This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3035823, IEEE Transactions on Visualization and Computer Graphics

Rainbow Dash: Intuitiveness, interpretability and memorability of the rainbow color scheme in visualization

Izabela M. Gołębiowska, Arzu Çöltekin

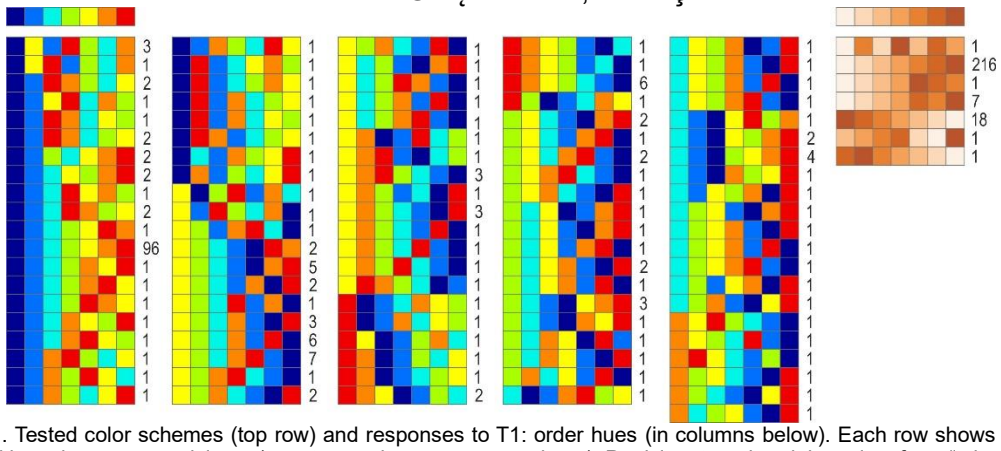

Figure 1. Tested color schemes (top row) and responses to T1: order hues (in columns below). Each row shows a unique sequence provided by at least one participant (counts are shown next to each set). Participants ordered the colors from "min to max" (left to right) purely based on their perceptual associations for the seven hues. We see 101 different orders for the rainbow and seven for the sequential schemes. The 'correct' sets are shown on top.

Abstract—After demonstrating that rainbow colors are still commonly used in scientific publications, we comparatively evaluate the rainbow and sequential color schemes on choropleth and isarithmic maps in an empirical user study with 544 participants to examine if a) people intuitively associate *order* for the colors in these schemes, b) they can successfully conduct perceptual and semantic map reading and recall tasks with quantitative data where order may have implicit or explicit importance. We find that there is little to no agreement in ordering of rainbow colors while sequential colors are indeed intuitively ordered by the participants with a strong *dark is more bias*. Sequential colors facilitate most quantitative map reading tasks better than the rainbow colors, whereas rainbow colors competitively facilitate extracting specific values from a map, and may support hue recall better than sequential. We thus contribute to *dark- vs. light is more bias* debate, and demonstrate why and when rainbow colors may impair performance, and add further nuance to our understanding of this highly popular, yet highly criticized color scheme.

Index Terms—Color, visualization, colormap, color perception, visual design

1 Introduction

SEMIOTIC importance of color has been acknowledged for decades, and it is well documented that color changes how viewers perceive and interpret the underlying data (*e.g.*, [1], [2], [3]). A number of empirical studies have been conducted to examine color use in visualization (*e.g.*, [4], [5]) and cartography (*e.g.*, [6], [7], [8], [9]). Besides these scientific efforts, post-Internet decades led to dozens of online channels about the importance of color, and how to use it correctly, as well as color recommender software (*e.g.*, [10]–[19]). Despite the plethora of information, empirical evidence, and tools, people violate recommended practices in color use. A particularly common 'offense' appears to be the use of the rainbow color scheme (RC) for presenting quantitative data. The RC, mimicking the rainbows in nature, is composed of spectrally ordered hues: blue, cyan, green, yellow, and red [20], [21]. While people enjoy rainbow colors [22], the RC has been shown to create problems in legibility and pattern detection when used in visualizing quantitative data [23], [24]. Arguably, this is at least partly because the global ordering of the hues in the RC is not intuitive. It has been argued by some authors that ordering colors locally (describing the relationship of a hue with its near neighbors) in selected parts of the RC might work without problems, though it is important to note that this statements are based on general reasoning or introspection, *i.e.*, without empirical evidence [5], [25], [26], [27]. As opposed to the RC, sequential color schemes (SC) are considered intuitive for ordering colors, as they cater to subconscious associations, such as the dark is more or the

- *I.M. Gołębiowska is with the Faculty of Geography and Regional Studies, University of Warsaw, ul. Krakowskie Przedmieście 30, 00-927 Warszawa, Poland. E-mail: i.golebiowska@uw.edu.pl.*
- *A. Çöltekin is with the Institute of Interactive Technologies, University of Applied Sciences and Arts Northwestern Switzerland, Bahnhofstrasse 5,5201, Brugg-Windisch. E-mail: arzu.coltekin@fhnw.ch.*

xxxx-xxxx/0x/$xx.00 © 200x IEEE Published by the IEEE Computer Society

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3035823, IEEE Transactions on Visualization and Computer Graphics

opposing light is more bias [28], [29]. Despite the relative intuitiveness of the SC, RC has been popular for a long time. More than two decades ago, Brewer [30] urged her readers that "...the public is learning this code" (p. 217) as more people used the RC in visualizations. Brewer's warning seems to have become a persistent challenge [31], as we see a surprisingly frequent use of the RC in practice also today (see *Section 3*). Despite the expert evidence discouraging the use of the RC, such as Borland and Taylor's seminal paper titled "Rainbow color map (still) considered harmful" [23], empirical studies that provide nuanced evidence on precise boundaries of how much RC hurts performance, and for which tasks, are rare. Thus, we conducted a user experiment comparing the RC to the SC for various tasks on two map types: Choropleth (Choro) and isarithmic (Isa). These map types can represent the same information, but Choro requires classification and represents discrete categories, while Isa uses continuous data. They also use color in different spatial configurations; Choro often does not have ordered colors on the map and shows the order in a legend, whereas Isa features colors always in the same order and the order is presented on the map itself. Comparing RC and SC with these two map types enables examining if the effect of the color scheme type remains stable across different representations. Also importantly, our review (see *Background*) suggests that whether the RC is 'good or bad' largely depends on task type: Value-varying schemes, such as the SC, might facilitate tasks that require ordering values by magnitude well (*e.g.*, interpreting general patterns, anomaly detection. comparing values); and hue-varying schemes, such as the RC, might be in general better for reading specific details. Therefore, we hypothesize that RC may lead to different degrees of performance loss depending on task type (for specific hypotheses, see *Study 2*) irrespective of map type. To test our hypotheses, we designed tasks that differed in cognitive processing requirements (see *Procedure*), and conducted a controlled study comparing the RC to the SC (Figure 1 top row) on the two map types (see *Experiment design*). Specifically, we answer the following five research questions (RQs – note that 1 and 2 split in two as A and B):

RQ1. A) Can participants intuitively order the RC? B) Is ordering affected by exposure to a color scheme?

RQ2. Do the effects of the RC remain stable across various A) map reading and B) visuospatial recall tasks, and across the two map types?

RQ3. Do participants like the RC more than the SC?

In addition to the RQs above, we explore how participants' self-declared task difficulty correlates with their task accuracy as an implicit measure of confidence. Confidence is interesting as it can signal "intuitiveness" or "obviousness" [32], [33]. We believe our study contributes to visualization research and practice in supporting design decisions where color plays a crucial role.

2 Background

The RC has received much criticism from experts for a long time for many reasons, e.g. [34], [35], [36], [37]. Nonetheless, this color scheme is often used, also in scientific publications. A detailed account of the critique and possible reasons why people still use the RC is presented in *Suppl. Mat., Sections 2.1.* and *2.2*. Borland and Taylor [23] estimated that 40-59% of the papers in IEEE Visualization conference proceedings between 2001-2005 used the RC. We examined the same question in planetary science and remote sensing journals more than 20 years later, and found even higher percentages in some cases (see *Section 3.2*).

2.1 User studies with the RC: Mixed evidence

Results from the studies examining the RC use in visualization do not always agree. In cartography, this may be partly due to different map types, *e.g.,* studies exist on Isa ([38], [39] after [40]), Choro ([39] after [40], [2], [40]), and other map types [21], [41]–[43]. Besides map types, task types appear to be important, *e.g.,* comparing areas [2] visible to the user might lead to different outcomes than tasks that rely on working from the memory [40]. The RC has been shown to impair user performance in some visuospatial tasks. For example, people cannot order spectral hues intuitively: Olson [44] has shown that no two participants produced the same order while creating a legend for a bivariate map. In other tasks, *e.g.,* estimating the degree of similarity between pairs of Isa maps [38], and for identifying high, medium and low data densities in Choro and Isa maps with no legends [39], the RC led to low task accuracy. This might be because of the perceptual uniformity issues mentioned earlier, specifically, because the central portion of the RC facilitates feature discrimination poorly [45], [46]. Also, in a relative color distance judgement task, the RC resulted in higher error rates and longer response times than the SC and a diverging scheme [47]. Discretization appears to be difficult with the RC too, *i.e.,* people were inconsistent in recognizing and placing boundaries between rainbow hue bands [42], further confirming that the RC is not perceptually uniform.

In contrast, some studies show that participants can achieve similar or higher visuospatial task accuracy with the RC than with its alternatives [2], [21], [40], [43], [48]. Ware [48] has shown that for extracting specific values, the RC leads to lower mean response error than the SC and divergent schemes. Reda *et al.* [21] further found that participants reached the highest accuracy with the RC compared to other alternatives in identifying locations of specified values in continuous color maps. In Reda *et al.*'s [21] study, RC resulted with one of the best accuracy scores for gradient (steepness) estimation. Brewer *et al.*'s [2] seminal study shows that participants, when retrieving values from Choro maps ('legend matching'), were significantly more accurate with the RC than a grayscale. On the other hand, participants' response accuracy was lower in the same task with the RC than with two diverging schemes [30]. Hyslop [41], too, demonstrated that participants achieve higher accuracy in map reading tasks with the RC than with a grayscale alternative. Using a modified version of the RC (yellow as the min value), Mersy [40] has shown that for reading specific details and for recalling particular values, RC users do not perform worse than SC users. Mersy [40] pos-

its that hue-varying color schemes with large contrasts, including RC, work well for reading specific details from maps, also citing Robinson's [49] explanation that the HVS is more sensitive to differences in hue than in value. Ware [48] and Mersy [40] both mention that the *simultaneous contrast effect* occurs less with hue than with value. Task type seems to be important: the RC did not facilitate tasks that require identifying general patterns well in either study [48],[40]. Mersy [40] notes that for reading and recall of general patterns on a map, and anomaly detection on Choro maps, the SC yields higher response accuracy than the RC. Recall tasks are especially interesting as people remember the colors that they can name better than colors they cannot [50], which could give the RC an advantage compared to the SC. In sum, the evidence for or against the RC is mixed and task type is important (see *Suppl. Mat., Section 2* for a systematic overview of the literature summarized above). Furthermore, people like the RC: Kumler and Groop [43] demonstrated that 73% of the participants preferred RC over other representations and Brewer *et al.*'s [2] participants rated the RC as the most 'pleasant' and 'easy to read' among eight schemes.

2.2 Original contributions

We provide original contributions on 1) the current prevalence of the RC in scientific publications, 2) intuitiveness of the RC *vs.* SC in color ordering tasks and a re-examination of the *dark is more* bias, and 3) a systematic comparison of the consequences of using the RC *vs.* the SC for various visuospatial tasks with two different map types, examining the robustness of their effects across varying conditions. Differing from previous work, we also evaluate participants' response speed (besides accuracy and preferences), which is a meaningful metric from an information processing perspective, and important for visualization use under time pressure.

3 Study 1: Prevalence of the RC

To document the *current* prevalence of the RC use in scientific publications, similarly to Brewer [30] and Borland and Taylor [23], we sampled scientific publications in familiar domains that use visualizations heavily, and examined if they use the RC, and for what kind of tasks.

3.1 Methods: Prevalence study

In two scientific fields with large audiences, *i.e.,* remote sensing and planetary science, we selected two journals, of wide outreach. For planetary science we browsed *Icarus* (IF 2019 3.51) and *Journal of Geophysical Research: Planets / JGR Planets* (IF 3.71), and for remote sensing, we searched *ISPRS Journal of Photogrammetry and Remote Sensing* (IF 7.32) and *Remote Sensing* (IF 4.51). We queried 355 papers published in late 2019 and early 2020, evenly spread across the two fields and journals. We marked if a paper contains at least one figure with the RC for representing *quantitative* data *i.e.,* we excluded the papers that do not contain figures representing quantitative data even if they used RC, as well as those with only black-and-white figures or photos. As a result, we retained 205 papers for further analysis.

3.2 Results: Prevalence study

Table 1 shows the prevalence of the RC in 205 papers for quantitative data visualizations: The RC is *still* very popular, with an average of 64% in planetary science and 48% in remote sensing. In one journal this value hits 70%, which clearly demonstrates that the RC is still a very common choice for visualization in the analysed disciplines. This share is remarkably higher than the 51% reported by Borland and Taylor [23] 13 years ago (though it is important to note that they analyzed IEEE Vis conference proceedings, i.e., our study is not an exact replication of theirs).

TABLE 1 Papers visualizing quantitative data with the RC

journal	# papers analyzed	% using the RC
Icarus	74	55
JGR Planets	76	70
Planetary Science discipline	150	64
ISPRS JPRS	80	46
Remote Sensing	75	50
Remote Sensing discipline	155	48

The RC is used in a variety of visualizations, *e.g.,* maps, plots, *etc.*, and for both raster and vector data. It is also used for a variety of topics, ranging from elevation, atmospheric characteristics (*e.g.,* wind stress); through physical indices (*e.g.,* gravity, mass) to specific indices (*e.g.,* motion correction, methane abundance) in planetary science, and camera-scene distance, estimation errors, vegetation mass change, etc. in remote sensing. Since this sampling verifies that RC is *still* popular, next we re-examine if, and in which tasks, it is (still) harmful [23].

4 Study 2: Comparing the RC to the SC

Based on the literature reviewed above, we formulate and test the following hypotheses:

H1) The RC will be overall inferior to the SC in color ordering tasks across the tested conditions (addressing RQ1A, see *Introduction* section)

H2) In map reading, RC will be superior for extracting specific values and SC for general pattern interpretation (addressing RQ2A)

H3) The RC will be a competitive alternative in recall tasks (RQ2B)

H4) Participants will rate the RC more likeable than the SC (RQ3)

H5) The expected effects (H1-H4) will persist across the two map types, irrespective of the task type (RQ2A&B).

Besides the above hypotheses, we *explore* two questions: 1) Would participants learn and remember the way the colors were used in the other experimental tasks *more* with the SC or the RC? (focusing on the RQ1B). We expected that they would, with both color scheme types, perhaps a little more with the RC due to nameable colors. 2) Would participants' perceived task difficulty match the actual task difficulty? We expected that they like the RC better, and they may be misguided by this and believe they did better with RC than the SC.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3035823, IEEE Transactions on Visualization and Computer Graphics

4.1 Experiment design

Using a mixed factorial design, we compared the RC *vs.* SC on two map types with various tasks (see *Procedure*) in a controlled experiment. Our independent variables are *color scheme* (RC *vs.* SC, main variable), *map type* (Choro *vs.* Isa) and *task type* (ordering, map reading, recall). Map and color scheme types were treated as between-subject variables, while task type was treated as a within-subject factor. As dependent variables, we measured two performance metrics *effectiveness* (response accuracy) and *efficiency* (response time) and two subjective metrics, *i.e.,* participants' rating of *task difficulty,* and *likeability* of the RC and SC.

4.2 Participants

Because maps and atlases are commonly used in education [51], we decided to work with high school students. This age group is old enough to have some map-use experience, but not yet biased by professional expertise. To recruit participants, we contacted the principals of 22 high schools which were selected randomly all over Poland. Participation was voluntary, no compensation was offered, and consent was obtained from the principals. Ethical approval was obtained from the ethics board of the Faculty of Geography and Regional Studies, University of Warsaw. Total 544 participants (50.9% female (f), 15-20 years-old) took part in the study. All of them completed courses covering geography and map reading. 16.2% declared that they use maps daily, 30.1% once a week, 20.7% several times a month, 18.5% once a month. 5.7% once a year and 7.8% less than once a year. None of them self-reported color vision deficiency, though in this age group, it is possible that they were not yet tested for color vision impairment, which typically is administered along with driving license tests. Since color vision deficiency can be up to 10% in men [37], we later conducted an outlier analysis to catch the possible effects of 'hidden' color deficiency issues. The participants were divided into four groups (n=140, 131, 134 and 139) counterbalanced for biological sex. Each group solved the same tasks using four maps (Figure 2). At the analysis stage, we excluded participants who did not complete all tasks and who provided answers in less than 3 seconds on average, as this suggests they did not read the task, and their response may have been random. We also excluded outliers, *i.e.,* participants whose average response accuracy and time over all tasks was more than +/-3SD away from the mean. We ended up analyzing the answers collected from 534 participants, in four groups: RC-Choro (n=139, 51.1% f), RC-Isa (n=127 49.6% f), SC-Choro (n=132, 53.0% f), SC-Isa (n=136, 50.7% f).

4.3 Apparatus and materials

We applied two 7-class color schemes: (1) RC (Figure 2a, 2b) with fully saturated hues, and (2) SC using Color Brewer's '7 class oranges' (Figure 2c, 2d) checked with Color Oracle [18] (see Figure 1 top row and for specifications of the hues see *Suppl. Mat., Table 2*). When selecting the scheme we decided to test colors that are 'neutral', *i.e.,* not clearly associated with any phenomenon (through custom or natural associations), *e.g.,* blue for hydrography, red for dangers, green for nature etc. We also excluded diverging schemes, since they are more complicated for ordering tasks, and are recommended only for a specific set of data with a critical value within the range [52]. Since having a critical value to diverge from is a special case, we decided to keep our data and stimuli for a 'more general' case of linear variation (the RC is applied frequently also for data with no such 'critical value'). Seven classes are the recommended maximum in cartography, guided by Miller's seminal "magical number seven plus minus two" study [53]. To keep seven classes also in the RC for comparability, we used the spectral order minus the violet hues. We assigned higher values at the red end, as found on other maps that use the RC. We decided to use a white background, and *dark is more* order for the SC, since these reflect the current common practice.

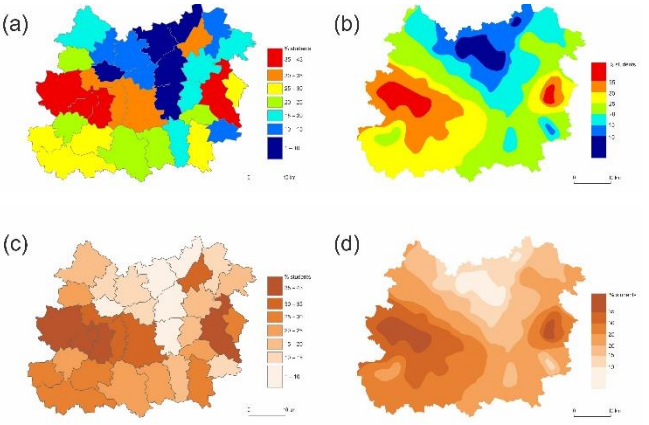

Figure 2. The stimuli: Rainbow (a: RC-Choro, b: RC-Isa) and sequential schemes (c: SC-Choro, d: SC-Isa) on two map types.

We then made four maps using the RC and the SC, both showing 'percentage of high school students working paid jobs' as we assumed this would be relatable for students. Reference data was sourced from the national database, whereas thematic data was fictitious. We used ArcMap and CorelDraw to create the maps, MS Excel and SPSS for data analysis. The resulting four maps are shown in Figure 2: Choro (2a,c) and Isa (2b,d) maps. We selected the area (a remote region of Poland) so that it would not be commonly known. We verified this by asking a few expert geographers to identify it (they could not). We presented the tasks in Google Chrome under constant lighting and procedural conditions.

4.4 Procedure

We administered the study in all 22 schools' computer rooms. We first briefed the students and instructed them to work on their own without interacting with each other. Students could ask questions during the briefing but not during the experiment. Participants started the tasks at the same time in each group, and solved them without a time limit. Tasks were presented digitally (see Table 2 for precise wording and *Suppl. Mat. Table 3* for further information on the tasks). We designed the tasks largely consistent with the literature (*e.g.*, [2], [21], [40], [43], [54]).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3035823, IEEE Transactions on Visualization and Computer Graphics

TABLE 2. TASKS IN THE STUDY

No	Task instruction	RQ
T1	Order hues starting from the one that should symbolize min values and finishing with hues symbolizing max values.	RQ1A RQ1B
T2	Select an example unit representing the **max** values (map presented without legend).	RQ1A
T3	Select an example unit representing the **min** values (map presented without legend).	RQ1A
T4	Which of the marked units (A or B) features the higher value of the presented phenomenon?	RQ1A
T5	What is the value range in the marked unit?	RQ2A
T6	Select an example unit featuring value range 10-15%	RQ2A
T7	Which profile (A/B/C) shows the values correctly along the marked black line?	RQ2A
T8	Which region (A/B/C) features the lowest average values?	RQ2A
T9	On the black-and-white map, all units that fall into a particular color category (on the color map you have just studied) are marked in black. Using the legend below, match the color with the marked units.	RQ2B
T10	On the black-and-white map, all units that fall into a particular color category (on the color map you have just studied) are marked in black. Indicate which value range is presented with marked units.	RQ2B
T11	Order hues starting from the one that should symbolize min values and finishing with hues symbolizing max values.	RQ1B

Our task design also partially matches Knapp's taxonomy of visual operators *locate, identify, compare and rank, associate* [55], and we added *recall* which fits with Çöltekin's modifications [54]. The majority of the resulting 11 tasks were about map reading (value retrieval, interpretation, recall) whereas two (T1 and T11) required perceptually ordering seven hues (Figure 1). Tasks were presented in the order shown in Table 2, and each task was solved only once. We opted not to randomize the task order to prevent participants from seeing legend-based tasks before the tasks that are not assisted with a legend. If they had seen the tasks with legend, this would be a threat to the experiment's internal validity. We later verified that there were no order effects. Tasks T2-T10 were accompanied by maps placed next to the instructions (see *Suppl. Mat. Figure 1*). Tasks T2-T4 also involve ordering colors, *i.e.*, for these tasks, maps do not include a legend, so participants needed to rely on their perceptual judgement alone to solve them. T5 and T6 require extracting specific values, based on visual search and comparison on maps with legends. T7 requires examining a line that crosses the map and matching the values of the colors across this line to the "profile" which shows the same data as a plot (*Suppl. Mat., Figure 1a*). T8 is a prototypical thematic map use task: Participants selected the region with the *lowest values* among three options. T9-T10 measure information recall based on hue and value. Participants were instructed to try to remember as much as possible from a map they viewed for 15 seconds. Then they matched some regions with a legend from memory, using a black and white map (see *Suppl. Mat., Figure 1b*). After each task, participants subjectively rated task difficulty. After all tasks were completed, likeability of the color schemes (T12-T15) on the two map types (T12 RC-Isa; T13 SC-Choro; T14 RC-Choro; T15 SC-Isa) using a 5-point Likert scale, addressing RQ3. The tasks T1-T4 and T11 refer to intuitive order (participants did not have access to legends), whereas T5-T10 are legend-based.

4.5 Results: Performance and preference study

We first report the main effects of color scheme and map type at aggregate level, then detailed analyses at the task level; based on task types (always linking it to color scheme and map type), and based on individual tasks.

4.5.1 Main effects of color scheme and map type

Color scheme (irrespective of map type, for all tasks) has statistically significant effects both on response accuracy and response time. On average, with the SC, participants are both more accurate (t=-7.74; p<.001; RC M=64.9, SD=21.8, SC M=77.1, SD=13.6); and faster (t=3.89; p<.001; RC: M=25.4 sec, SD=6.79 sec, SC: M=23.3sec, SD=5.90 sec) than with the RC. Map type (irrespective color scheme, for all tasks) does not lead to statistical significant differences at the aggregate level neither for response accuracy (t=-1.15, p>.05, Choro: M=70.1, SD=18.8, Isa: M=72.0, SD=19.5) nor for response time (t=.146, p>.05, Choro: M=24.4 sec, SD=7.15 sec, Isa: M=24.3 sec, SD=5.63sec).

4.5.2 Task-level analyses

We organized the task-level analyses as: RQ1) Color ordering, RQ1A on an abstract display (T1) and on a map without legend (T2, 3, 4), RQ1B learning (T11 *vs.* T1), RQ2) RQ2A (T5, 6, 7, 8) map reading, RQ2B (T9, 10) recall hue and value, and RQ3) likeability ratings (T12-15). The RQs and related sub-questions are treated as subsections after we provide an overview of the findings at aggregate level.

Aggregate task analysis

Because T1 & 11 are not map related, and T12-15 are subjective ratings, we treat them later. The aggregate results for the rest of the tasks for RQ1A (ordering), RQ2A (map reading), and RQ2B (recall) are shown in Figure 3.

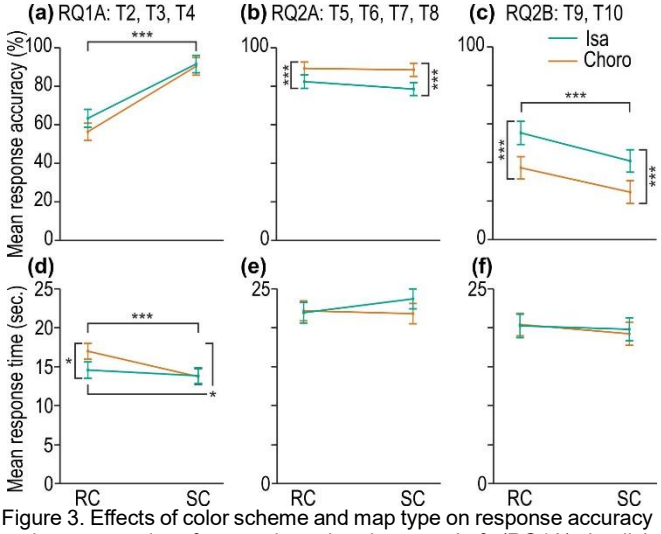

Figure 3. Effects of color scheme and map type on response accuracy and response time for map-based task types. Left (RQ1A): Implicit color ordering tasks on maps, Middle (RQ2A): Map reading tasks with a legend, Right (RQ2B): Recall tasks. Color scheme and map type interact in (d). *p < .05, ***p ≤ .001. Error bars: ±2SEM.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3035823, IEEE Transactions on Visualization and Computer Graphics

Figure 3 (left) shows that participants perform better with the SC than with the RC in color ordering tasks on maps without a legend both in terms of response accuracy ($F=158$ $p<.001$, $\eta^2_p=.230$) and speed ($F=13.6$, $p<.001$, $\eta^2_p=.025$). Map type does not matter for accuracy in these tasks, but participants are slower with the RC than the SC ($F=13.6$, $p<.001$, $\eta^2_p=.025$), more so with the Choro map than with Isa ($F=4.55$, $p<.05$, $\eta^2_p=.009$). Thus, map type interacts with color scheme ($F=6.46$, $p<.05$, $\eta^2_p=.012$) in terms of response speed. Figure 3 (middle) shows that participants' accuracy ($F=1.69$, $p=.194$, $\eta^2_p=.003$), or time ($F=1.66$, $p=.197$, $\eta^2_p=.003$) do not differ based on color scheme. However, their overall accuracy with the Choro is higher than with the Isa ($F=24.6$, $p<.001$, $\eta^2_p=.044$), irrespective of the color scheme. Figure 3 (right) demonstrates that, first of all, the recall tasks are overall harder than the others. Participants have a higher response accuracy with the RC than with the SC ($F=21.1$, $p<.001$, $\eta^2_p=.038$), and with the Isa than with the Choro ($F=36.1$, $p<.001$, $\eta^2_p=.064$) in recall tasks. Neither color scheme nor map type affect the response speed for aggregated recall tasks.

Color ordering tasks (RQ1)

Responding to RQ1A and 1B we report response accuracy and time (Figure 4) for color ordering tasks (see detailed statistics in *Suppl. Mat., Section 4*). We first present an overview of T1-4 and T11, then analyze individual tasks. Note that there is not an *objectively* accurate order for the RC, while for the SC, one can argue for two, based on *dark-* or *light is more* bias [28], [29]. When we say 'accuracy' in this section, we mean consistency of responses with our legend. We assigned the order for the RC based on its common use of it in maps. We also present agreement among participants as a measure of intuitive color ordering later.

Figure 4a shows that for *every* color ordering task, the SC led to higher response accuracy than the RC, and participants needed less time in five out of eight trials, and never took longer to order colors with the SC than with the RC (Figure 4b). Chi-square goodness of fit (for response accuracy) and Mann–Whitney U tests (for response time), reveal that all differences in participants' accuracy and response time based on *color scheme* are statistically significant (Figure 4, for more statistical detail including effect sizes see *Suppl. Mat., Section 4*). While *color scheme* leads to consistent differences in accuracy in color ordering tasks, *map type* does not: We only see a difference in the two tasks which require identification of extreme values (T2 and T3, max/min). In T2, when selecting the max value, Isa group has a higher accuracy than the Choro group for either color scheme (RC: $\chi^2=3.97$, $\varphi=.122$, $p=.046$; SC: $\chi^2=17.1$, $\varphi=.252$, $p=.000$). When selecting the min value, the Isa group has lower accuracy than the Choro ($\chi^2=19.8$, $\varphi=.273$, $p=.000$), but only with the SC. A Mann–Whitney U test revealed that for map-based tasks T2-4, participants were slower with the RC *only* in the Choro group. In the Isa group, color scheme did not matter (Figure 4b). In tasks T2 and T4, the Isa group was faster than the Choro group while working with the RC (T2: $U=6844$, $r=.194$, $p=.002$; T4: $U=7188$, $r=.161$, $p=.009$). Next, we present task specific observations in two categories 1) ordering hues (T1 and T11) and 2) implicit color ordering on a map without a legend (T2-4).

T1 and T11 With the RC, responses to T1 are 38% consistent with our map legend (Figure 4a, *Suppl. Mat., Table 4*). While solving T1, participants have not seen the maps, thus this shows participants' intuition *vs.* common practice in the RC maps. Remaining solutions exhibit great variability with the ordering of the RC (Figure 1, *Suppl. Mat., Figure 2*). After excluding incomplete responses where participants did not use all seven hues (9.63% for the RC, 8.43% for the SC), we found 101 unique sets with the RC (Figure 1), where 77 of them were by one participant only. 53.3% of participants started with dark blue as the min value, and 20.7% with yellow. Otherwise, participants placed six out of seven individual hues for either min or max values at least once. With the SC, responses are much 'tidier': 81.7% match with the map legend (Figure 4a, also see *Suppl. Mat., Table 4*), and only 7 different solutions (Figure 1, *Suppl. Mat., Figure 2*). With the SC, in T1 91.4% assume *dark is more*, and 7.8% *light is more*.

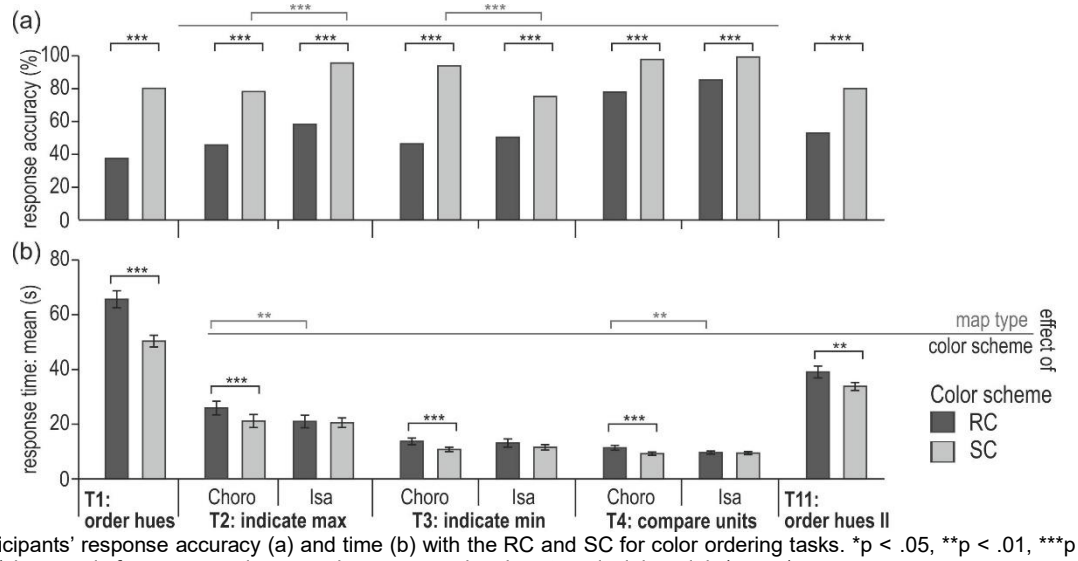

Figure 4. Participants' response accuracy (a) and time (b) with the RC and SC for color ordering tasks. *p < .05, **p < .01, ***p ≤ .001. Error bars: ±2SEM (shown only for response times, as the accuracy data is categorical, i.e., right/wrong).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3035823, IEEE Transactions on Visualization and Computer Graphics

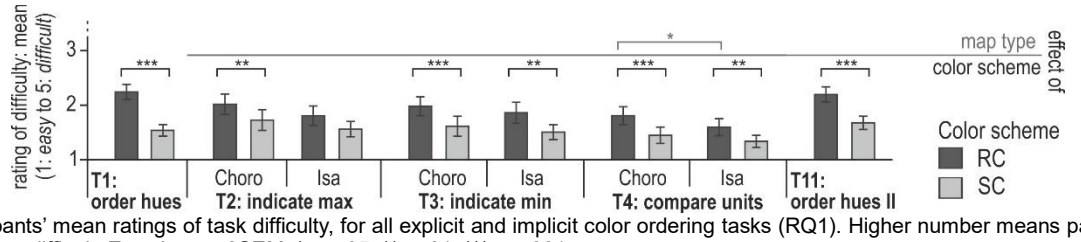

Figure 5. Participants' mean ratings of task difficulty, for all explicit and implicit color ordering tasks (RQ1). Higher number means participants rated the task more difficult. Error bars: ±2SEM, *p < .05, ** < .01, ***p ≤ .001.

Responses to T11, *after* having worked with the maps, were similar for the SC: 80.2% consistent with the legend. For the RC, answer accuracy increased to 53.4%. The variability in the ordered sets were smaller for the RC in T11 too; 53 unique sets, 42 by one participant only; while it remained the same for the SC (statistical detail in *Suppl. Mat., Section 4*). A McNemar's matched pair test comparing T1 (before) and T11 (after) revealed that participants' answer accuracy increased for the RC (χ^2=21.3, φ=.469, p=.000). For the SC this difference is not statistically significant (χ^2=.145, φ=.267, p=.703). For both conditions, according to a Wilcoxon signed-ranks test for related pairs, there is a speed gain from T1 to T11: Participants were 26.4s faster with the RC (Z=12.9, r=.793, p=.000), 16.5s with the SC (Z=12.4, r=.756, p=.000) after the experiment (Figure 4b).

T2, T3 and T4 T2 and T3 require implicit color ordering, *i.e.,* a series of comparisons on a map without a legend to identify the min and max values in a set, based on participants' preconceived perceptual associations. Both tasks were very difficult with the RC: Only about half the participants 'correctly' indicated hues representing max and min values (as we assigned them) with the RC (Figure 4a). We also analyzed the frequency of hues that were selected as max and min values. For tasks T2 and T3, demonstrates some variability in the responses with the RC, in fact, each color is selected at least once, whereas we see a clear agreement among the participants with the SC. With the RC, bright red appears to be strongly associated with the max value, which is in accordance with our legend. Dark blue is the second most frequently picked color for the max value, which is the opposite of what we assumed. It appears that participants associated the brightest hues (cyan, green and especially yellow) with *min values*. These patterns were similar for the two map types (Choro and Isa). With the SC, variability is near zero; lighter colors are associated with less and darker with more. There are only few, seemingly negligible deviations in sorting the SC by perceptual association (further details in *Suppl. Mat., Section 4*). T4 required pairwise-comparisons, for which the results are clearly in favor of the SC (Figure 4).

Task difficulty: Color ordering tasks. Participants' subjective ratings of task difficulty for color ordering tasks based on a 5-point Likert scale are presented in Figure 5. Participants rated the color ordering tasks overall as 'easy' (rarely higher than two out of five), whereas they rated the RC as 'more difficult' in all. A Mann–Whitney U test revealed that these differences were statistically significant in all tasks except T2 with Isa (Figure 5). Map type overall does not affect the difficulty ratings in color ordering tasks (RQ1) except for T4 (U=7677, r=.125, p<.05) where participants in the RC group marked that it was harder to work with the Choro map than with the Isa.

Map reading and recall tasks (RQ2)

We present the tasks that we study under RQ2 in two categories 1) *Map reading* (T5-8) and 2) *Recall* tasks (T9-10).

T5-T8 Map reading tasks T5-8 varied in levels of complexity where participants always had access to a legend, resulting in more than 80% response accuracy with one exception (Figure 6a): T7 was difficult for the Isa group with either color scheme (58% for both).

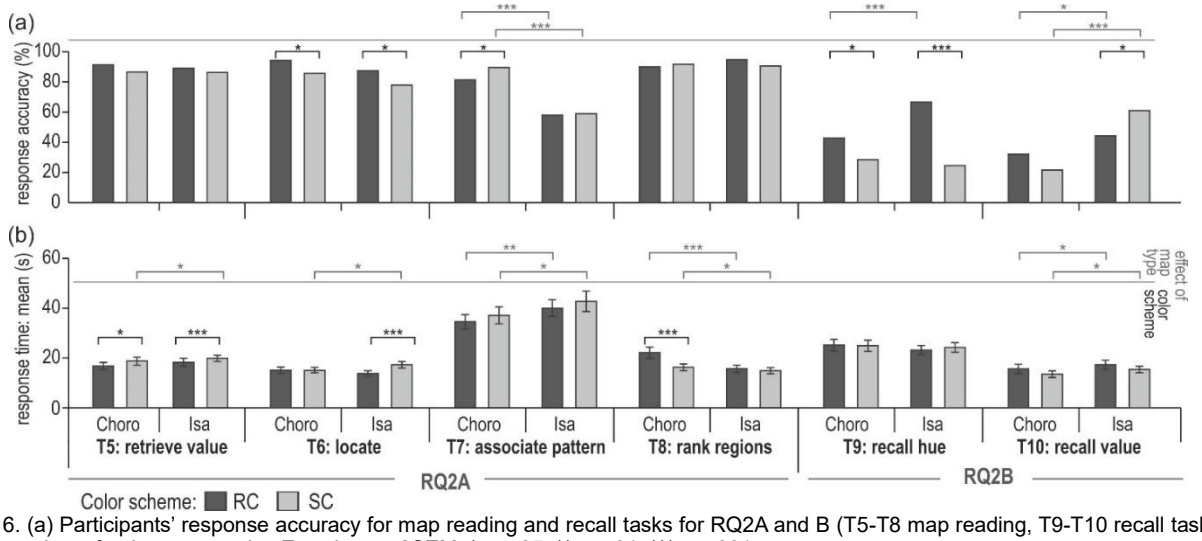

Figure 6. (a) Participants' response accuracy for map reading and recall tasks for RQ2A and B (T5-T8 map reading, T9-T10 recall tasks). (b): Response times for the same tasks. Error bars: ±2SEM, *p < .05, **p < .01, ***p ≤ .001.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3035823, IEEE Transactions on Visualization and Computer Graphics

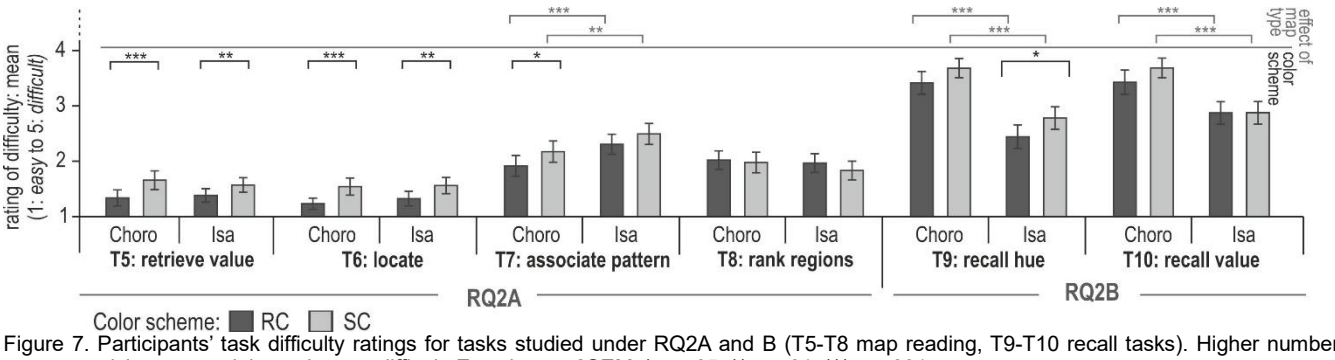

Figure 7. Participants' task difficulty ratings for tasks studied under RQ2A and B (T5-T8 map reading, T9-T10 recall tasks). Higher number means participants rated the task more difficult. Error bars: ±2SEM, *p < .05, **p < .01, ***p ≤ .001.

Participants did well in tasks that require reading specific map details (T5-6) irrespective of map type. The RC appears to work better than the SC for T6 with either map type. For the two tasks that require interpretation of general information (T7-8), participants' overall accuracy is not affected by color scheme, except in the Choro group where the SC was superior to the RC. In T7, map type has an effect with both color schemes: Choro group has higher response accuracy than the Isa in this task (RC: $\chi 2$=16.9, φ=.252, p=.000; SC: $\chi 2$=35.7, φ=.365, p=.000). Response time analysis based on color scheme yielded statistically significant differences for T5 for both map types, for T6 for Isa, and for T8 for Choro. For tasks that require extracting details from the map (T5-T6), participants were faster with the RC than with the SC, although for T5 accuracy did not differ (Statistical details provided in *Suppl. Mat., Section 4*).

Map type affects response time as well (Figure 6b): Participants took longer with the SC with Isa than with Choro for T5 and T6 (T5: U=7497, r=.142, p<.05; T6: U=7418, r=.150, p<.05), even though this is not reflected in their response accuracies. With T7, map type matters on both counts: Participants were faster with the Choro map than with the Isa using either color scheme (RC: U=7108, r=.168, p<.01; SC: U=7637, r=.129, p<.05) and they were also more accurate with it (Figure 6a). For T8, which requires interpreting a more general pattern, participants in the Choro group took longer to respond with the RC than with the SC. Participants in the Choro group are overall faster in T8 than those in the Isa group (RC: U=5632, r=.313, p<.001, SC: U=7673, r=.125, p<.05).

T9-T10 The recall tasks resulted in lower response accuracy than map reading tasks (Figure 6a). However, for these tasks, the RC overall facilitated higher response accuracy than in other tasks. In T9, participants' accuracy was 67.7% with the RC with Isa maps, but remained at 23.5% with the SC. The pattern is persistent with the Choro maps where RC facilitates 42.4% accuracy, and SC only 28.8%. The differences in accuracy with different color schemes in task T9 are statistically significant for both map types. Participants' response accuracy is still considerably higher with the RC (32.4%) than with the SC (22%) with Choro maps, but this difference is statistically not significant (p=.055). With Isa, the effect is reversed for T10; participants' recall accuracy is higher with the SC (59.6%) than with the RC (44.9%), and the difference between the two is statistically significant. When we focus on map types, we see that in T9, participants in Isa group responded more accurately than in Choro group (χ^2=17.1, φ=.253, p=.000) with the RC. Isa facilitates higher accuracy also for T10, but for both color schemes (RC: χ^2=4.39, φ=.128, p=.036, SC: χ^2=39.1, φ=.382, p=.000). Response time analysis for T9 and T10 (Figure 6b) did not reveal an effect of color scheme, but for T10 they are somewhat faster with the Choro maps (RC: U=7224, r=.157, p<.05; SC: U=7552, r=.137, p<.05). The inferential analyses of the effects of color scheme, including effect sizes, can be seen in *Suppl. Mat., Section 4*.

Task difficulty: Map reading and recall tasks. T5-6 were rated as the easiest among map reading and recall tasks (Figure 7). The difficulty rating is somewhat higher for T7 and T8, though with T8 we do not see differences based on color scheme or map type. The two recall tasks T9-T10, on the other hand, were overall rated as 'difficult' compared to the other tasks (and they were difficult, given the response accuracy and response time data). Color scheme did not matter in T9-10 difficulty ratings, though map type had a peculiar effect: Over 50% of participants of Choro rated T9 and T10 as 'difficult' or 'rather difficult'. Participants working with the Isa rated these tasks 'easy', in fact, less than 15% of the participants marked the tasks as 'difficult'/'rather difficult' irrespective of color scheme. A Mann–Whitney U tests revealed statistically significant differences between participants' rating of the RC and the SC in reading details from the map (T5 and T6) for both map types, and a pattern analysis task (T7) for Choro (detailed in *Supp. Mat., Section 4*). The recall tasks were rated similarly difficult with the exception of T9 for Isa, for which the RC was rated easier than the SC. Map type (Isa *vs.* Choro) comparison revealed an effect on T7 (RC: U=6697, r=.218, p<.001, and SC: U=7367. r=.162, p<.01), T9 (RC: U=5140, r=.367, p<.001, and SC: U=5196, r=.377, p<.001) as well as T10 (RC: U=6512, r=.231, p<.001, and SC: U=5450, r=.351, p<.001).

Measured *vs.* perceived task difficulty

Comparing the participants' performance with their task difficulty ratings gives us an indirect measure of confidence. Earlier work on naïve cartography and realism suggest that people might not always 'know' what works well [32], [33]. In this study we find a general alignment in actual task difficulty based on performance data, and task difficulty ratings (Figure 8). While response accuracy increases, rated difficulty decreases (Figure 8 left), *e.g.* plotting response accuracy *vs.* difficulty (Figure 8 right) also

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3035823, IEEE Transactions on Visualization and Computer Graphics

reveals this general trend, even though the correlation between the two variables is not statistically significant (Spearman's $r(11)=-.536$, $p=.089$).

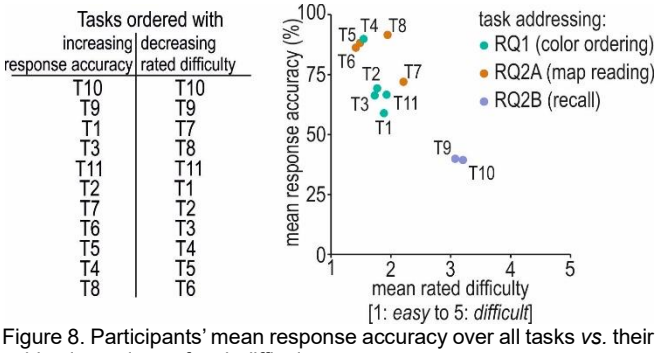

Figure 8. Participants' mean response accuracy over all tasks *vs.* their subjective ratings of task difficulty.

Subjective color scheme likeability (RQ3)

We examined *likeability* of the color schemes in all four visualization conditions (Figure 9) on a 5-point Likert scale (1: 'I don't like it', 5: 'I like it'). Almost half (47%) of the participants marked 1 or 2 for the RC-Choro (shown in *Supp. Mat., Section 4*). For the RC-Isa, SC-Choro and SC-Isa, more than 60% of the participants marked 4 and 5 on the scale. A Wilcoxon matched pairs signed-ranks test shows that participants like the SC-Choro more than the RC-Choro ($Z=7.97$, $r=.345$, $p<.001$), whereas for the Isa, we see no difference ($Z=1.36$, $r=.059$, $p>.05$) (Figure 9). Isa-RC are rated more likeable than Choro-RC ($Z=8.63$, $r=.373$, $p<.001$), while with the SC, likeability of the map types does not differ ($Z=.473$, $r=.020$, $p>.05$).

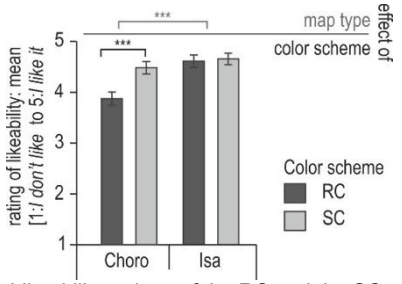

Figure 9. Likeability ratings of the RC and the SC with Choro and Isa. Error bars: ±2SEM, ***p ≤ .001.

5 Discussion

To understand the effects of using rainbow colors in visualizations with a greater nuance than a binary "good or bad" statement, we conducted a large study (n=534) examining a series of visuospatial tasks on two map types. We posed three RQs (see *Introduction*) on participants' ability to intuitively order the RC hues for the magnitude they represent (RQ1A); implicit learnability of this order from the maps (RQ1B); effects of the RC when solving map reading (RQ2A) and recall tasks (RQ2B), and likeability of it across two map types (RQ3), and five related hypotheses (see *Section 4)*. As a baseline, we obtained the same measures using the SC in all tasks.

For color ordering tasks (RQ1), our observations regarding the color schemes are very clear. In line with the consensus in literature, our participants ordered the SC with over 90% agreement in T1 (*Supp. Mat., Figure 2*), before they knew about the tasks or seen the maps. This may be interpreted as *intuition* or subconsciously learned association. Furthermore, participants exhibited a clear *dark is more* bias in T1 (91.8%), and even stronger in T11 (96.2%) with the SC. In contrast, with the RC, variation is high (101 different responses) and it is difficult to argue for a specific bias. Many participants picked dark blue as the min value with the RC (T1 53.3%, T11 81%). Perhaps this is partially explained by proposition that there is a 'natural order' of hues from cold (blue) to warm (red) [56], as it is found in rainbows in nature. This cold-to-warm order is also likely learned from temperature maps which use dark blue as min value (and for our T11, from our own legend too). While the dominant "dark blue as min value" answer suggests that the *dark is more* bias breaks down with the RC, we find it notable that 20.7% of the participants choose yellow (the lightest hue in the set) as the min value in T1 and most of them placed darker hues as *more* in these sets. This choice goes against the spectral order, in which yellow would be in the middle, suggesting that for some participants *dark is more* association persists despite having used the RC in the previous tasks, and/or having seen actual rainbows in the nature or being exposed to spectral order in physics classes. Other variations, while interesting, are not repeated often enough to interpret as a pattern. With T2-4, more implicit, map-based color ordering tasks, the SC is also consistently superior to the RC (Figure 4). Changes from T1 to T11 (RQ1B) are negligible for the SC (consistency with the map legend went from 81.7% to 80.2%), though we observe some learning for the RC (38% to 53.4%, *i.e.,* 15.4% increase) and number of different sets is nearly halved (now 53 instead of 101). From T1 and T11, not surprisingly, both groups got faster, but markedly more with the RC (SC: 24.4-17.7=6.7s gain, RC: 65.3-38.9=26.4s gain). This is interesting but provides limited support for the use of the RC, *i.e.,* even after some (implicit) learning in our study, or 'learning the code' from other graphics and media [30], many still do not associate the RC colors with a specific order. In fact, our results are "worse for the RC" than what Olson [44] and Cuff [39] reported decades ago. Olson [44] asked participants to order nine hues in a complex two-variable choropleth map legend. We asked seven hues to be ordered for one variable, yet this task is still clearly difficult ("unintuitive") with the RC. In sum, the answer to RQ1A is that rainbow hues do not have an intuitive order, whereas sequential hues do, with a strong *dark is more* bias. The answer to RQ1B suggessts that there is considerable learnability for the RC. However, in this study we did not test *how long* this learning effect persists, or how it may progress with longer exposure to the RC. For example, scientists in the fields discussed in *Section 3* may have had a longer exposure to the RC scheme and it may have an even stronger, more persistent effect on its intuitiveness, readability and memorability. Taken together, as we hypothesized (H1) *the RC is overall inferior to the SC in color ordering tasks* across varying conditions, providing clear support to the critical voices on hue-varying schemes (*e.g.,* the RC) for representing magnitude, as formulated by many classics (*e.g.,* [1], [57], [36]).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3035823, IEEE Transactions on Visualization and Computer Graphics

Next, we explored the effect of the RC on map reading (RQ2A) and recall (RQ2B) tasks. Map reading is a complex task and can require different modes of information processing, *e.g.,* extracting specific details or interpreting overall patterns. In some of these tasks, the RC can be competitive, even superior. For example, in T5: *retrieve value* in our study, participants' response accuracy did not differ, but they were faster with the RC than the SC. Also in favor of the RC, T6:*locate* results in higher accuracy with the RC than the SC, and higher (with Isa) or comparable (with Choro) speed. These findings (T5-6) might be because large contrasts between the rainbow hues facilitate extracting specific map details well [2], [21], [40]. However, for the tasks concerning more general information processing (T7: *associate pattern* and T8: *rank regions*), the effect of the color scheme is not clear: The SC appears favorable in T7 for one map type (Choro), while there are no differences in T8. Thus, we can only partially accept H2: *In map reading, RC will be superior for extracting specific values and SC for general pattern interpretation*. The data only supports RC's superiority for extracting specific values. In contrast, our findings regarding general pattern interpretation broadly suggest that the RC can compete with the SC.

Recalling (or not being able to recall) information from a map can have important implications from an applied perspective (i.e., in education, journalism, wayfinding), as well as in studying spatial cognition. With this in mind, we examined the RC's effect on recall across our experimental conditions (RQ2B, T9 and 10). Overall performance was low with this task (Figure 4, *Suppl. Mat. Tables 7 & 8*) compared to other map tasks. However, for T9: *recall hue*, the RC led to higher accuracy than SC with both map types (differences are 44% with Isa, and 13.6% with Choro). We speculated that this may happen as it has been shown that that nameable colors are easier for people to remember [50]. With T10: *recall value*, where colors are much harder to name, the results shifted. We observed no differences based on color scheme for the Choro, while with the Isa, the SC was superior. This is possibly explained by the fact that, with the SC-Isa, values are ordered on the map and on the legend similarly while this is not the case with the Choro (Figure 2). We can thus only partially accept H3: *The RC will be a competitive alternative in recall tasks*, *i.e.,* for recalling hues. When we take the results of all map reading and recall tasks together (RQ2) the answer to whether people are effective and efficient with the RC in map tasks is "it depends": For some tasks (extracting specific details, recall hue) it works well, but for some tasks (general pattern interpretation, recall value) the SC works better. Among the 11 tasks (T1-11), the RC was superior to SC in two (T6: *locate*, T9: *recall value*) across conditions, was competitive in two (T5: retreive value, T8: rank regions) where participants' response accuracy was comparable but they were faster with the RC. In the remaining tasks, the SC was superior to the RC, with two exceptions where there was no effect based on map type.

Map types were included in the study to mainly examine the robustness of the effect of the color scheme. In four out of nine tasks, the observed effects hold for both map types (T4, 5, 6, 8), while for the others we have mixed results (T2, 3, 7, 9, 10). In map-based ordering tasks, map type matters for locating min/max values, but only with the SC (T2 and 3, Figure 4). For some map reading and recall tasks (T7, T9 and 10) we see differences based on map type, where for T7 Choro appears to facilitate better, and for T9 and 10, Isa, although it is difficult to argue for a pattern. Choro and Isa can represent the same information, while a key difference between them is that Isa contains an order within the map which is comparable to the legend while Choro does not. Part of the differences might be explained by *e.g.,* possible difficulty in distinguishing colors in lighter shades of the SC, or a hidden simultaneous contrast effect based on surrounding colors. In sum, against our prediction, the effect of map type can be stronger than the effect of the color scheme in map-related tasks as demonstrated in a fair number of cases in our study. Thus, we reject our H5 *The expected effects will persist across map types, irrespective of task type*. When formulating opinions for or against the RC, one should consider the *context of use*.

Besides the performance metrics discussed above, we examined two subjective metrics: Participants' ratings of likeability (RQ3) and task difficulty as an implicit measure of confidence (additional exploratory question). Interestingly, our likeability results contradict Brewer *et al.*'s [2] findings that demonstrate the RC as a highly likeable color scheme: Our participants overall like the SC more than the RC (no difference for Isa, in favor of the SC for Choro). The fact that we see a difference based on map types also here supports the notion that the spatial context plays an important role when evaluating the color [8]. We thus reject our working hypothesis H4 that *participants will rate the RC more likeable than the SC*. Likeability is important to study because if people like something they may be 'fooled' by this feeling and think that thing is better for them [32], [33]. Therefore, we explored participants' task difficulty ratings, and if these match with actual task difficulty (response accuracy). For map reading tasks, both color scheme and map type matter for task difficulty ratings in about half of the pairs we compared. An additional analysis (Figure 9) shows that the task difficulty ratings broadly match participants' performance with both color schemes. In other words, in the case of the RC use, we did not detect signs of 'naïve cartography' [32] in this study. This leaves the popularity of the RC in media, conferences and scientific papers (see *Section 3*) somewhat more puzzling. Possibly it is not that people do not understand that RC may be harmful to their performance, but the *nuance* that it is bad for some tasks and not-so-bad for some others does not allow for building patterns of preference.

All experiments suffer from certain limitations and so does this study. For example, we opted for not randomizing the order of the tasks, motived by nature of the tasks we had in the study (first on maps with no legend, then with legend). Lack of randomization brings the risk of learning/order effects. To make sure that this was not the case, we examined the task accuracy *vs.* time in the order the tasks were solved, and found no detectable trend. While we cannot fully rule out the possibility of a hidden

learning effect, that would be true for both examined variables (color scheme and map type), so we believe those comparisons are justified. We also are aware that data type presented using different color schemes might matter. For example, future studies that use a color scheme other than shades of orange would most definitely be very useful to verify it the baseline SC would affect the outcomes. Similarly, one can test diverging colors to further verify if the effects we observe here remain consistent. Furthermore, there are some perceptual properties of, and associations with, color that might affect the intuitiveness when presenting magnitudes, *e.g.,* red as danger, or green as calming. Also, we tested only two map types among many possible data visualizations. Some of the effects of the RC on the performance shown in this study may hold while using other types of visualizations (e.g., we believe these can be replicated for other SC color schemes), whereas others may differ from our result (e.g., if the map type or data distributions change). While our study is large and we believe data is robust, more work is needed to test its generalizability to other maps and visualization types. Participant characteristics might be an issue too, *e.g.,* we had high school students under 20 years old. As perceptual and cognitive abilities change with age, we hope to repeat the study with different age cohorts to further examine our findings.

6 Conclusions

In this paper, we examined if the RC was still popular, and what are the consequences of using it for representing quantitative data. Our findings show that the RC is a) still very popular, b) is not intuitive, and clearly harms performance for tasks that require ordering colors, c) task type and map type matters *i.e.,* the context matters (different from abstract color ordering tasks). Specifically, we demonstrated that the RC does not clearly support general pattern interpretation. Thus, common and important visualization tasks such as pattern recognition or comparing regional differences are not properly assisted by the RC. On the other hand, for tasks requiring reading details, the RC can be competitive. Nonetheless, even for tasks where the RC might facilitate, it is not appropriate for people with color vision deficiencies. At this point in time, color encoding advice is *still* largely based on conventions or people's convictions rather than empirical evidence [21]. As Brewer [30] noted, experimental work is the most convincing way to challenge conventional wisdom. Our work further establishes in which situations the RC is clearly not recommendable, but also demonstrates why one should examine if a variable is good or bad in its specific context.

Acknowledgments

Corresponding author: Izabela M. Gołębiowska (i.golebiowska@uw.edu.pl). We thank Izabela Karsznia, Jolanta Korycka-Skorupa, Tomasz Nowacki, Tomasz Panecki, and Katarzyna Słomska for their help with data collection. This work was supported by the National Science Centre, Poland under the grant no. UMO-2016/23/B/HS6/03846.

References

[1] J. Bertin, *Semiology of Graphics: Diagrams, Networks, Maps. Madison*. Madison: University of Wisconsin Press, 1983.

[2] C. A. Brewer, A. M. MacEachren, L. W. Pickle, and D. Herrmann, "Mapping mortality: Evaluating color schemes for choropleth maps," *Ann. Assoc. Am. Geogr.*, 87(3): 411–438, 1997.

[3] A. Brychtova and A. Coltekin, "Discriminating classes of sequential and qualitative colour schemes," *Int. J. Cartogr.*, 1(1): 62–78, 2015.

[4] D. A. Szafir, "Modeling Color Difference for Visualization Design," *IEEE Trans. Vis. Comput. Graph.*, 24(1): 392–401, 2018.

[5] R. Bujack, T. L. Turton, F. Samsel, C. Ware, D. H. Rogers, and J. Ahrens, "The Good, the Bad, and the Ugly: A Theoretical Framework for the Assessment of Continuous Colormaps," *IEEE Trans. Vis. Comput. Graph.*, 24(1): 923–933, 2018.

[6] J. M. Olson and C. A. Brewer, "An evaluation of color selections to accommodate map users with color-vision impairments," *Ann. Assoc. Am. Geogr.*, 87(1): 103–134, 1997.

[7] A. Brychtova and A. Coltekin, "An Empirical User Study for Measuring the Influence of Colour Distance and Font Size in Map Reading Using Eye Tracking," *Cartogr. J.*, 53(3): 202–212, 2016.

[8] A. Brychtová and A. Çöltekin, "The effect of spatial distance on the discriminability of colors in maps," *Cartogr. Geogr. Inf. Sci.*, 44(3): 229–245, 2017.

[9] A. Çöltekin, A. Brychtová, A. L. Griffin, A. C. Robinson, M. Imhof, and C. Pettit, "Perceptual complexity of soil-landscape maps: a user evaluation of color organization in legend designs using eye tracking," *Int. J. Digit. Earth*, 10(6): 560–581, 2016.

[10] C. A. Brewer, G. W. Hatchard, and M. A. Harrower, "ColorBrewer in Print: A Catalog of Color Schemes for Maps," *Cartogr. Geogr. Inf. Sci.*, 30(1): 5–32, 2003.

[11] C. A. Brewer, M. Harrower, B. Sheesley, A. Woodruff, and D. Heyman, "Color Brewer 2.0. Color advice for cartography," 2013. http://colorbrewer2.org/.

[12] A. Brychtová, J. Doležalová, and O. Štrubl, "Sequential Scheme Generator," 2015. http://eyetracking.upol.cz/color.

[13] A. Brychtová and J. Doležalová, "Designing Usable Sequential Color Schemes for Geovisualizations," in *Proc. of the 1st ICA Europ. Symp. on Cartogr.*, 2015, pp. 31–32.

[14] G. Aisch, "Chroma.js Color Palette Helper," 2020. https://vis4.net/palettes/.

[15] "Colorgorical." http://vrl.cs.brown.edu/color.

[16] C. C. Gramazio, D. H. Laidlaw, and K. B. Schloss, "Colorgorical: Creating discriminable and preferable color palettes for information visualization," *IEEE Trans. Vis. Comput. Graph.*, 23(1): 521–530, 2017.

[17] R. Stauffer, G. J. Mayr, M. Dabernig, and A. Zeileis, "Somewhere Over the Rainbow: How to Make Effective Use of Colors in Meteorological Visualizations," *Bull. Am. Meteorol. Soc.*, 96(2): 203–216, 2015.

[18] B. Jenny and N. V. Kelso, "Color Oracle." https://colororacle.org.

[19] B. Jenny and N. V. Kelso, "Color Design for the Color Vision Impaired," *Cartogr. Perspect. Perspect.*, 57:61–67, 2007.

[20] B. E. Rogowitz and A. D. Kalvin, "The 'Which Blair Project': A quick visual method for evaluating perceptual color maps," *Proc. IEEE Vis. Conf.*, pp. 183–188, 2001.

[21] K. Reda, P. Nalawade, and K. Ansah-Koi, "Graphical perception of Continuous quantitative maps: The effects of spatial frequency

and colormap design," *Conf. Hum. Factors Comput. Syst. - Proc.*, 2018.

[22] L. Zhou and C. D. Hansen, "A Survey of Colormaps in Visualization," *IEEE Trans. Vis. Comput. Graph.*, 22(8): 2051–2069, 2016.

[23] D. Borland and R. M. Taylor, "Rainbow color map (still) considered harmful," *IEEE Comput. Graph. Appl.*, 27(2): 14–17, 2007.

[24] M. A. Borkin *et al.*, "Evaluation of artery visualizations for heart disease diagnosis," *IEEE Trans. Vis. Comput. Graph.*, 17(12): 2479–2488, 2011.

[25] R. Bujack, T. L. Turton, D. H. Rogers, and J. P. Ahrens, "Ordering Perceptions about Perceptual Order," in *2018 IEEE Scientific Visualization Conf. (SciVis)*, 2018, pp. 32–36.

[26] M. Green, "Toward a Perceptual Science of Multidimensional Data Visualization : Bertin and Beyond," *ERGO/GERO Hum. Factors Sci.*, 8: 1–30, 1998.

[27] B. E. Trumbo, "A Theory for Coloring Bivariate Statistical Maps," *Am. Stat.*, 35(4): 220–226, 1981.

[28] A. Silverman, C. Gramazio, and K. Schloss, "The dark is more (Dark+) bias in colormap data visualizations with legends," *J. Vis.*, 16(12): 628, 2016.

[29] K. B. Schloss, C. C. Gramazio, A. T. Silverman, M. L. Parker, and A. S. Wang, "Mapping Color to Meaning in Colormap Data Visualizations," *IEEE Trans. Vis. Comput. Graph.*, 25(1): 810–819, 2019.

[30] C. A. Brewer, "Spectral Schemes: Controversial Color Use on Maps," *Cartogr. Geogr. Inf. Sci.*, 24(4): 203–220, 1997.

[31] A. Çöltekin, S. Bleisch, G. Andrienko, and J. Dykes, "Persistent challenges in geovisualization – a community perspective," *Int. J. Cartogr.*, 3(1): 115–139, 2017.

[32] M. Hegarty, H. S. Smallman, A. T. Stull, and M. S. Canham, "Naïve cartography: How intuitions about display configuration can hurt performance," *Cartogr. Int. J. Geogr. Inf. Geovisualization*, 44(3): 171–186, 2009.

[33] H. S. Smallman and M. S. John, "Naive realism: misplaced faith in realistic displays," *Ergon. Des. Q. Hum. Factors Appl.*, 13(3): 6–13, 2005.

[34] T. Munzner, *Visualization analysis and design*, CRC Press, 2014.

[35] K. Moreland, "Why we use bad color maps and what you can do about it," *Hum. Vis. Electron. Imaging 2016, HVEI 2016*, pp. 262–267, 2016.

[36] A. M. MacEachren, *How maps work: Representation, visualization, and design*. New York: Guilford, 1995.

[37] C. Ware, *Information visualization: perception for design*. Elsevier, 2012.

[38] H. H. McCarty and N. E. Salisbury, *Visual comparison of isopleth maps as a means of determining correlations between spatially distributed phenomena*. Iowa City: Studies in Geogr. No. 3, 1961.

[39] D. J. Cuff, "The Magnitude Message: A Study of the Effectiveness of Color Sequences on Quantitative Maps.," The Pennsylvania State Univ., 1972.

[40] J. E. Mersy, "Colour and Thematic Map Design: The Role of Colour Scheme and Map Complexity in Choropleth Map Communication," *Cartogr. Int. J. Geogr. Inf. Geovisualization*, 27(3): 1–167, 1990.

[41] M. D. Hyslop, "A Comparison of User Performance on Spectral Colour and Grayscale Continuous-Tone Maps," Michigan State Univ., 2007.

[42] P. S. Quinan, L. M. Padilla, S. H. Creem-Regehr, and M. Meyer, "Examining implicit discretization in spectral schemes," *Comput. Graph. Forum*, vol. 38, no. 3, pp. 363–374, 2019.

[43] M. P. Kumler and R. E. Groop, "Continuous-tone mapping of smooth surfaces," *Cartogr. Geogr. Inf. Syst.*, 17(4): 279–289, 1990.

[44] J. M. Olson, "Spectrally encoded two-variable maps," *Ann. Assoc. Am. Geogr.*, 71(2): 259–276, 1981.

[45] C. Ware, T. L. Turton, F. Samsel, R. Bujack, and D. H. Rogers, "Evaluating the perceptual uniformity of color sequences for feature discrimination," *EuroRVVV EuroVis Work. Reprod. Verif. Valid. Vis.*, 2017.

[46] C. Ware, T. L. Turton, R. Bujack, F. Samsel, P. Shrivastava, and D. H. Rogers, "Measuring and Modeling the Feature Detection Threshold Functions of Colormaps," *IEEE Trans. Vis. Comput. Graph.*, 25(9): 2777–2790, 2019.

[47] Y. Liu and J. Heer, "Somewhere over the rainbow: An empirical assessment of quantitative colormaps," *Conf. Hum. Factors Comput. Syst. - Proc.*, 2018.

[48] C. Ware, "Color/Maps Color Sequences for Univariate Maps: Theory, Experiments, and Principles," *IEEE Comput. Graph. Appl.*, 8(5): 41–49, 1988.

[49] A. H. Robinson, "Psychological Aspects of Color in Cartography," *Int. Yearb. Cartogr.*, 7: 50–61, 1967.

[50] E. H. Lenneberg, "Color Naming, Color Recognition, Color Discrimination: A Re-Appraisal," *Percept. Mot. Skills*, 12(3): 375–382, 1961.

[51] R. Schnürer, R. Sieber, and A. Çöltekin, "The next generation of atlas user interfaces: A user study with 'Digital Natives,'" in *Modern Trends in Cartogr.*, 2015, pp. 23–36.

[52] C. A. Brewer, "Guidelines for selecting colors for diverging schemes on maps," *Cartogr. J.*, 33(2): 79–86, 1996.

[53] G. A. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information.," *Psychol. Rev.*, 63(2): 81–97, 1956.

[54] A. Çöltekin, "What contributes to the complexity of visuospatial displays?," in *Abstraction, Scale and Perception. Intern. Cartogr. Assoc. Joint Comm. Workshop*, 2019, pp. 1–2.

[55] L. Knapp, "A Task Analysis Approach to the Visualization of Geographic Data," in *Cognitive Aspects of Human-Computer Interaction for Geographic Information Systems*, Dordrecht: Springer Netherlands, 1995, pp. 355–371.

[56] K. Moreland, "Diverging Color Maps for Scientific Visualization," 2009, pp. 92–103.

[57] M. Monmonier, *How to lie with maps*. Chicago: The Univ. of Chicago Press, 1991.

I.M. Gołębiowska, PhD, is an Assistant Professor at the University of Warsaw, Faculty of Geography and Reg. St. Her research interests include topics related to GIScience, geovisualization, perception and cognition covering user studies with maps and geovisualization tools. She is also involved in utilizing GIS and geovisualization within other fields, such as planetary geology and analog Mars simulation.

A. Çöltekin, PhD, is a Professor at the University of Applied Sciences and Arts Northwestern Switzerland, Department of Computer Science, Institute of Interactive Technologies. She is also affiliated with the Seamless Astronomy group at Harvard University, chairs the working group "Visualization, Augmented and Virtual Reality" with the ISPRS and co-chairs the visual analytics commission with the ICA. Previously she was a group leader at the University of Zurich, Department of Geography. She has been the PI and co-PI in more than a dozen national and international grants, authored/co-authored more than hundred publications, edited journal special issues and books. She was awarded Google faculty research award in 2015 and her work with her PhD advisees has been awarded a best short paper in EuroVis2015 and third best poster award in Spatial Cognition 2018.