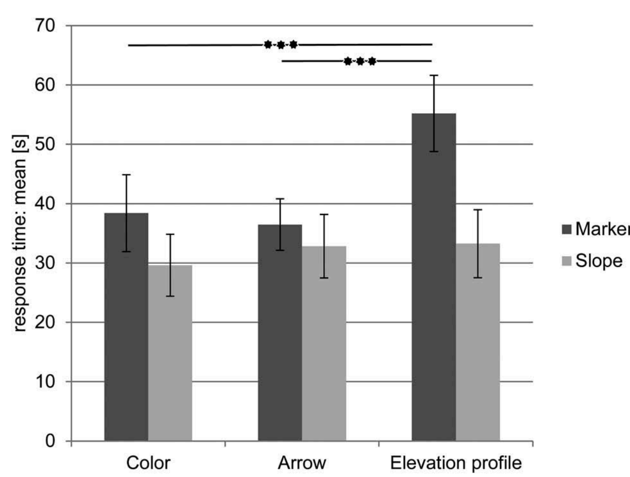

Figure 8. Mean response time in seconds for the Slope and Marker tasks across symbolization types. Error bars: ±2SEM, ***p < .001.

Table 1. Correlation analysis between overall response time and response accuracy across map use tasks.

Correlation	Both tasks	Marker	Slope
	0.018 (0.911)*	−0.009 (0.955)*	−0.002 (0.991)*

*Pearson correlation coefficient *r* (*p*-value).

the Marker trials. On average, participants take almost a minute (M = 55.2 s, SD = 6.42 s) to solve the Marker task with the map-extrinsic Elevation Profile. The Arrow trials not only yield the most accurate, but also the fastest responses (M = 36.5 s, SD = 4.3 s), closely followed by the Color (M = 38.4 s, SD = 6.5 s). These differences are statistically significant ($F(2,78) = 18.222$, $p < .001$, $n^2_p = 0.318$). Pairwise comparisons reveal that participants are statistically significantly faster with both map-intrinsic symbolization methods (Arrow and Color), than the map-extrinsic Elevation Profile ($p < .001$). The response times of the two intrinsic methods do not differ significantly ($p > .05$). Spatial ability does not affect response times in the Marker task condition ($p > .05$). On average, women need considerably more time (M = 48.4 s, SD = 12.8) for the Marker task than men (M = 37.5 s, SD = 10.8 s), and this difference is statistically significant ($F(1,39) = 8.371$, $p < .01$, $n^2_p = 0.177$).

Since participants on average spend more time to respond to the Marker tasks than to the Slope tasks across all symbolization methods, we further analyzed whether spending more time on a task might lead to higher response accuracy. However, respective computed correlations do not suggest any speed-accuracy trade-offs, as shown in Table 1 above.

In summary, participants perform worse – both in terms of response accuracy and response time – with the map-extrinsic Elevation Profile compared to the map-intrinsic Color and Arrow symbolization methods. Furthermore, we observe that gender and spatial ability influence map-use performance only for the Marker task. Next, we turn to the analysis of the eye movement recordings to better understand how and why these performance differences might have occurred.

Eye movement analysis

We subjected participants' eye movements to an AOI analysis, to investigate which and how specific parts of the map display (or design elements) might have influenced viewers' decision-making processes. We segmented the stimuli into "task-relevant" and "task-irrelevant" AOIs, that is, relevant/irrelevant to finding the correct answer. We recorded the proportion of time spent fixating at a particular AOI with respect to the total time spent looking at the entire map display.

On average, participants fixate the task-relevant AOIs for over 75% of their total time studying the stimuli in the Arrow condition (M = 75%, SD = 2%). The average fixation durations are lower for the task-relevant AOIs in the Color and the Elevation Profile conditions ($M_{Color} = 67\%$, $SD_{Color} = 4\%$, $M_{ElevationProfile} = 63\%$, $SD_{ElevationProfile} = 4\%$) than for the Arrow. We find a statistically significant difference between time spent looking at task-relevant AOIs across the three studied elevation symbolization methods ($F(2,76) = 13.613$, $p < .001$, $n^2_p = 0.264$). Pairwise comparisons indicate that time spent on the task-relevant AOIs in the Arrow condition is significantly longer ($p < .05$) compared to the Color and the Elevation Profile methods. Interestingly, we could not find a statistically significant difference in viewing durations of task-relevant AOIs between the Color and the Elevation Profile conditions ($p > .05$), even though people need significantly more time to respond in the Elevation Profile. One reason for this surprising effect might have to do with additional screen estate relevant for answering the test questions: the legend information. Therefore, we delineated two additional AOIs: One covering the whole map area, and a second one covering the legend area (refer to Figure 1 to see the type of legend areas). The results of this AOI analysis are summarized in Table 2 below.

Table 2 shows that average AOI fixation durations for the Arrow and Color conditions are ten times longer on the map than on the legend. Conversely, in the Elevation Profile condition, the average viewing time is almost equally divided between the map area and the legend area, where the Elevation Profile is located. The AOI for the map area was, on average, focused longer (17–20 s) than the legend AOI (1–22 s). A repeated measures ANOVA indicates that the viewing duration differences between map and legend AOIs are statistically significant ($F(2,76) = 132.609$, $p < .001$, $n^2_p = 0.777$) across all experimental conditions ($p < .05$). These fixation time differences could potentially support the hypothesized split attention effect, suggestive of performance decrease in the Elevation Profile condition.

Table 2. Mean eye fixation durations for the areas of interest Legend and Map across symbolization types.

Symbolization type	Eye fixation duration* Legend	Map
Color	2.4	20.9
Arrow	1.2	17.0
Elevation profile	22.7	18.3

*mean in seconds.

User preferences

To systematically assess whether participants' intuitions about symbolization methods and their symbolization type preferences matched their performance, we asked them to state their symbolization type preferences at three instances during the experiment: once *before* the experiment (overall preferences, not task-specific), and twice *after* the experiment (overall preferences again, and then the task-specific preferences). Below we detail this comparative preference rating assessment. First, we present the results of the overall display preferences, collected before and after the experiment (Table 3).

As can be seen in Table 3, the majority (61%) of the participants (26) do not change their preferred symbolization type, while almost 40% (17) do, as a result of their experience with the experimental tasks and stimuli. Seven out of these 17 are high-spatial participants (41%), and 10 are low-spatial (59%). Irrespective of spatial abilities, we see that nearly half (49%) of the participants (21) prefer the Elevation Profile prior to the experiment, and keep this preference also after completing the study, even though our analysis shows that participants perform worse overall with this elevation symbolization. Furthermore, it is interesting to note that six participants (13%) switch their overall preference to Elevation Profile after the experiment (four of these six participants prefer the Color before the experiment, and two prefer the Arrow).

Task-specific preferences are summarized in Figure 9. The majority (60%) of the participants prefers the Elevation Profile specifically to identify steepness in the Slope task, followed by 23% that prefer the Color type (10), and lastly the Arrows are preferred by 12% of the participants (5). Two participants do not state any preference. This pattern changes for the Marker task, which better reflects participants' performance: Only 27% of the participants (12) prefer the Elevation Profile, while 44% (19) prefer Arrow for reading off and comparing heights. A few more participants prefer the Elevation Profile over Color, which indeed is not indicative of their actual performance.

Table 3. Change of overall symbolization type preferences before and after the experiment. Numbers indicate raw totals of participants (N = 43). Gray Shaded cells: no preference changes.

Symbolization type preference before the experiment	Symbolization type preferences after the experiment			
	Color	Arrow	Elevation profile	No preference
Color	2 (5%)	2 (5%)	4 (9%)	1 (2%)
Arrow	3 (7%)	3 (7%)	2 (5%)	0
Elevation profile	3 (7%)	1 (7%)	21 (49%)	1 (2%)

Figure 9. Self-reported symbolization type preferences across map use tasks. Numbers indicate raw totals of participants (N = 43).

Preference versus performance

To further study whether and how symbolization preferences might relate to actual performance, we grouped participants based on their symbolization preferences. Table 4 lists the average percentages of correct answers overall for each participant group, based on their stated preferences. The two participants who did not state any preference were excluded from this analysis (N = 41).

Table 4 shows that the eight participants who prefer the Color (first row) are fastest and most accurate with the Arrow, that is, not with their preferred symbolization type. Interestingly, however, we see a significant positive correlation between response time and accuracy in the Color condition for these participants, but not with the other symbolization types. Taking more time in the Color trials yields higher accuracy. The six participants who prefer the Arrow (second row) are indeed fastest with the Arrow, but slightly (~3%) more accurate with Color. The 27 participants who prefer the Elevation Profile (third row) are most accurate with Color, and fastest with Arrow. For this group, we detect a strong positive correlation between response time and accuracy with Elevation Profiles. Taking more time in the Elevation Profile trials also yields higher accuracy. In this case, response speed and accuracy

Table 4. Correlations of response times and response accuracy across self-reported symbolization type preferences and symbolization conditions.

Preference(N)	Color	Arrow	Elevation profile
Color (8)	29.1/81#	28.6/91#	36.9/83#
	0.76(.03)*	0.12(.6)*	-0.230(.58)*
Arrow (6)	34.3/90#	29.1/87#	35.8/86#
	-0.23(.67)*	-0.50 (.31)	0.94(.01)*
Elevation profile (27)	38.2/75#	37.3/62#	46.2/71#
	-0.51(.01)*	-0.197 (.33)*	0.40(.037)*

#Response times in seconds/response accuracy in percent.
*Pearson correlation coefficient *r* (*p*-value).

negatively correlate in the Color condition, and positively correlate for the Elevation Profile, both at a moderate level.

Overall, participants' performance does not directly match their preferences (or performance intuitions). They are not faster nor more accurate with their overall preferred symbolization type (Elevation Profile). Only when they use more time compared to the other symbolization methods, does their accuracy increase with their preferred symbolization type.

Below we report on symbolization preferences vs. performance across tasks. Table 5 shows response accuracy and response times.

Studying the response accuracy patterns in Table 5, it seems that irrespective of their symbolization preferences, participants' response accuracy is consistently higher with the Marker task, suggesting that the Marker task was overall easier for them. For participants who preferred the Arrow and Color, response accuracy is about 15% higher for the Marker task compared to the Slope task. This difference is only about 10% for those who preferred the Elevation Profile. A different pattern emerges for the response time analysis. Interestingly, people who prefer the Color symbolization type need about the same amount of time for responding, irrespective of task type. The time difference is only 3%. The response time difference across tasks is much larger for the Arrow (~16%) and for the Elevation Profile (~15%). The Marker task takes participants much longer to solve, no matter which symbolization type they prefer, suggesting that the Marker task was not necessarily equally easy in all conditions.

We again computed correlations to investigate whether and how symbolization preferences might relate to actual performance (i.e. response accuracy and response time), specifically for the Slope task. This analysis, also reported in Table 5, does not reveal statistically significant results.

These analyses show that there were differences in participant performance in map-based decision-making based on the symbolization type (see Figures 5–8), and, overall, participants prefer a display type which did not facilitate their decision-making performance, that is, the Elevation Profile. However, when we study the relationship between performance and preference more in depth, we do not observe a clear Naïve Cartography pattern in all conditions. This is especially true for the Arrow condition. Participants who prefer the Arrow are faster with it. Even though they are slightly more accurate with Color, this can be mostly explained by the task difference. Color appears to work better for the Slope task compared to the Arrow. Participants seem to be somewhat aware that they perform better with the Arrow when they solve the Marker task, and in this case they are able to judge which of the symbolization types is more helpful to them. Nonetheless, the clear overall preference for the Elevation Profile irrespective of task type may indicate that they might not have enough experience using the intrinsic symbolization types for the Marker task, as to influence their general symbolization preference.

Speed-accuracy relationships show another interesting pattern that emerges from this analysis. For example, even though participants who prefer the Elevation Profiles achieve best performance results with the Color, their response accuracy *does* increase when they spend more time with the Elevation Profile. In fact this is true also for the Color condition. Hence, in two out of three conditions, as participants use more time with their preferred symbolization type, their accuracy increases, suggesting a more engaged decision-making process with their preferred display type.

Table 5. Correlations of response times and response accuracy across self-reported symbolization type preferences and task conditions.

Preference	Both tasks	Marker	Slope
Color	33.8 /82#	39.2 /88#	35.8 /75#
	0.16(.69)*	−0.07(.02)*	−0.01(.97)*
Arrow	31.6 /81#	43.5 /88#	27.9 /73#
	−0.22(.67)*	0.13(.61)*	0.15(.81)*
Elevation profile	39.6 /80#	46.4 /83#	31.1 /74#
	−0.06(.76)*	−0.22(0.49)*	0.07(.73)*

#Response times in seconds/response accuracy in percent.
*Pearson correlation coefficient *r* (*p*-value).

Discussion

We set out to empirically study how people assess elevation change depicted along a route by systematically comparing three linear elevation symbolization methods: Color (color hue), Arrow (color-coded arrows), and Elevation Profile; which are commonly used in bicycle maps. We identified three key dimensions relevant for this empirical study: the elevation depiction method (i.e. symbolization type), typical map use tasks for planning a bicycle trip (i.e. identifying relative spot heights, and assessing slope along a route), as well as participant characteristics (i.e. gender and spatial ability). Below we discuss our findings based on these evaluation dimensions.

Our results support prior findings by Irvankoski (2012), suggesting that visualization characteristics do affect response accuracy specifically for judging the height differences between two points. Overall, as hypothesized, participants are more accurate and faster in their decision-making with the map-intrinsic elevation symbolization types: Arrow and Color, compared to the map-extrinsic Elevation Profile (compare Figures 5 and 7). The eye movement analysis, based on fixation durations of task-relevant AOI in the map, provides a potential explanation for participants' low performance with the extrinsic Elevation Profile. As prior literature suggests (Opach, Gołębiowska, and Fabrikant 2013), this map-extrinsic display of task-relevant information might require extra perceptual and cognitive effort to overcome a split of visual attention needed to integrate the locational information on the map separated from the elevation graph in the legend. This split of attention effect might, in the worst case, be even interrelated with change blindness (Rensink 2009). Note that we added reference points at segment divisions both on the path and the profiles (see Figure 1, as suggested by Sutula 2007) in order to support the integration of location information in the map with the elevation information in the legend.

Participants' decision-making performance with respect to the two intrinsic elevation symbolization methods (Color and Arrow) is best discussed in the context of the two typical bicycle route planning tasks we selected for this study. Prior map design studies suggest that users' performance differences are rarely due to map design or symbolization type alone, but are typically interrelated with task type or task complexity. Confirming prior work (Wilkening and Fabrikant 2011), map use context, specifically task complexity, did influence participants' decision-making performance. For the arguably less complex Marker task, participants performed most accurately and completed the task most rapidly with the Arrow, closely followed by Color, and they took longest and were least accurate with the Elevation Profile, as suggested above, we believe this is because the Elevation Profile is a map-extrinsic symbolization method (Opach, Gołębiowska, and Fabrikant 2013).

Overall, participants achieved higher accuracy with the Marker task, but spent more time solving this task compared with the Slope task across all symbolization types (see Figures 6 and 8). Only for the Marker task, requiring them to simply read off and compare relative spot heights, the tested symbolization types lead to statistically significant differences in accuracy and response times compared to the arguably more complex Slope task (see Figure 6 and 8). Slopes are not explicitly depicted in the map, but have to be inferred by evaluating segment length (i.e. distance along a path) together with symbolized elevation change, further explained in the legend. This added complexity might explain why participants overall make more errors in the Slope task compared to the Marker task, irrespective of the symbolization type (Figure 6).

Our results empirically support Huffmann's (2009) contentions that color alone is not enough for elevation related tasks, such as finding the elevation of a point, similar to our Marker task. As contended by Huffmann (2009), employed arrows effectively and redundantly encode not only qualitative differences of elevation change by using different color hues (e.g. magenta for uphill, green for downhill), but also show navigation direction along a route (Kurata and Egenhofer 2008).

It is important to note that the Marker task involved summing the symbolized ±10 m elevation change classes from the start of a route to an indicated marker. It seems that the arrows, in essence compact point symbols along a route, can be counted and summed more easily than counting extended colored line segments. Eye movement analysis indicates that the arrow heads may have indeed guided participants more effectively to the thematically relevant path segments, compared to the other symbolization types. This is possibly due to arrows supporting the rapid detection of the thematically relevant start and end points of a route. In the Color and Elevation Profile conditions, participants spent more time on task-irrelevant segments compared to the Arrow condition. This confirms prior findings by Chittaro and Burigat (2004).

To systematically assess how participant characteristics might have influenced the study outcomes, we analyzed individual and group differences based on spatial abilities and gender. As already documented in various previous studies involving map-based decision-making tasks (Wilkening and Fabrikant 2011), we observed that high-spatial participants, irrespective of display design decisions, were overall more effective than the low-spatial participants in their map-based decision-making. Interestingly, high-spatial participants overall used more decision time with the Elevation Profile than the low-spatial participants, suggesting that they might have more carefully performed the task, given the greater perceptual and cognitive load of the task, while at the same time mitigating split attention effects more easily. This finding is somewhat counter intuitive, as high-spatial participants have been shown to be faster than low-spatial participants when working with visual displays (Hegarty et al. 2006). On the other hand, our results could be explained perusing Kahneman's (2011) central thesis of a dichotomy of thought between an intuitive "System 1" based on fast, but often biased and thus inaccurate heuristics, compared to a slower, more deliberate "System 2," requiring more logical thinking.

High-spatial participants thus might have better understood the complexity of the task, and thus literally given more and longer thought to solving it as accurately as possible. One might also note that the employed Hidden Patterns Test to group participants on spatial ability, in essence, measures the ability to identify an embedded linear pattern inside other linear structures. This arguably is most relevant within a complex map pattern, and it might have less relevance for the Elevation Profile condition in comparison to the Color and Arrow conditions, where the elevation information was embedded within the street network in the map.

As hypothesized, overall, we did not find any significant differences in decision-making based on gender. However, surprisingly, at the task level, women seem to be more effective than men with the map-extrinsic Elevation Profile with the Marker task. Conversely, men are more effective than women with the intrinsic Arrow and Color symbolization types. This might suggest that perhaps spatial ability (based on the perceptual Hidden Pattern Test) might have played a role. However, we did not find any correlations between spatial ability and gender. Women also used more time for their responses compared to men for the Marker task, and this lead to better accuracy, irrespective of the symbolization type with this task. Spatial abilities and expertise levels are similar between men and women, thus these cannot explain the observed differences. However, male overconfidence is a known phenomenon in the psychology literature for many domains of life (Bengtsson, Persson, and Willenhag 2005), and has been shown to play a role also in map-based decision-making (Wilkening and Fabrikant 2011). Were women indeed to be more deliberate and thus choosing to engage Kahneman's slower but logical System 2 over the rapid and intuitive System 1 for decision-making in the seemingly easier Marker task, it appears to have been the better strategy for accuracy, especially also when working with the more demanding Elevation Profile symbolization.

Other human factors to be considered in empirical map-based decision-making studies are people's attitudes, intuitions, and map design preferences. Overall, participants preferred the Elevation Profile both before and after the experiment, even if they made more errors with it than with the map-intrinsic elevation symbolization types (Arrow and Color), as we hypothesized. This, in essence, confirms Hegarty et al. (2009) Naïve Cartography theory, suggesting that participants' intuitions/self-reports about own performance and preferences might not always match actual performance. This might be partly explained by, for example, prior familiarity with tested symbolization types. In the context of elevation depictions, our participants were considerably more familiar with the Elevation Profile (98% of the participants) compared to Color and Arrow (25% of the participants). Another explanation might be the intuitively understandable direct visual cue for elevation changes and slopes in the Elevation Profile. Elevation profiles make the three dimensional structure of the terrain visually explicit, and slopes perceptually salient (Sutula 2007), thereby possibly afford an intuitive impression of steepness. Huffmann (2009) reports that map users expect to see the topography directly, and Sutula (2007) contends that elevation profiles give people a good understanding of the characteristics of a trail of interest. These might explain why participants clearly prefer the Elevation Profile specifically for the Slope task, while their preferences are more varied for the Marker task. Overall, participants prefer the Arrow for the Marker task, with which they also perform most accurately. Similarly to Wilkening and Fabrikant (2011), this is also an example where empirical findings do not fully support Naïve Cartography theory.

Interestingly, participants spent more time with their preferred symbolization type in the case of the Color and the Elevation Profile, which in turn seems to have improved response accuracy. One interpretation might be that if participants prefer a certain symbolization, they might wish to engage with it more, and thus might spend more time with it, which might lead to better accuracy. User preferences can also change during display use and task completion (Levy et al. 1996). While participants seem to have more nuanced and malleable preferences at the task level, their overall preference for the Elevation Profile did not change in the course of the experiment. Contrary to the findings of Smallman and Cook (2011), we find that more of the low-spatial participants change their preference during experiment compared to the proportion of high-spatial participants.

Conclusions and outlook

We detailed an empirical study in which we comparatively assessed three symbolization types for depicting elevation change in bicycle maps: map-intrinsic Color (color coded line segments), Arrow (color-coded arrows), and map-extrinsic Elevation Profiles. We tested these symbolization methods in the context of basic bicycle route planning tasks: Marker task (relative spot-height identification) and Slope task (slope assessment). Our rare empirical findings offer much needed new insights into the function and appropriateness of common elevation symbolization methods, specifically to identify elevation change in bicycle route planning tasks in urban areas where map display real estate is already very limited. As hypothesized, participants performed best with the map-intrinsic and space-efficient color-coded arrows, but only for relative spot-height identification, whereas elevation symbolization types did not significantly influence map-use performance for slope identification. Overall, individual differences such as spatial ability *do* influence visuospatial decision-making with maps as our study shows, even beyond display design as prior work has suggested (Wilkening and Fabrikant 2011). However, the potential role gender might play in this context emerges less clearly in our research. While participants prefer the more familiar Elevation Profile, they perform worst with this elevation symbolization method overall, thus giving support to the Naïve Cartography theory (Hegarty et al. 2009) despite the nuanced findings at the task level. The eye movement data collection method turned out to be critical to better understand how participants arrive at a particular decision. Armed with respective decision theories (Schulte-Mecklenbeck, Kühberger, and Ranyard 2011), this analysis could be expanded to further assess potential decision-making strategies across different participant groups and task contexts. As we embrace more and more digital and interactive maps on mobile navigational assistance devices, a future empirical study could focus on interactive versions of these linear elevation symbolization types. One might imagine moving the tested, static, map-extrinsic Elevation Profile into the map display using advanced interaction methods. All tested methods may potentially benefit from interactive querying mechanisms, including highlighting, brushing and linking methods. A future empirical study could test symbolization methods on mobile displays on the move, and in the wild. While many more empirical design studies are possible, our results demonstrate that not only design choices might have an impact on map user performance or preferences, but importantly, also participant characteristics.

Acknowledgements

We would like to thank all participants for taking part in our experiment. We would also like to thank the reviewers for their helpful and in-depth comments that helped improve our paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Annina Brügger http://orcid.org/0000-0003-3517-664X
Sara Irina Fabrikant http://orcid.org/0000-0003-1263-8792
Arzu Çöltekin http://orcid.org/0000-0002-3178-3509

References

Bengtsson, C., M. Persson, and P. Willenhag. 2005. "Gender and Overconfidence." *Economics Letters* 86 (2): 199–203. doi:10.1016/j.econlet.2004.07.012.

Bertin, J. 1967. *Sémiologie Graphique: Les Diagrammes, Les Réseaux, Les Cartes*. Paris: Gauthier-Villars.

Bigler, S., A. Brügger, F. Utzinger, and K. F. Richter. 2014. "Up, Down, Turn Around: Assisted Wayfinding Involving Level Changes." In *Spatial Cognition IX - Internactional Conference, Spatial Cognition*, edited by C. Freska, B. Nebel, M. Hegarty, and T. Barkowsky, 176–189. Cham, Switzerland: Springer International. http://link.springer.com/chapter/10.1007/978-3-319-11215-2_13.

Boér, A., A. Çöltekin, and K. C. Clarke. 2013. "An Evaluation of Web-based Geovisualizations for Different Levels of Abstraction and Realism – What do users predict?". Proceedings of the 26th International Cartographic Conference – From Pole to Pole, Dresden, August 25–30. http://icaci.org/files/documents/ICC_proceedings/ICC2013/_extendedAbstract/284_proceeding.pdf.

Brunyé, T. T., S. A. Gagnon, D. Waller, E. Hodgson, S. Tower-Richardi, and H. A. Taylor. 2012. "Up North and down South: Implicit Associations between Topography and Cardinal Direction." *Quarterly Journal of Experimental Psychology* 65 (10): 1880–1894. doi:10.1080/17470218.2012.663393.

Chittaro, L., and S. Burigat. 2004. "3D Location-Pointing as a Navigation Aid in Virtual Environments." In *Proceedings of the Working Conference on Advanced Visual Interfaces - AVI '04*, 267–274. New York: ACM. doi:10.1145/989863.989910.

Çöltekin, A., I.-E. Lokka, and A. Boér. 2015. "The Utilization of Publicly Available Map Types by Non-Experts - A Choice Experiment." *Paper presented at the 27th International Cartographic Conference, 16th General Assembly*, Rio de Janeiro, August 23–28. http://www.icc2015.org/trabalhos/13/593/T13-593_1430419913.pdf.

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.

Crease, P., and T. Reichenbacher. 2011. "Designing Usable Cartographic Representations of Geographic Relevance for LBS Users." 25th International Cartographic Conference, International Cartographic Association, Paris, July 3–8. http://icaci.org/files/documents/ICC_proceedings/ICC2011/Oral%20Presentations%20PDF/B1-Location%20based%20services%20and%20ubiquitous%20cartography/CO-073.pdf.

Fabrikant, S. I., D. R. Montello, and D. M. Mark. 2010. "The Natural Landscape Metaphor in Information Visualization: The Role of Commonsense Geomorphology." *Journal of the American Society for Information Science and Technology* 61 (2): 253–270. doi:10.1002/asi.21227.

Freksa, C. 1999. "Spatial Aspects of Task-Specific Wayfinding Maps." In *Visual and Spatial Reasoning in Design*, edited by J. S. Gero and B. Tversky, 15–32.

Sydney: Key Centre of Design Computing and Cognition, University of Sydney.

French, J. W., R. B. Eksstrom, and L. A. Price. 1963. *Manual for Kit of Reference Tests for Cognitive Factors*. Princeton, NJ: Educational Testing Service.

Garlandini, S., and S. I. Fabrikant. 2009. "Evaluating the Effectiveness and Efficiency of Visual Variables for Geographic Information Visualization." In *Proceedings Spatial Information Theory: 9th International Conference, COSIT 2009* Lecture Notes in Computer Science 5736, edited by K. S. Hornsby, C. Claramunt, M. Denis, and G. Ligozat, 195–211. Berlin: Springer. http://link.springer.com/chapter/10.1007%2F978-3-642-03832-7_12.

Gregory, R. L. 1987. *The Oxford Companion to the Mind*. 2nd ed. Oxford: Oxford University Press.

Hegarty, M., M. S. Canham, and S. I. Fabrikant. 2010. "Thinking about the Weather: How Display Salience and Knowledge Affect Performance in a Graphic Inference Task." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36 (1): 37–53.doi:.10.1037/a0017683

Hegarty, M., D. R. Montello, A. E. Richardson, T. Ishikawa, and K. Lovelace. 2006. "Spatial Abilities at Different Scales: Individual Differences in Aptitude-Test Performance and Spatial-Layout Learning." *Intelligence* 34 (2): 151–176. doi:10.1016/j.intell.2005.09.005.

Hegarty, M., A. E. Richardson, K. Lovelacea, and I. Subbiaha. 2002. "Development of a Self-Report Measure of Environmental Spatial Ability." *Intelligence* 30 (5): 425–447. doi:10.1016/S0160-2896(02)00116-2.

Hegarty, M., H. S. Smallman, A. T. Stull, and M. S. Canham. 2009. "Naïve Cartography: How Intuitions about Display Configuration Can Hurt Performance." *Cartographica: The International Journal for Geographic Information and Geovisualization* 44 (3): 171–186. doi:10.3138/carto.44.3.171.

Hochmair, H. H. 2004. "Decision Support for Bicycle Route Planning In Urban Environments." In *Proceedings, 7th AGILE Conference on Geographic Information Science*, 697–706. Heraklion: Crete University Press.

Horn, R. E. 1998. *Visual Language*. Bainbridge Island, WA: MacroVU.

Huffmann, D. 2009. "A Technique for Encoding Elevation Changes Along a Route." *Cartographic Perspectives* 63: 83–86. doi:10.14714/CP63.165.

Irvankoski, K. M. 2012. "Visualisation of Elevation Information on Maps: An Eye Movement Study." Master's Thesis, University of Helsinki.

Kahneman, D. 2011. *Thinking, Fast and Slow*. New York: Macmillan.

Kurata, Y., and M. J. Egenhofer. 2008. "The Arrow-Semantics Interpreter." *Spatial Cognition & Computation* 8 (4): 306–332. doi:10.1080/13875860802148843.

Larkin, J. H., and H. Simon. 1987. "Why a Diagram Is (Sometimes) Worth Ten Thousand Words." *Cognitive Science* 11 (1): 65–100. doi:10.1111/j.1551-6708.1987.tb00863.x.

Levy, E., J. Zacks, B. Tversky, and D. J. Schiano. 1996. "Gratuitous Graphics? Putting Preference in Perspective." In *CHI '96 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, edited by M. J. Tabuer, 42–49. ACM: Washington, DC. doi:10.1145/238386.238400.

Lobben, A. K. 2004. "Tasks, Strategies, and Cognitive Processes Associated With Navigational Map Reading: A Review Perspective." *The Professional Geographer* 56 (2): 270–281.

Montello, D. R., K. L. Lovelace, R. G. Golledge, and C. M. Self. 1999. "Sex-Related Differences and Similarities in Geographic and Environmental Spatial Abilities." *Annals of the Association of American Geographers* 89 (3): 515–534. doi:10.1111/0004-5608.00160.

Muehrcke, P. C., and J. O. Muehrcke. 1992. *Map Use: Reading, Analysis and Interpretation*. 3rd ed. Madison, WI: JP Publications.

Opach, T., I. Gołębiowska, and S. I. Fabrikant. 2013. "How Do People View Multi-Component Animated Maps?" *The Cartographic Journal* 51 (4): 330–342. doi: 10.1179/1743277413Y.0000000049.

Rensink, R. A. 2009. "Attention : Change Blindness and Inattentional Blindness." In *Encyclopedia of Consciousness*. Vol. 1, edited by W. Banks, 47–59. New York: Elsevier.

Schulte-Mecklenbeck, M., A. Kühberger, and R. Ranyard. 2011. *A Handbook of Process Tracing Methods for Decision Research: A Critical Review and User's Guide*. New York: Psychology Press.

Slocum, T. A., R. B. McMaster, F. C. Kessler, and H. H. Howard. 2009. *Thematic Cartography and Geovisualization: International Edition*. 3rd ed. Upper Saddle River, NJ: Prentice Hall.

Smallman, H. S., and M. B. Cook. 2011. "Naïve Realism: Folk Fallacies in the Design and Use of Visual Displays." *Topics in Cognitive Science* 3 (3): 579–608. doi:10.1111/j.1756-8765.2010.01114.x.

Su, J. G., M. Winters, M. Nunes, and M. Brauer. 2010. "Designing a Route Planner to Facilitate and Promote Cycling in Metro Vancouver, Canada." *Transportation Research Part A: Policy and Practice* 44 (7): 495–505. doi:10.1016/j.tra.2010.03.015.

Sutula, D. P. 2007. Hiking Staff, Ski Pole, or Boat Paddle, with Integrated Topographical Representations of Trails and or Terrain. *US Patent 7261564*.

Tobler, W. R. 1987. "Experiments In Migration Mapping By Computer." *The American Cartographer* 14 (2): 155–163. doi:10.1559/152304087783875273.

Tversky, B. 2001. "Spatial Schemas in Depictions." In *Spatial Schemas and Abstract Thought*, edited by M. Gattis, 79–111. Cambridge, MA: MIT Press.

Wessel, N., and M. Widener. 2015. "Rethinking the Urban Bike Map for the 21st Century." *Cartographic Perspectives* (81): 6–22. doi:10.14714/CP81.1243.

Wilkening, J., and S. I. Fabrikant. 2011. "The Effect of Gender and Spatial Abilites on Map Use Preferences and Performance in Road Selection Tasks." 25th International Cartography Conference, International Cartographic Association, Paris, July 3–8. http://icaci.org/files/documents/ICC_proceedings/ICC2011/Oral%20Presentations%20PDF/C2-Mapping%20network%20and%20route/CO-232.pdf.

Wilkening, J., and S. I. Fabrikant. 2013. "How Users Interact with a 3D Geo-Browser under Time Pressure." *Cartography and Geographic Information Science* 40 (1): 40–52. doi:10.1080/15230406.2013.762140.

Woolfolk, M. E., W. Castellan, and C. I. Brooks. 1983. "Pepsi Versus Coke: Labels, Not Tastes, Prevail." *Psychological Reports* 52 (1): 185–186. doi:10.2466/pr0.1983.52.1.185.