Received: 15 April 2016 | Revised: 23 January 2018 | Accepted: 23 January 2018
DOI: 10.1002/acp.3396

RESEARCH ARTICLE

WILEY

Integration processes during frequency graph comprehension: Performance and eye movements while processing tree maps versus pie charts

Lynn Huestegge | Tristan Herbert Pötzsch

Würzburg University, Würzburg, Germany

Correspondence
Lynn Huestegge, Department of Psychology III, University of Würzburg, Röntgenring 11, 97070 Würzburg, Germany.
Email: lynn.huestegge@uni-wuerzburg.de

Summary

Frequency graph types differ in the way how data are translated into visual representations. We compared 2 visualization methods, a traditional circular representation (pie chart) and a rectangular representation (constant column width tree map), which were hypothesized to differ regarding the cognitive ease of visual comparison processes. Performance was evaluated in tasks involving proportion and comparison judgments under both highly controlled and more realistic circumstances. The results showed performance benefits (in terms of reduced response times or error rates) for rectangular representations. Additional eye movement analyses revealed that this benefit was mainly due to a facilitation of scanning the graph for relevant information. The results suggest that facilitating comparison processes by representing the critical variable in less complex visual dimensions (i.e., straight length with constant orientation instead of surface area or curved length) eventually enhances the efficiency of integration processes during graph comprehension.

KEYWORDS

eye movements, frequency graph comprehension, pie charts, tree maps

1 | INTRODUCTION

The increase of computational power and networking over the last decades resulted in a massive increase of available quantitative data—a phenomenon commonly referred to as Big Data. This calls for efficient data reduction techniques (Press, 2013), for example, by utilizing visualization methods including various types of graphs. Graphs are a common tool to display large sets of quantitative information within relatively small space (Larkin & Simon, 1987) and do not only occur in all fields of science but also in public print media and in the classroom. Therefore, studying the cognitive processes underlying graph comprehension along with the aim to optimize the display of data in a theory-driven manner represents an important mission within the field of cognitive science. This study focuses on visualization techniques for one of the most widely used type of quantitative data depicted in graphs, namely, categorical (relative) frequency data. Pie charts and bar graphs are the most commonly used graph types in this context and are readily supported by standard spreadsheet software (Harris, 1999). Here, we will compare graph comprehension performance resulting from two graph types that presumably differ in the extent to which they draw on cognitive processes, namely, traditional (circular) pie charts and a newly developed (rectangular) alternative, a specific variant of the increasingly popular tree maps (see www.cs.umd.edu/hcil/treemap-history).

Guidelines for graph design were often derived by postulating intuitively plausible principles, for example, maximizing the "data-ink ratio" (Tufte, 2001; see also Kosslyn, 1994). However, in recent years, many aspects of graph design were also evaluated empirically (e.g., Carpenter & Shah, 1998; Carswell, Frankenberger, & Bernhard, 1991; Fischer, 2000; Huestegge & Philipp, 2011; Körner, Höfler, Tröbinger, & Glichrist, 2014; Peebles & Cheng, 2003; Ratwani, Boehm-Davis, & Trafton, 2008; Riechelmann & Huestegge, in press; Shah & Carpenter, 1995; Siegrist, 1996; Spence, 1990; Zacks, Levy, Tversky, & Schiano, 1998). Prior to presenting a brief overview of empirical findings relevant for our present study, we will first outline some theoretical considerations regarding the cognitive processes underlying graph comprehension to derive relevant implications for the representation of frequency data.

2 | COGNITIVE PROCESSES DURING GRAPH COMPREHENSION

Many theoretical accounts of graph comprehension have been proposed about 20–30 years ago based on both formal task analyses and empirical research (e.g., Bertin, 1983; Cleveland & McGill, 1984, 1986, 1987; Gillian & Lewis, 1994; Kosslyn, 1989; Lohse, 1991, 1993; Pinker, 1990). One central concept in many of these theories, which is important in the context of this study, is the notion of cognitive integration processes, which are essential for graph comprehension whenever distinct graphs elements and/or their meanings need to be linked. They are especially relevant when several elements in a graph need to be compared (e.g., to make comparison and proportion judgments, see Simkin & Hastie, 1987). For example, Carpenter and Shah (1998) proposed a flexible model that can account for such integration processes. Specifically, it suggests three processing stages, namely, a pattern recognition stage involving the encoding of a visual pattern by chunk formation, and two interpretative stages. The first interpretative stage involves translating the visual pattern into its quantitative and qualitative interpretation and also comprises arithmetic operations on encoded values and the comparison of spatial relations of indicators or shares (i.e., integrating visual features, see Gillian & Lewis, 1994). The second interpretative stage relates these decoded patterns to the referents in the graph (see Huestegge & Philipp, 2011; Zacks & Tversky, 1999). The efficiency of this latter process is also determined by graph design features, for example, direct labeling versus a legend (affecting data-legend integration). According to Carpenter and Shah (1998), these three stages are assumed to be repeated in a cyclical fashion, which is reflected in corresponding graph readers' gaze transitions between respective graph regions (e.g., main graphical pattern, axes, legend, and title; see Bertin, 1983). Building on this framework, Ratwani et al. (2008) further distinguished between visual and cognitive integration processes: Whereas visual integration refers to pattern recognition and visual cluster formation, cognitive integration involves further reasoning about the graph based on specific task instructions. Finally, another line of research utilizing oculomotor analyses demonstrated that integration performance is not only determined by specific graph types but also by user characteristics (Peebles, 2008; Peebles & Cheng, 2001, 2003).

3 | INTEGRATION PROCESSES IN PIE CHART COMPREHENSION

The pros and cons of different types of graphs depicting frequency data can be related to the various integration processes referred to above. However, it is important to note that graphs depicting frequency data (such as pie charts) can serve different purposes, each involving different sets of mental operations. Probably the two most basic tasks for understanding frequency graphs (e.g., Simkin & Hastie, 1987) are proportion judgments (i.e., estimating the size of one element relative to the whole) and comparison judgments (i.e., comparing the size of two elements to judge which one is larger). Principally, both tasks can be supported by utilizing specific design features, such as direct labeling of relative element size in percent (for proportion judgments) and size-based ordering of elements (e.g., in a clockwise manner for pie charts) for comparison judgments.

However, several drawbacks of pie charts still remain: First, it should generally be more difficult to compare surface areas of pie slices, which are unusually complex geometrical figures involving both curved and straight lines. Thus, it should be more difficult to judge the size of pie slices or the length of a curved outer line (both requiring definition in two spatial dimensions) when compared to a situation that can rely on the comparison of a simpler, one-dimensional feature only, such as straight line length (e.g., in segmented bar graphs using only segment length as relevant information). Second, the comparison of pie slices can also be considered more complex in that it may involve mental rotation. Specifically, on the one hand, all slices are connected at the center of the pie, and thus, any comparison between shares are in line with the proximity compatibility principle (Wickens & Carswell, 1995), which assumes that when a task requires the integration of multiple sources of information (e.g., for comparison purposes), performance will only be optimal when that information is displayed in close spatial proximity. On the other hand, however, the comparison is made difficult because the relevant information (either the whole slice surface area or the length of the curved outer line segment) needs to be transformed (i.e., mentally rotated) to be compared. Finally, pie charts come with a disadvantage regarding the extraction of the referents related to the slices due to font alignment issues. Specifically, it appears difficult to implement an easy to read, upright labeling of slices (i.e., category names or additional information of relative frequency in percent) in many instances, especially when many (and thus small) slices are present.

Some of these theoretical considerations have also been tested empirically. For example, the question of whether pie charts are suited to display percentages (i.e., for proportion judgments) was already examined by Eells (1926), who found that bar graphs were superior. Based on similar observations, Cleveland and McGill (1984) concluded that pie charts should always be replaced by bar charts. Spence and Lewandowsky (1991), on the other hand, found that combinations of shares could sometimes be compared more easily when participants were provided with a pie chart than with a bar graph. In line with this finding in favor of pie charts, another study considered pie charts useful for relative proportion judgments (Hollands & Spence, 2001). As a response to some of the difficulties associated with comparison processes in pie charts, Gillian and Callahan (2000) proposed the use of specifically aligned pie charts (i.e., presenting the two to-be-compared slices side by side, each aligned to the same 12 o'clock starting point), although these never became widespread (see also Hollands & Spence, 2001, on the issue of alignment). Other studies reported similar proportion judgment accuracy for pie charts and bar graphs (whereas performance for multiple stacked bar graphs was less accurate; Simkin & Hastie, 1987). Further research comparing both objective and subjective effects suggested that even when bar graphs yielded lower comprehension than pie charts, participants still subjectively preferred bar graphs (Fagerlin et al., 2007). Taken together, the empirical evidence regarding the two visualization types so far is rather mixed and appears to strongly depend on specific graph characteristics and the task at hand.

One advantage that may explain some effects in favor of pie charts referred to above is that they contain some visual cues that may facilitate comprehension. As Eells (1926) already pointed out, graph readers not only compare slices to the whole in a pie chart but also to reference portions of a circle (e.g., a quarter, half, and three quarters) to estimate shares. This concept was later referred to as anchoring and has been shown to have significant impact on accuracy and reaction times in graph comprehension (e.g., Chandrasekaran & Lele, 2010; Simkin & Hastie, 1987; Spence & Lewandowsky, 1991). In pie charts, this mechanism appears to counteract the problem that visual angle judgments in general are known to be quite inaccurate (Cleveland & McGill, 1985; Simkin & Hastie, 1987).

4 | CONSTANT COLUMN WIDTH TREE MAPS (CCW-TREE MAPS) AS A POTENTIAL PIE CHART ALTERNATIVE

Based on the rather mixed empirical evidence in the previous studies reviewed above, we reasoned that it should be possible to combine many benefits of both major graph types (pie charts and bar graphs) while avoiding most of the respective drawbacks. To this end, we specified a new graph type alternative for the display of frequency data. In recent years, many new visualization techniques benefitted from modern web technologies such as HTML5 and CSS3 and were incorporated into modern business intelligence software or open code libraries. A particularly successful variant of these newly emerging graph techniques was the tree map (Shneiderman, 1992; see Figure 1). The tree map was conceptualized as a "two-dimensional (2-d) space-filling approach in which each node [entity] is a rectangle whose area is proportional to some attribute such as node size" (Shneiderman, 1992, p. 1). Although Shneiderman emphasized the possibility of displaying hierarchical data in tree maps (i.e., one can usually zoom into specific rectangles to look at child nodes), their probably most common application so far has been as a substitute for pie charts. For example, some news websites present their content according to coverage and group information accordingly into news clusters such as business, culture, and sports (e.g., www.newsmap.jp), and recent versions of popular spreadsheet software (by Microsoft, Google) now include tree maps as a standard display option. Furthermore, several variants of the tree map were derived (e.g., the clustered tree map, Wattenberg,1999, or the squarified tree map, Bruls, Huizing, & van Wijk, 2000).

Several theoretically informed design choices were made to come up with a viable competitor to pie charts. First, we reasoned that one major disadvantage of the original tree map is that to estimate the size of a share, two dimensions (height and width) need to be considered (similar to the difficulties associated with estimating surface areas of pie slices, see Cleveland & McGill, 1984; Kosslyn, 1994; Carswell, 1992). Therefore, we concluded that a reduction to one relevant dimension should increase computational efficiency (Spence, 1990). Thus, we decided to utilize only one dimension (height) to represent relevant information, whereas the other (width) is held constant (similar to bar width in bar graphs, see, e.g., Cleveland & McGill, 1984; Simkin & Hastie, 1987; Spence, 1990; Friel, Curcio, & Bright, 2001). Comparison of spatial relations between shares in such tree maps should be easier because at least some of the relevant length information is displayed in proximity in the same spatial orientation, rendering complex and potentially error-prone mental rotations (aligning) unnecessary. In contrast, comparison performance in pie charts is known to depend on the angular and size differences between segments (Gillian & Callahan, 2000). In the light of this salient design feature, we will from now on refer to the specific tree maps used in our study as CCW-tree maps.

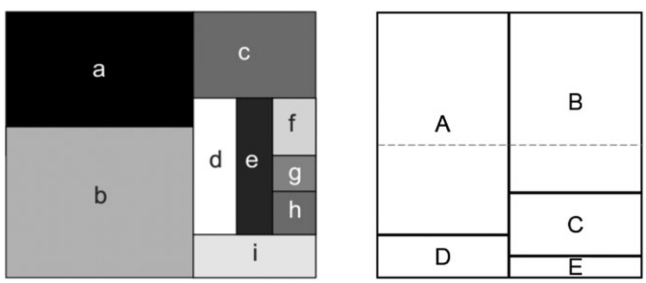

FIGURE 1 Example of a traditional tree map (Shneiderman, 1992; left panel) and a constant column width tree map (including visual anchors indicating quarters) used in this study (right panel)

Second, the principle of providing visual anchors was adapted by including the display of quadrants in the rectangular-sized tree maps. Third, one potential general advantage of tree maps relates to labeling options. Usually, tree maps are labeled directly, which is much easier to implement in rectangular than in circular graphs. Utilizing direct labeling is typically considered to increase efficiency of integration processes (e.g., Gillan, Wickens, Hollands, & Carswell, 1998; Kosslyn, 1994). Furthermore, direct labeling also implies that CCW-tree maps can be considered more space efficient than pie charts (which often include labels located outside the pie), which may also reduce the need for extended eye movements (and associated processing time). Additionally, CCW-tree map labeling allows for more natural reading patterns when compared with the relatively unusual circular arrangement of text (implying clockwise reading) in pie charts (see Huestegge & Radach, 2012; Huestegge, Radach, Kunert, & Heller, 2002, for evidence suggesting that visual search typically follows reading-like paths). Finally, direct labelling should prevent unnecessary attentional "cycling" back and forth between the legend and the indicators/shares (Carpenter & Shah, 1998; Ratwani et al., 2008; Trafton, Marshall, Mintz, & Trickett, 2002). The issue of labeling will be explicitly addressed in Experiment 3.

5 | THIS STUDY

Taken together, our theoretically informed design decisions led to the development of the CCW-tree maps as a potential competitor for pie charts (see Figure 1). CCW-tree maps were designed to combine the benefits of pie charts (e.g., by providing visual anchors), original tree maps (comparatively easier labeling and thus enhanced space efficiency), and bar graphs (accurate, uni-dimensional length code). Based on our design decisions, we reasoned that CCW-tree maps should principally be easier to process than pie charts, resulting in shorter response times (RTs) and/or fewer errors. However, it is important to note that experience and previous knowledge (see Pinker, 1990) also play an important role for the efficiency of integration processes. For example, both familiar graph layouts and data patterns reflecting expectations based on real-world knowledge yielded faster and/or more accurate graph comprehension (Fischer, Dewulf, & Hill, 2005; Gattis & Holyoak, 1996; Shah, 1995). In the context of the present research, it is thus a relevant question whether the presumable efficiency gains regarding integration processes in CCW-tree maps outweigh potential drawbacks associated with a less familiar visualization (when compared with the more frequently used pie charts). The overall aim of this study was thus to compare the two types of graphs in the context of two different tasks (proportion and comparison judgments), and under both experimentally controlled as well as more natural conditions.

Four experiments were conducted with graph type (pie chart vs. CCW-tree map) as a within-subject factor (i.e., all blocks of trials contained both pie charts and CCW-tree maps). Whereas Experiment 1 examined the accuracy of participants' proportion judgments, Experiment 2 addressed comparison judgments. Whereas the first two experiments involved highly controlled stimulus material, Experiment 3 aimed at determining whether any advantage of CCW-tree maps also transfers to more realistic visualization circumstances by taking typical graph labeling (semantic context) into account. Apart from overall performance-related parameters (RTs, error rates), we additionally recorded eye movements for exploratory analyses in these three experiments. We reasoned that more difficult visual integration processes for pie charts could be observable in terms of greater oculomotor effort (e.g., Carpenter & Shah, 1998; Huestegge & Philipp, 2011; Körner et al., 2014; Peebles & Cheng, 2003; Ratwani et al., 2008; Renshaw, Finlay, Tyfa, & Ward, 2004). Finally, Experiment 4 focused on more realistic data sets that cannot be evenly divided into two 50% shares (which was a prerequisite of designing perfectly rectangular CCW-tree maps in Experiments 1–3) and also involved single-stacked bar graphs as a third type of graph layout.

6 | EXPERIMENT 1: PROPORTION JUDGMENTS

Experiment 1 compared pie charts and CCW-tree maps with respect to proportion judgment performance. Specifically, participants were asked to estimate the size of a single share (in %). Given that judging the relative size of a single share implicitly also involves a comparison with the size of the remaining portion of the graph (single-whole comparison), we reasoned that the ease of integration processes should also play a role for performance in this task. Thus, if our measures to facilitate integration processes via the CCW-tree maps are effective and their positive effects outweigh any advantages associated with a greater familiarity with pie charts, we expected that performance (in terms of RTs, judgment accuracy, and/or oculomotor effort) should be better for CCW-tree maps than for pie charts.

6.1 | Method

6.1.1 | Participants

Experiments 1–3 were conducted in fixed sequence with the same set of participants, which were recruited via social media. All participants (N = 15, 6 male, 9 female, mean age = 22 years, SD = 2.6, range: 19–28) were university students or had recently completed their study program (most of them were enrolled in Psychology). All participants were fluent German speakers.

6.1.2 | Stimuli

For creating the stimulus material, 30 frequency data sets were constructed. These data sets were used to create 30 pie charts and 30 corresponding CCW-tree maps, respectively (i.e., the same data set was displayed in both ways; see Figure 2). The area covered by each pie (slice) and the corresponding CCW-tree map (rectangles) was kept constant.

The data sets involved five, seven, or nine indicators (10 data sets for each condition). All indicators of one data set referred to different percentage values. None represented a share less than 5% (to avoid issues related to very small segments/labels), and indicators were always dividable into two groups of 50% each. This restriction was due to an inherent limitation of CCW-tree maps for our present purpose: To achieve an overall rectangular shape of the graph that is similar to the coherent circular shape of pie charts, the shares needed

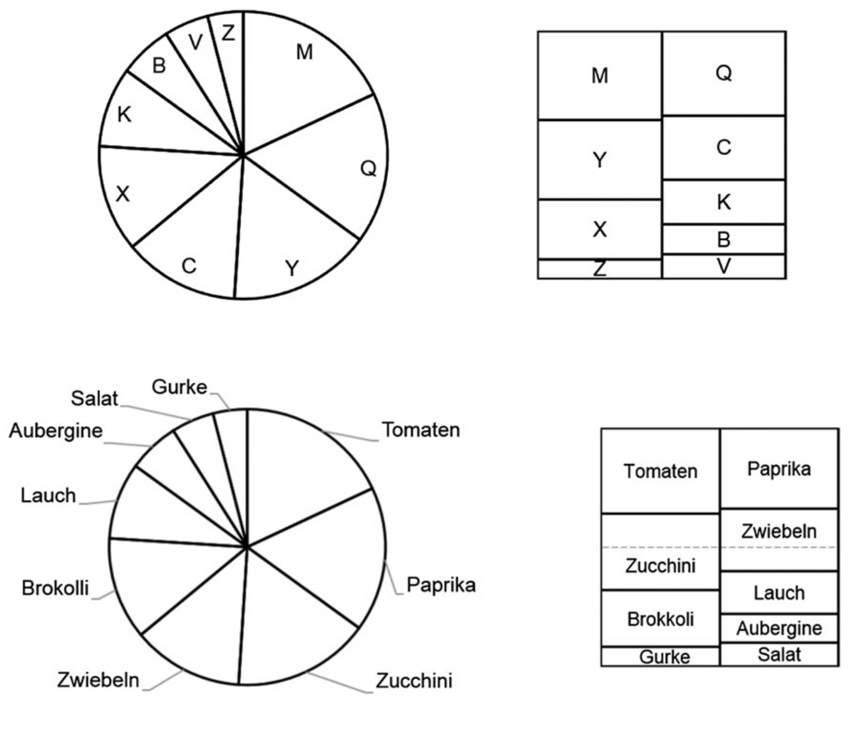

FIGURE 2 Examples of stimuli used in Experiments 2 (upper panel) and 3 (lower panel, German labels referring to vegetables). Experiment 1 involved similar stimuli as in Experiment 2, but with additional visual anchors (see text for details). Furthermore, in Experiment 1 (unlike in Experiment 2), shares were labeled in alphabetical order

to be organized into two columns of identical width (both columns representing 50%).

All indicators were labeled directly using letters, and percentage values were not explicitly displayed. In each graph, all shares were sorted according to their size (starting from the top and in clockwise orientation for pie charts, and from top to bottom for CCW-tree maps). In Experiment 1, the biggest indicator was always labeled “A,” the second biggest “B,” etc. All stimuli (pie charts and CCW-tree maps) in Experiment 1 (but not those in Experiment 2) had additional visual anchors (dashed lines at 90°, 180°, and 270° for pie charts, and a dashed horizontal line in the middle of CCW-tree maps). The experiment consisted of 120 trials; therefore, each data set was used four times (i.e., each pie chart and each tree map was presented two times). Trial order was randomized but remained constant for all participants.

A brief subjective follow-up survey administered immediately after the run of the first three experiments additionally examined which of the two visualization methods (forced choice) was perceived as (a) more pleasant and (b) a more accurate basis for judgments. Although the response rate for this survey was only 73%, 55% (vs. 45%) of the responders stated that pie charts were “more pleasant to use” than CCW-tree maps, whereas only 27% (vs. 73%) stated that pie charts “led to more accurate judgments” than CCW-tree maps. These introspective results suggest that CCW-tree maps tended to be perceived as supporting more accurate performance than pie charts, even though the underlying sample size is too small for meaningful statistical conclusions.

6.1.3 | Procedure

For eye tracking purposes, an EyeLink 1000 (SR Research, Ontario, Canada) system was used in a stationary setup. The experiments were programmed using Experiment Builder software (SR Research). Participants sat 50 cm in front of a screen with a display resolution of 1024 * 768 pixels. Participants were first familiarized with the concept of the newly developed CCW-tree maps (all participants reported to be familiar with pie charts). Specifically, participants were instructed how CCW-tree maps were constructed and how they code shares. Note, however, that we did not reveal any information regarding size ordering of the shares for either one of the two graph types.

Each experiment started with a calibration of the eye tracker. At the beginning of each experimental trial, a question was presented on the screen (e.g., “How big is share A?”). The identity of the probed share varied from trial to trial in a random manner. After encoding the question, participants pressed the space key, which triggered a central fixation cross in the middle of the screen (1,500 ms), followed by the graph. The midpoint of each graph was always located at the screen center. The onset of the oral answer was used to calculate RTs using the built-in voice key functionality in the Experiment Builder software. The content of the answer was recorded by the investigator. The investigator also noted failed trials (when other sounds than the answer triggered the voice key or when the participant's response was too quiet to trigger the voice key), which were excluded from data analysis. On average, the three experiments lasted 50 min in total, including short breaks between experiments.

6.1.4 | Design and data analysis

The independent variables in this experiment were graph type (pie chart vs. CCW-tree maps) and the number of indicators per graph (five, seven, or nine). The dependent variables were accuracy (defined as the absolute difference between estimated and actual indicator size) and RTs; 2 × 3 repeated measures ANOVAs were conducted unless otherwise indicated. Apart from standard parameters of eye movement analyses (number of fixations, mean fixation duration, and mean saccade amplitude), we further analyzed mean (horizontal and vertical) fixation position and fixation dispersion as an additional index of oculomotor effort as well as relative directional distributions of saccades (i.e., the relative frequency of leftward, rightward, upward, and downward saccades) as indices of search systematicity.

6.2 | Results and discussion

The ANOVA for absolute accuracy revealed a significant main effect of the number of indicators, $F(2, 28) = 12.25$, $p < .001$, $\eta_p^2 = .467$, indicating that in the context of the present proportion judgment task, accuracy increased with the number of indicators in the graph (mean absolute deviation = 2.42, 1.69, and 1.71 for five, seven, and nine indicators, respectively; see Figure 3). Although there was no main effect of graph type, $F < 1$, we observed a significant disordinal interaction of number of indicators and graph type, $F(2, 28) = 6.99$, $p < .01$, $\eta_p^2 = .333$. Specifically, although accuracy tended to be greater for pie charts than for CCW-tree maps when there were only few indicators, this difference was reversed in the condition with a maximum number of indicators (nine), where the results suggest an advantage of CCW-tree maps over pie

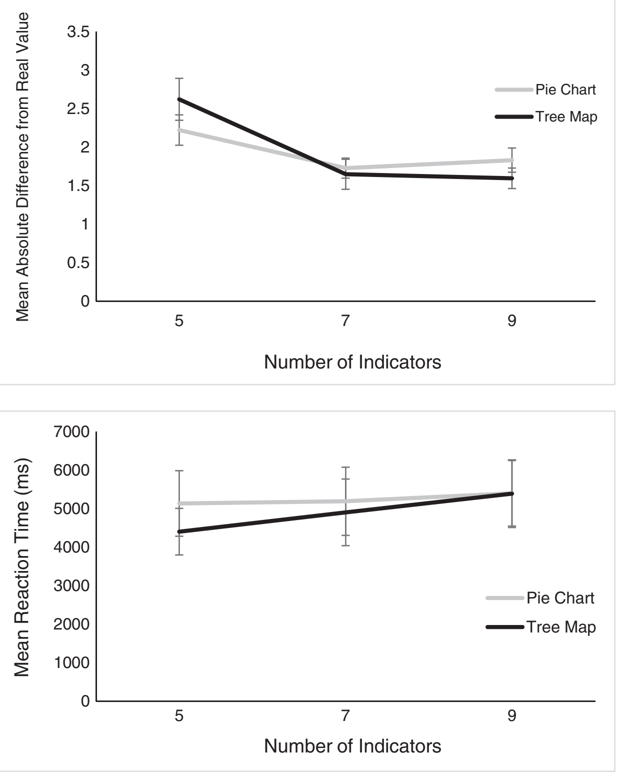

FIGURE 3 Accuracy and response time results in Experiment 1 as a function of graph type and number of indicators. Error bars represent standard errors of the mean

charts. However, post hoc pairwise comparisons revealed no significant simple main effects of graph type in the individual number of indicator conditions (p = .059, p = .741, and p = .165 for three, five, and seven indicators, respectively).

As an alternative accuracy analysis, we additionally looked at the relative (instead of absolute) deviation of the estimated value from the real value. In this analysis, there was no longer a significant main effect of the number of indicators, $F(2, 28) = 2.09$, $p = .142$, nor a significant main effect of graph type, $F(1, 14) = 1.24$, $p = .284$, but still a significant interaction, $F(2, 28) = 11.89$, $p < .001$. Here, post hoc pairwise comparisons revealed a significant simple main effect of graph type in the five indicators condition (tree maps: 19.3% mean deviation, SE = 2.6; pie charts: 11.5% mean deviation, SE = 0.9, p = .004), but no significant differences in the seven indicators condition (tree maps: 13.1% mean deviation, SE = 1.4; pie charts: 14.1% mean deviation, SE = 1.2, p = .598), nor in the nine indicators condition (tree maps: 15.9% mean deviation, SE = 1.3; pie charts: 17.7% mean deviation, SE = 1.6, p = .249).

The RT analysis revealed a somewhat different picture (Figure 3). Again, there was a significant main effect of the number of indicators, $F(2, 28) = 6.83$, $p < .01$, $\eta_p^2 = .328$, indicating that RTs increased with an increasing number of indicators in the graph (M = 4,768, 5,048, and 5,389 ms). The RT data also revealed a significant main effect of graph type, $F(1, 14) = 9.12$, $p < .01$, $\eta_p^2 = .394$, suggesting faster overall RTs for CCW-tree maps (M = 4,899 ms) as opposed to pie charts (M = 5,238 ms). A significant interaction of number of indicators and graph type, $F(2, 28) = 3.73$, $p = .037$, $\eta_p^2 = .210$, suggests that the advantage of CCW-tree maps was only present for graphs with a small number of indicators (i.e., in the five and seven indicators conditions, $p < .03$ and $p < .03$, respectively), but not in the condition with the maximum number of (nine) indicators, $t < 1$.

To come up with a more comprehensive, performance-related interpretation of the data, it is important to consider the results regarding both accuracy and RT at the same time. For example, regarding the graphs of low complexity (i.e., those with five indicators), it is difficult to come up with a final conclusion regarding overall performance benefits of one graph type over the other due to a speed-accuracy trade-off (i.e., CCW-tree maps were processed faster, but with a tendency towards lower accuracy than pie charts). A similar ambiguous picture emerges for the more complex graphs (seven and nine indicators), where no clear performance advantages for pie charts or CCW-tree maps regarding both parameters (accuracy and RTs) emerged. Overall, we thus conclude that the data in Experiment 1 suggest roughly comparable overall performance for CCW-tree maps and pie charts. However, these results are far from conclusive, and it will thus be interesting to examine the extent to which a clearer advantage might show up under task conditions that further emphasize comparison processes (see Experiment 2).

Regarding the effect of the number of indicators, it appears plausible to assume that the increase in time spent on the more complex graphs (i.e., those with more indicators) ultimately allowed participants to come up with greater overall proportion judgment accuracy. Furthermore, it appears conceivable that greater segmentation (that inevitably goes along with an increased number of indicators) generally increases the ability of participants to give accurate proportion estimates (albeit at the cost of increased processing time).

6.2.1 | Eye movement analyses

For the sake of brevity, the eye movement analyses reported in this section focused on selected major oculomotor processing differences between pie charts and CCW-tree maps (see Table 1 for a complete set of means and statistical parameters, including effects of the number of indicators). Overall oculomotor effort was increased for the processing of pie charts versus CCW-tree maps, as indicated by an increased number of fixations on the graph, prolonged saccade amplitudes, and greater horizontal and vertical dispersion of fixations. Furthermore, search in CCW-tree maps was characterized by more downward saccades and fewer left/right saccades compared to pie charts. Additionally, mean fixation position was shifted further to the right and to the upper part of the graph in CCW-tree maps as compared to pie charts. Together, these observations indicate different scanning strategies for the two graph types. In sum, the greater oculomotor effort for pie charts further specifies our central assumption that the underlying integration processes are more difficult for pie charts than for CCW-tree maps.

6.2.2 | Correlational analyses

Further variables were examined regarding their correlation with RTs. However, we only found a significant correlation between the number of fixations and RTs ($r = .88$ $p < .001$), which is to be expected given that longer processing time on a graph inevitably increases the time window for further fixations to occur (thus, we will not further report this correlation in the following experiments). There was no significant correlation between the position of the relevant indicator in the graph (i.e., first, second, and third) and RTs ($r = .11$, $p = .70$) and no evidence for a learning effect over the course of the experiment in terms of a significant negative correlation between trial number and RTs ($r = -.17$, $p = .54$). There was also no significant relationship between RTs and accuracy within each graph type (pie charts: $r = .18$, $p = .52$; CCW-tree maps: $r = -.02$, $p = .94$).

7 | EXPERIMENT 2: COMPARISON JUDGMENTS

The second experiment examined differences between CCW-tree maps and pie charts when making comparison judgments. Participants were asked to verbally indicate the larger one of two given indicators. Because, unlike in Experiment 1, comparison processes are explicitly instructed in Experiment 2 (and not just an implicit feature of a part-whole comparison), we reasoned that this greater emphasis on comparison should yield a more pronounced advantage of CCW-tree maps, which were explicitly developed to facilitate comparison processes (see Introduction). Many methodological details were similar to Experiment 1 (i.e., regarding participants, stimuli, overall procedure, design, and data analysis). Thus, the method section will only focus on differences to Experiment 1.

TABLE 1 Results of eye movement analyses in Experiment 1

Dependent variable	Indicators	Pie charts: Mean (*SE*)	Tree maps: Mean (*SE*)	*F*	*df*	*p*	η_p^2
Number of fixations (*N*)	5	14.8 (2.3)	13.0 (1.8)	Graph type: 5.38	1, 14	.037	.293
	7	15.1 (2.5	14.4 (2.5)	Indicator: 7.71	2, 28	.006	.372
	9	15.7 (2.3)	16.1 (2.5)	Interaction: 4.07	2, 28	.044	.238
Mean fixation duration (ms)	5	325 (23)	330 (26)	Graph type: 0.10	1, 14	.759	.007
	7	329 (24)	329 (22)	Indicator: 1.61	2, 28	.218	.103
	9	317 (17)	316 (21)	Interaction: 0.31	2, 28	.736	.022
Mean saccade amplitude (°)	5	2.74 (.10)	2.33 (.13)	Graph type: 67.98	1, 14	<.001	.839
	7	2.62 (.09)	2.21 (.10)	Indicator: 5.57	2, 28	.010	.300
	9	2.56 (.09)	2.29 (.11)	Interaction: 2.35	2, 28	.122	.153
Horizontal fixation dispersion (*SD* of x coordinates per trial)	5	52.6 (2.1)	46.4 (2.0)	Graph type: 97.83	1, 14	<.001	.883
	7	53.1 (1.6)	43.2 (1.8)	Indicator: 2.54	2, 28	.109	.163
	9	55.5 (2.1)	47.8 (2.8)	Interaction: 1.22	2, 28	.307	.086
Vertical fixation dispersion (*SD* of y coordinates per trial)	5	58.4 (4.0)	52.9 (6.0)	Graph type: 12.41	1, 14	.004	.488
	7	57.4 (3.5)	49.7 (5.0)	Indicator: 2.80	2, 28	.083	.177
	9	61.3 (3.9)	51.0 (4.3)	Interaction: 1.41	2, 28	.263	.098
Mean horizontal (x) coordinate (px)	5	494 (13)	520 (15)	Graph type: 44.88	1, 14	<.001	.775
	7	494 (14)	508 (15)	Indicator: 7.39	2, 28	.010	.362
	9	489 (15)	524 (15)	Interaction: 12.63	2, 28	.001	.493
Mean vertical (y) coordinate (px)	5	370 (16)	412 (16)	Graph type: 82.94	1, 14	<.001	.864
	7	384 (16)	422 (14)	Indicator: 10.42	2, 28	.001	.445
	9	383 (17)	410 (15)	Interaction:4.28	2, 28	.034	.248
Leftward saccades (%)	5	25.4 (1.3)	21.0 (1.5)	Graph type: 43.65	1, 14	<.001	.771
	7	26.5 (1.7)	21.0 (1.5)	Indicator: 1.05	2, 28	.365	.075
	9	26.7 (1.6)	19.0 (1.5)	Interaction: 2.01	2, 28	.165	.134
Rightward saccades (%)	5	26.4 (1.1)	23.6 (1.8)	Graph type: 6.83	1, 14	.021	.345
	7	26.1 (1.2)	23.2 (2.0)	Indicator: 0.19	2, 28	.795	.014
	9	26.2 (1.2)	23.2 (1.8)	Interaction: 0.01	2, 28	.986	.001
Upward saccades (%)	5	24.9 (1.7)	25.0 (2.0)	Graph type: 0.78	1, 14	.392	.057
	7	24.4 (1.8)	25.5 (2.4)	Indicator: 0.81	2, 28	.431	.059
	9	24.9 (1.9)	26.7 (1.8)	Interaction: 0.60	2, 28	.535	.044
Downward saccades (%)	5	23.3 (2.2)	30.4 (2.6)	Graph type: 63.03	1, 14	<.001	.829
	7	23.0 (2.1)	30.4 (2.3)	Indicator: 0.03	2, 28	.955	.003
	9	22.2 (2.2)	31.1 (2.6)	Interaction: 0.71	2, 28	.706	.052

Note. p-values based on two-tailed tests, α = .05.

7.1 | Method

7.1.1 | Stimuli and procedure

For Experiment 2, the stimuli from Experiment 1 were used again, but two changes were introduced to avoid ceiling effects in performance: First, the assignment of letter labels to shares was now randomized (instead of being sorted alphabetically from largest to smallest share). Second, the additional visual anchors (dashed lines) were removed from all graphs. At the beginning of each trial, a question regarding the comparison of two indicators was displayed. Specifically, participants were asked to indicate the larger one of two given segments (e.g., "Which segment is larger, R or Z?"). The relevant indicators were chosen in a way that an equal distribution of differences between two indicators was achieved across all trials (i.e., there was an equal amount of trials referring to differences smaller than 5%, 5–9%, 10–14%, etc.).

7.1.2 | Design and data analysis

The independent variables in this experiment were graph type (pie chart vs. CCW-tree maps), number of indicators per graph (5, 7 or 9), and the size of the difference between the relevant two indicators as an index of comparison difficulty (for statistical purposes, this variable was dichotomized into differences <5% and ≥5 percentage points). The dependent variables in this experiment were the mean accuracy (%) and RTs; 2 × 3 × 2 repeated measure ANOVAs were computed unless otherwise indicated.

7.2 | Results and discussion

Regarding response accuracy (error percentage), the ANOVA revealed a significant main effect of graph type, $F(1, 14) = 5.95$, $p < .05$, $\eta_p^2 = .298$, indicating an accuracy advantage for CCW-tree maps (mean error rate = 2.36%) over pie charts (mean error rate = 5.03%). There was also a significant main effect of the size of the difference between indicators, $F(1, 14) = 9.42$, $p < .05$, $\eta_p^2 = .402$, suggesting a decrease in accuracy (i.e., increase in error rate) when comparisons involved small differences and thus greater difficulty (from mean error rate = 1.80% to 5.58%, see Figure 4). There was no significant main effect of the number of indicators, $F < 1$. None of the two-way (or three-way) interactions were statistically significant, all *ps* > .18. Nevertheless, visual inspection of Figure 4 indicates that participants especially appear to encounter problems in pie charts when small differences were involved. To further investigate this issue, we went back to the raw data in order to find specific examples of graph layouts that produced strong accuracy differences between corresponding pie charts and CCW-tree maps. Figure 5 displays two pairs of such stimuli (Examples 2a and 2b; note that each pair was based on the same data set).

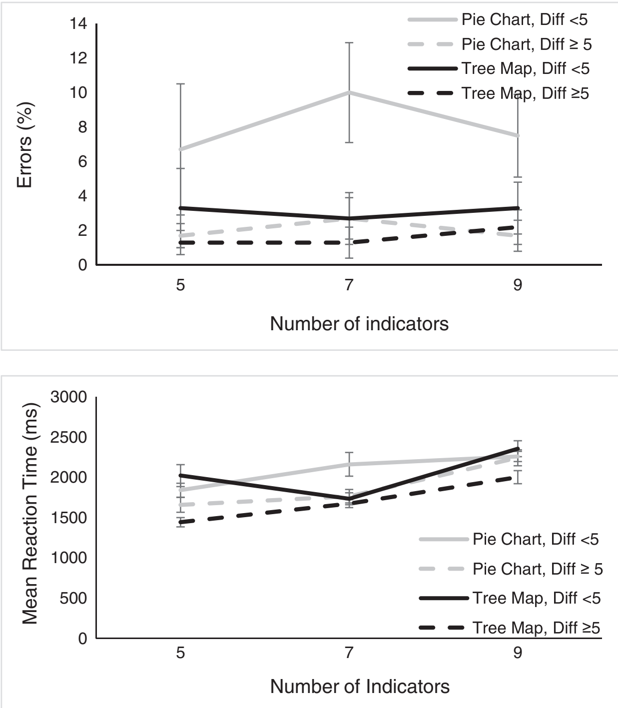

FIGURE 4 Accuracy and response time results in Experiment 2 as a function of graph type, number of indicators, and numerical difference ("Diff") between relevant shares. Error bars represent standard errors of the mean

Interestingly, even though the pie chart segments were generally size ordered (in clockwise fashion), participants obviously did not use (or learn) this regularity to come up with correspondingly accurate comparison judgments. As expected during our design of the CCW-tree maps, it appears to be easier for participants to compare the respective segment length in CCW-tree maps than to compare the size of the pie segments (or the length of the curved outer lines). The present observations also suggest that size ordering of pie segments does not automatically facilitate comparison judgments.

Regarding RTs, the ANOVA revealed a significant main effect of graph type, $F(1, 14) = 10.61$, $p < .01$, $\eta_p^2 = .431$, indicating that CCW-tree maps were processed faster (M = 1873 ms) than pie charts (M = 1990 ms; see Figure 4). There was also a significant main effect of the number of indicators, $F(2, 28) = 33.90$, $p < .001$, $\eta_p^2 = .708$. Specifically, an increase of the number of indicators also led to an increase in RTs (M = 1,742, 1,835, and 2,216 ms). Finally, we observed a significant effect of the size of the difference, $F(1, 14) = 38.05$, $p < .001$, $\eta_p^2 = .731$, indicating prolonged RTs for the comparison of small (vs. large) differences (2,064 vs. 1,799 ms). There was also a significant interaction of graph type and the number of indicators, $F(2, 28) = 3.94$, $p < .05$, $\eta_p^2 = .220$, reflecting the observation that the CCW-tree map advantage was especially pronounced in the seven indicators condition (average advantage of 260 ms; see Figure 4). All other interactions were not significant, all $ps > .10$, except for the three-way interaction, $F(2, 28) = 7.03$, $p < .01$, $\eta_p^2 = .334$. Taken together, the RT results are nicely in line with the accuracy data and, in line with our hypotheses, suggest a substantial CCW-tree map benefit over pie charts.

7.2.1 | Eye movement analyses

The eye movement analyses were carried out in the same way as in Experiment 1 (see Table 2). As in the previous experiment, overall oculomotor effort was increased for the processing of pie charts versus CCW-tree maps, as indicated by an increased number of fixations on the graph, prolonged saccade amplitudes, and greater horizontal and vertical dispersion of fixations. Note that both dispersion measures were numerically higher than in Experiment 1, most likely because the instructions in the present experiment explicitly focus on two spatially separated indicators (instead of just one). Similar to Experiment 1, we again found that CCW-tree maps were scanned with fewer leftward saccades but more downward saccades. Additionally, we observed more rightward saccades and fewer upward saccades for CCW-tree maps, which might hint at a more reading-like scanning pattern. In sum, these analyses again indicate greater oculomotor effort associated with the more demanding comparison processes for pie charts than for CCW-tree maps.

Finally, we further analyzed the number of gaze transitions between the relevant to-be-compared shares for the two stimulus variants referred to above (see Figure 5, Examples 2a and 2b), which yielded strong accuracy disadvantages for pie charts (note that only participants who contributed uncompromised eye movement data in the respective trials were considered). Although for the first stimulus variant we found a similar number of gaze transitions between the to-be-compared shares for both graph types (pie chart: 3.5, tree map: 3.4, $t < 1$), there were more gaze transitions for the pie chart (vs. the tree map) in the second variant (3.4 vs. 2.1, $t(11) = 2.11$, $p = .029$ using a one-tailed test). This post hoc analysis suggests that at least in some cases, difficult comparison processes (here: for pie charts) go along with an increased number of gaze transitions between the relevant shares in the graph.

7.2.2 | Correlational analyses

Further analyses revealed no significant correlation of RTs and number of errors ($r = .18$, $p = .52$ in the CCW-tree map condition and $r = -.20$, $p = .47$ in the pie chart condition). Given the random assignment of indicator letters to shares, we did not analyze the correlation of RT and stimulus position (see Experiment 1, where this analysis was more informative but revealed no significant effect). Instead, we here analyzed whether participants responded faster when the smaller indicator was named first in the question, which was not the case ($r = .07$, $p = .80$). As in Experiment 1, there was no evidence for decreasing RTs over the course of the experiment ($r = -.04$, $p = .89$).

8 | EXPERIMENT 3: NATURAL LABELING

The third experiment compared CCW-tree maps and pie charts in a more realistic setting to assess the extent to which the CCW-tree map advantage still holds in more complex scenarios. Participants solved the same task as in Experiment 2 (comparison judgments), but instead of random letters, we utilized graphs involving meaningful labels from various semantic fields for the indicators. Based on the overall procedural similarity to Experiment 2, the method section will selectively focus on the relevant differences.

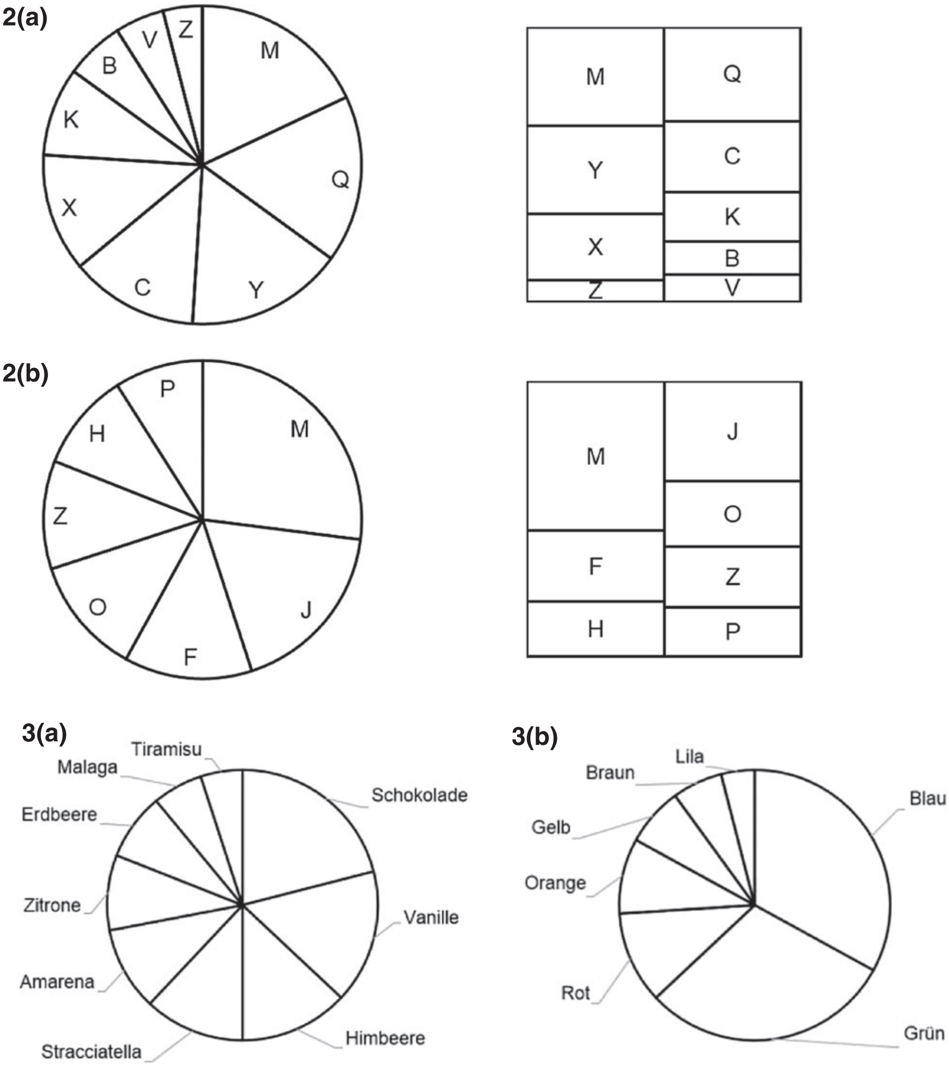

FIGURE 5 Critical stimulus examples in Experiments 2 and 3 yielding a strong comparison accuracy disadvantage for pie charts (vs. tree maps). In the upper two examples (a,b) from Experiment 2, participants mistakenly judged shares "X"/"O" as being larger than "C"/"F," respectively. In the examples from Experiment 3 (a,b), participants mistakenly judged shares related to "Grün"/"Zitrone" as being larger than "Blau"/"Amarena"

8.1 | Method

8.1.1 | Stimuli

Participants were asked to indicate the larger one of two given segments (e.g., "Which one is more popular: green or blue?"; "Which one is more popular: onion or broccoli?"). The indicators in each stimulus were associated with meaningful labels (e.g., types of fruits, vegetables, noodles, and bread). Unlike the previous experiments and more in line with common usage, labels for pie charts were located outside of the graph with a marker line connected to the respective share, whereas we utilized direct labelling for CCW-tree maps. Note that a direct labeling of the pie chart segments (see Ratwani et al., 2008; Gillan et al., 1998; Kosslyn, 1994, for benefits of direct labelling) seemed unfeasible as it would have required rotated, hard to read labels. Examples of the stimuli are depicted in Figure 5.

8.1.2 | Procedure

Due to the complexity involved in designing the graphs (each graph displayed a different semantic field), this experiment consisted of 24 trials only. Twelve of the 30 data sets from the previous experiments were used (four with five indicators, four with seven indicators, and four with nine indicators). There were two comparison tasks for each data set (one as a CCW-tree map, one as a pie chart).

8.1.3 | Design and data analysis

The independent variables in this experiment were graph type (pie chart vs. CCW-tree map) and the number of indicators per stimuli (five, seven, or nine). Due to the small set of trials, we did not include the size of the difference (which varied in a similar manner as in the previous experiment) as an additional independent variable in the design. The dependent variables were the mean number of errors per trial and RTs; 2 × 3 repeated measure ANOVAs were used for statistical analyses unless otherwise indicated.

8.2 | Results and discussion

The accuracy analysis revealed a significant main effect of graph type, $F(1, 14) = 19.31$, $p < .001$, $\eta_p^2 = .580$. In fact, not a single error was committed in the CCW-tree map condition, whereas on average 11.11% errors occurred during pie chart processing (see Figure 6).

TABLE 2 Results of eye movement analyses in Experiment 2

Dependent variable	Indicators	Pie charts: Mean (*SE*)	Tree maps: Mean (*SE*)	*F*	*df*	*p*	η_p^2
Number of fixations (*N*)	5	7.2 (0.3)	6.7 (0.3)	Graph type: 26.83	1, 14	<.001	.674
	7	8.2 (0.3)	7.0 (0.2)	Indicator: 20.41	2, 28	<.001	.611
	9	9.9 (0.7)	8.8 (0.3)	Interaction: 0.66	2, 28	.460	.048
Mean fixation duration (ms)	5	279 (9.8)	280 (9.3)	Graph type: 11.94	1, 14	.004	.479
	7	274 (9.1)	293 (14.7)	Indicator: 3.84	2, 28	.048	.228
	9	263 (8.3)	279 (9.6)	Interaction: 1.68	2, 28	.211	.115
Mean saccade amplitude (°)	5	4.12 (0.19)	3.46 (0.16)	Graph type: 116.04	1, 14	<.001	.899
	7	3.85 (0.18)	3.14 (0.13)	Indicator: 31.78	2, 28	<.001	.710
	9	3.74 (0.15)	3.03 (0.11)	Interaction: 0.17	2, 28	.815	.013
Horizontal fixation dispersion (*SD* of x coordinates per trial)	5	60.0 (2.3)	53.8 (1.7)	Graph type: 60.25	1, 14	<.001	.823
	7	64.4 (2.1)	55.9 (1.9)	Indicator: 14.99	2, 28	<.001	.535
	9	67.4 (2.0)	55.8 (1.5)	Interaction: 4.56	2, 28	.025	.260
Vertical fixation dispersion (*SD* of y coordinates per trial)	5	75.5 (4.2)	66.5 (3.6)	Graph type: 62.24	1, 14	<.001	.827
	7	75.2 (4.0)	57.4 (3.1)	Indicator: 11.94	2, 28	<.001	.479
	9	79.6 (3.6)	62.0 (2.7)	Interaction: 15.80	2, 28	.001	.549
Mean horizontal (x) coordinate (px)	5	473 (18)	508 (17)	Graph type: 202.12	1, 14	<.001	.940
	7	469 (17)	503 (17)	Indicator: 9.79	2, 28	.001	.429
	9	466 (18)	509 (15)	Interaction: 3.88	2, 28	.039	.230
Mean vertical (y) coordinate (px)	5	379 (15)	426 (18)	Graph type: 204.96	1, 14	<.001	.940
	7	394 (17)	429 (16)	Indicator: 17.13	2, 28	<.001	.568
	9	394 (17)	428 (17)	Interaction:10.74	2, 28	.001	.452
Leftward saccades (%)	5	23.9 (1.5)	19.6 (1.2)	Graph type: 20.10	1, 14	.001	.607
	7	25.9 (1.5)	24.4 (1.4)	Indicator: 11.41	2, 28	< .001	.467
	9	27.0 (1.2)	21.4 (0.9)	Interaction: 3.73	2, 28	.046	.223
Rightward saccades (%)	5	20.8 (1.2)	24.7 (1.1)	Graph type: 43.60	1, 14	<.001	.770
	7	22.7 (1.0)	27.7 (1.2)	Indicator: 10.65	2, 28	.001	.450
	9	21.6 (1.0)	24.3 (0.9)	Interaction: 1.11	2, 28	.346	.078
Upward saccades (%)	5	31.8 (1.7)	24.3 (1.3)	Graph type: 55.60	1, 14	<.001	.811
	7	28.3 (1.4)	21.7 (1.3)	Indicator: 8.09	2, 28	.002	.383
	9	29.5 (1.2)	25.2 (1.3)	Interaction: 3.29	2, 28	.054	.202
Downward saccades (%)	5	23.4 (1.0)	31.4 (1.2)	Graph type: 48.44	1, 14	<.001	.788
	7	23.2 (1.0)	26.3 (1.7)	Indicator: 7.74	2, 28	.003	.373
	9	22.0 (1.1)	29.1 (1.3)	Interaction: 4.67	2, 28	.019	.264

Note. p-values based on two-tailed tests, α = .05.

There was also a significant main effect of the number of indicators, $F(2, 28) = 8.38$, $p < .001$, $\eta_p^2 = .374$: Similar to Experiment 2, errors increased with increasing number of indicators. The interaction effect was also significant, $F(2, 28) = 8.38$, $p < .001$, $\eta_p^2 = .374$, suggesting that the effect of the number of indicators was only present for pie charts (mean error rates = 1.67%, 16.67%, and 15.00% for five, seven, and nine indicators, respectively).

As in Experiment 2, we further looked at those graphs that were associated with substantial accuracy differences between both graph types (Figure 5, Examples 3a and 3b). Again, the errors associated with the pie charts occurred despite the fact that the indicators were size ordered. The size difference between the indicators was below five percentage points in these trials, which seems to be a prerequisite for the occurrence of these specific errors.

Because this particular phenomenon was not explicitly investigated in previous research, we can only speculate about potential causes of these mistakes. Probably, the position of the pie chart segments in relation to visual anchors may play a role here. For example, Gillian and Callahan (2000) suggested that in pie graphs, performance is strongly related to the difference between the size of the target segment and the closest-sized anchor. In line with this observation, we found that the wrongly chosen indicators were typically located at anchor positions (even though these anchors were not visibly depicted) and extended towards both directions of the anchor. Probably, this configuration elicits the illusion of larger segment size. Note, however, that this explanation does not readily hold for all examples (e.g., see Experiment 2). Another observation is that throughout all salient trials, the wrongly judged indicator was located at the bottom left of the pie chart, which may hint towards position-dependent attention deployment that may lead to wrong judgments. Taken together, further research is clearly needed to follow up on the mechanisms of these performance deficits (but see the final part of eye movement analyses below).

Regarding RTs, the ANOVA only revealed a significant main effect of the number of indicators, $F(2, 28) = 31.70$, $p < .001$, $\eta_p^2 = .694$ (M = 2,052, 2,387, and 3,269 ms for five, seven, and nine indicators, respectively), but not of graph type, $F(1, 14) = 2.38$, $p = .15$. There was no significant two-way interaction, $F(2, 28) = 1.37$, $p = .27$ (see Figure 6). Altogether, the RT data in Experiment 3 do not compromise the interpretation of the accuracy data. Thus, the more realistic setting in Experiment 3 generally confirmed a performance advantage for CCW-tree maps over pie charts.

8.2.1 | Eye movement analyses

The eye movement analyses were carried out in the same way as in the previous experiments (see Table 3). Again, we observed increased

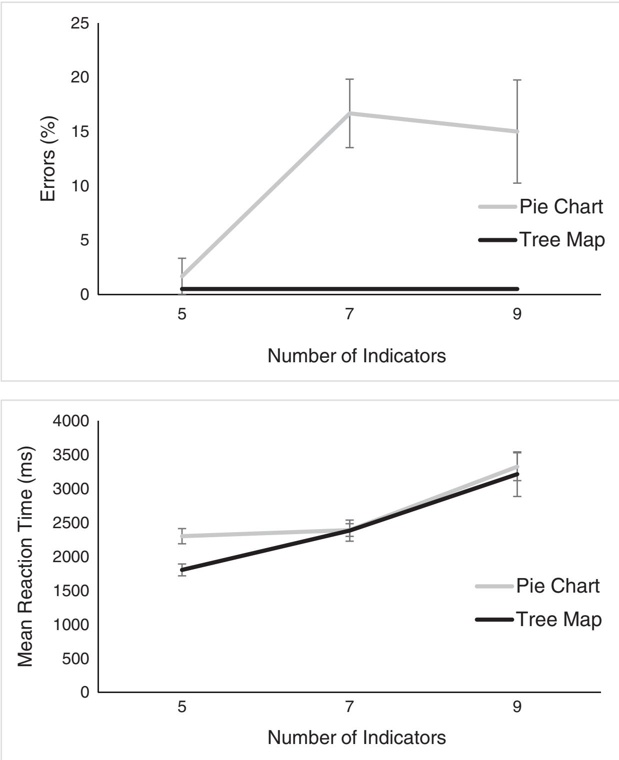

FIGURE 6 Accuracy and response time results in Experiment 3 as a function of graph type and number of indicators. Error bars represent standard errors

overall oculomotor effort for the processing of pie charts versus CCW-tree maps, as indicated by an increased number of fixations on the graph, substantially prolonged saccade amplitudes, and much greater horizontal and vertical dispersion of fixations. The substantial effects on saccade amplitudes and fixation dispersion are clearly attributable to the fact that the pie charts were labeled outside the segment area, which empirically supports the assumption that this suboptimal labeling method cannot be compensated for by drawing on, for example, parafoveal processing abilities. As in Experiments 1 and 2, we again found that CCW-tree maps were scanned with fewer leftward saccades but more downward saccades. In sum, the eye movement analyses in this experiment support the assumption of an additional disadvantage of pie charts (vs. tree maps) with respect to labeling under more realistic conditions. In line with the reported processing difficulties associated with shares at the bottom left of pie charts, we found that mean fixation positions were shifted accordingly more to the bottom and to the left (compared to CCW-tree maps, see Table 3).

Finally, we further analyzed the number of gaze transitions between the relevant to-be-compared shares for the two stimulus variants referred to above that yielded strong accuracy disadvantages for pie charts vs. tree maps (see Figure 5, Examples 3a and 3b). Although, for the first example, we did not find a significant difference in the mean number of gaze transitions between the pie chart (2.1) and the tree map (1.5), $t < 1$, there were significantly more gaze transitions for the pie chart (1.85) versus the tree map (0.93) in the second example, $t(12) = 2.41$, $p = .016$ (one-tailed test). Again, this post hoc analysis suggests that, at least in some cases, difficult comparison processes (here: for pie charts) go along with an increased number of gaze transitions between the relevant shares in the graph.

8.2.2 | Correlational analyses

As in the previous experiments, we computed the correlation between RT and accuracy, but only for pie charts (because there was no variance in the error data for tree maps). As a result, and in line with the observations in the previous experiments, there was no significant correlation of RTs and accuracy for the pie charts ($r = -.05$, $p = .86$). Further analyses (e.g., learning effects) were not considered meaningful due to the low number of trials in this experiment.

9 | EXPERIMENT 4: UNEVEN DATA SETS

This final experiment was designed to address a few remaining issues. First, we asked whether the superiority of CCW-tree maps over pie charts is still present when using data sets that cannot be split into two sets of 50%, thus with figures that can no longer be designed as perfect rectangles. Second, another alternative to pie charts would be the implementation of a single stacked bar graph. This design option would be similar to CCW-tree maps (the latter consisting of two columns) except for the constraint that all bar segments are stacked within one column. If the presence of a one-dimensionally defined feature (straight line length in constant orientation) as a code of the relevant information is the main reason for the superiority of CCW-tree maps over pie charts (see our reasoning in the introduction), stacked bar graphs should also result in better performance than pie charts. To answer these issues, we set up a final (noneye tracking) experiment.

9.1 | Method

9.1.1 | Participants

Twenty participants (16 female, 19 right-handed, mean age = 27; 11 years, $SD = 8.6$) took part in the experiment. Most of them were students enrolled in Psychology. All participants were fluent German speakers.

9.1.2 | Stimuli, procedure, and design

The previous experiments showed that the disadvantage of pie charts is especially pronounced when the difference in segment size is small. Thus, in this experiment—which in general was comparable to the previous experiments—we only implemented the small difference condition. Stimuli were designed based on 20 data sets. Each data set was combined with three different comparison tasks, which always related to small differences in segment size (<5%). The resulting 60 combinations were displayed as CCW-tree maps, stacked bar graphs, and pie charts, respectively (white on black background, see Figure 7a–c). The overall size of the graph types and the number of shares (seven) were held constant. Segments were randomly ordered, that is, their position was not ordered with respect to segment size. Together, there were 180 trials (equivalent to a duration of 30 min for the entire experiment). The experiment was programmed using the software PsychoPy. The only independent variable was graph type (CCW-tree map, stacked bar, and pie). Independent variables included RTs and

TABLE 3 Results of eye movement analyses in Experiment 3

Dependent variable	Indicators	Pie charts: Mean (*SE*)	Tree maps: Mean (*SE*)	*F*	*df*	*p*	η_p^2
Number of fixations (*N*)	5	10.3 (0.4)	8.0 (0.4)	Graph type: 12.27	1, 14	.004	.467
	7	11.6 (0.4)	10.0 (0.6)	Indicator: 35.90	2, 28	<.001	.719
	9	14.5 (0.7)	13.1 (1.0)	Interaction: 0.29	2, 28	.682	.020
Mean fixation duration (ms)	5	219 (6.9)	238 (7.0)	Graph type: 28.40	1, 14	<.001	.670
	7	220 (4.6)	254 (8.5)	Indicator: 2.88	2, 28	.078	.171
	9	217 (6.3)	229 (5.7)	Interaction: 2.17	2, 28	.134	.134
Mean saccade amplitude (°)	5	5.94 (0.22)	3.77 (0.14)	Graph type: 435.79	1, 14	<.001	.969
	7	5.86 (0.23)	3.55 (0.17)	Indicator: 22.90	2, 28	<.001	.621
	9	5.10 (0.16)	3.26 (0.12)	Interaction: 2.46	2, 28	.109	.150
Horizontal fixation dispersion (*SD* of x coordinates per trial)	5	144 (4.2)	67 (2.1)	Graph type: 885.87	1, 14	<.001	.984
	7	143 (4.2)	66 (1.7)	Indicator: 0.10	2, 28	.896	.007
	9	140 (3.8)	71 (2.0)	Interaction: 1.88	2, 28	.174	.119
Vertical fixation dispersion (*SD* of y coordinates per trial)	5	102 (4.5)	77 (3.6)	Graph type: 190.01	1, 14	<.001	.931
	7	121 (5.1)	69 (3.1)	Indicator: 7.46	2, 28	.003	.348
	9	110 (4.4)	77 (3.1)	Interaction: 44.99	2, 28	<.001	.763
Mean horizontal (x) coordinate (px)	5	446 (4.6)	515 (5.3)	Graph type: 59.88	1, 14	<.001	.810
	7	502 (9.1)	513 (4.8)	Indicator: 7.36	2, 28	.008	.344
	9	485 (9.1)	508 (7.3)	Interaction: 26.93	2, 28	<.001	.658
Mean vertical (y) coordinate (px)	5	358 (5.6)	381 (6.3)	Graph type: 167.82	1, 14	<.001	.923
	7	383 (5.7)	408 (6.3)	Indicator: 21.03	2, 28	<.001	.600
	9	393 (9.4)	419 (10.3)	Interaction:0.18	2, 28	.814	.013
Leftward saccades (%)	5	31.5 (2.1)	21.6 (2.0)	Graph type: 13.97	1, 14	.002	.499
	7	31.3 (1.4)	26.9 (1.8)	Indicator: 4.15	2, 28	.036	.229
	9	28.5 (1.5)	22.1 (1.9)	Interaction: 1.60	2, 28	.226	.103
Rightward saccades (%)	5	28.6 (1.5)	24.3 (1.8)	Graph type: 2.79	1, 14	.117	.166
	7	28.9 (1.3)	27.3 (2.2)	Indicator: 4.76	2, 28	.020	.254
	9	23.9 (1.0)	23.9 (0.9)	Interaction: 1.61	2, 28	.219	.103
Upward saccades (%)	5	20.9 (1.4)	25.9 (2.3)	Graph type: 0.05	1, 14	.834	.003
	7	21.1 (1.4)	17.9 (2.2)	Indicator: 4.51	2, 28	.026	.244
	9	24.3 (1.7)	23.5 (1.8)	Interaction: 5.52	2, 28	.012	.283
Downward saccades (%)	5	19.1 (1.1)	28.2 (1.7)	Graph type: 37.62	1, 14	<.001	.729
	7	18.6 (0.9)	27.9 (2.0)	Indicator: 3.93	2, 28	.051	.219
	9	23.4 (1.1)	30.4 (2.2)	Interaction: 0.46	2, 28	.597	.032

Note. p-values based on two-tailed tests, α = .05.

error rates. Additionally, we assessed which graph type participants subjectively preferred after the experiment.

9.2 | Results and discussion

The subjective preference rating resulted in 45% preference votes for pie charts, 40% for CCW-tree maps, whereas only 15% preferred the stacked bar variant. This preference distribution was statistically nonrandom (p = .002 based on a Chi square test), suggesting that stacked bar graphs were clearly less preferred than the other options. An ANOVA for RTs revealed no significant effect, $F(2, 38) = 1.852$, $p = .171$ (M = 2871 ms, SE = 145 for CCW-tree maps, M = 3006 ms, SE = 110 for stacked bar graphs, M = 2949 ms, SE = 122 for pie charts, respectively). However, there was a significant effect on error rates, $F(2,38) = 4.890$, $p = .02$, $\eta_p^2 = .205$. Mean error rates amounted to 18.5% (SE = 1.11) for CCW-tree maps, 19.2% (SE = 1.12) for stacked bar graphs, and 21.9% (SE = 1.22) for pie charts. Post hoc (nondirectional) pairwise comparison tests revealed that performance for pie charts was significantly worse than for CCW-tree maps (p = .029) and stacked bar graphs (p = .003), whereas there was no significant difference between CCW-tree maps and stacked bar graphs (p = .57). Taken together, the error data corroborated the main finding from the previous experiments, namely, better performance for CCW-tree maps than for pie charts. Notably, stacked bar graphs did not perform significantly worse than CCW-tree maps, probably due to the fact that performance in both types of representation rely on the same one-dimensional (straight line length in constant spatial orientation) comparison operations. Interestingly, a comparison with the data pattern in Experiment 2 suggests overall longer RTs, higher error rates, and a smaller error rate difference between pie charts and CCW-tree maps. Probably, the decision to focus only on very small depicted share size differences and the need to switch between three different graph types has substantially increased overall task difficulty in Experiment 4 (compared with Experiment 2).

10 | GENERAL DISCUSSION

This study experimentally addressed the mechanisms underlying integration processes during the comprehension of frequency graphs. Based on theoretical considerations, we developed a constrained variant of the increasingly popular tree map as a potential alternative to the ubiquitous pie charts for the display of (relative) frequency data. We hypothesized that these tree maps avoid some well-known major drawbacks associated with pie charts (i.e., especially those that hamper integration processes during graph processing) while maintaining or

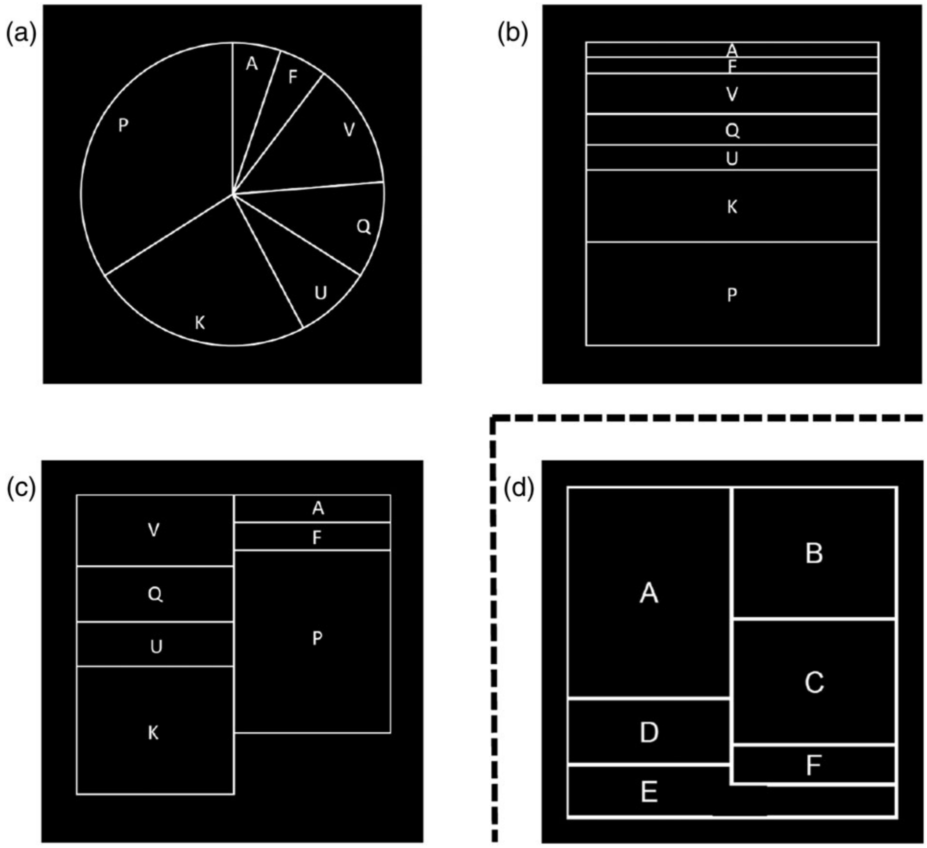

FIGURE 7 Stimuli for uneven data sets used in Experiment 4 (pie chart: a; stacked bar graph; b, constant column width tree map: c) and an alternative design solution (not implemented in the present experiment) maintaining a rectangular shape (d)

even improving (via the enhanced possibility of direct labelling) their overall space effectiveness. Previous research regarding the usefulness of pie charts has produced rather mixed results, and researchers that highlighted the shortcomings of pie charts usually recommended the use of bar graphs instead (e.g., Cleveland & McGill, 1984).

However, some studies suggested that bar graphs may not always be superior to pie charts (e.g., Hollands & Spence, 2001; Spence & Lewandowsky, 1991), which might be one of the reasons why pie charts are still very popular. Crucially, our newly developed tree map variant was empirically evaluated to test the extent to which the associated facilitation of integration processes actually translates into measurable performance benefits under different (prototypical) task demands. Thus, our theory-driven optimization of graph design should have advantageous effects on performance that should ideally outweigh a remaining important advantage of pie charts, namely, the overall familiarity of graph readers with the pie chart format (e.g., Pinker, 1990; Shah & Carpenter, 1995, for a discussion of effects of previous experience with graphs on performance).

To address these issues, we compared pie charts and tree maps in a set of four experiments. Participants' performance was evaluated in tasks involving proportion (Experiment 1) and comparison (Experiments 2, 3, and 4) judgments under both highly controlled (Experiments 1, 2, and 4) and more realistic (Experiment 3) circumstances. Additionally, we measured participants' eye movements (except for Experiment 4) to achieve first insight into oculomotor control underlying frequency graph comprehension in general and performance differences between both graph types in particular.

Overall, the results showed no clear performance differences between tree maps versus pie charts for proportion judgments. This result was probably to be expected because proportion judgments should be less influenced by integration processes than task demands that explicitly refer to specific comparisons between selected indicators. The latter was at stake in Experiments 2, 3, and 4, where we observed substantial performance benefits for CCW-tree maps over pie charts in line with our expectations. Additional analyses of eye movements confirmed that the performance advantage of the rectangular design went hand in hand with decreased oculomotor effort in terms of number of fixations, saccade amplitudes, and spatial fixation dispersion. Furthermore, our post hoc analysis of selected items involving the comparison of small differences revealed a tendency towards fewer gaze transitions between the two to-be-compared shares for CCW-tree maps than for pie charts. The analysis of saccade directions revealed further consistent differences between the scanning of the two graph types: CCW-tree maps were scanned in a more reading-like pattern, involving fewer leftward and more downward saccades compared to pie charts.

Overall, the results additionally suggest that the CCW-tree map advantage especially comes into play for graphs containing a larger number of indicators, because most of the CCW-tree map benefits in Experiments 2 and 3 were smallest in the condition with the smallest set of indicators. Furthermore, the post hoc analyses of those graphs that resulted in pronounced error rates in Experiments 2 and 3 suggested that comparison judgments especially suffer for pie charts when the difference in pie segment size is very small.

It is interesting that the effect of graph complexity (number of indicators) on accuracy tended to point into opposite directions for proportion judgments (Experiment 1) and comparison judgments (Experiments 2 and 3): Although proportion judgments appeared to be facilitated with an increasing number of indicators, comparison judgments rather suffered. Probably, the extra time needed to find two relevant indicators (vs. just one) for comparison judgments explains the performance decrement when more indicators are present, whereas proportion judgments might generally benefit from the presence of a more graphically structured graph layout that comes along with the presence of more indicators, which may serve as additional visual anchors for proportion judgments.

Our observation that the processing advantage for graphs with lower demands on integration processes (i.e., CCW-tree maps) is especially pronounced for complex graphs (i.e., those involving many indicators) nicely fits with previous research by Ratwani et al. (2008), who postulated that integration can be subdivided into visual integration (using perceptual features to build visual clusters) and cognitive integration (higher-level comparison of clusters). Crucially, they argued that integration processes were more demanding as visual graph complexity increased.

The purpose of use usually determines which display layout represents the most compatible choice (Sparrow, 1989; Vessey, 1991, 1994; Washburne, 1927; Wickens & Andre, 1990; Wickens & Carswell, 1995). Tables are used for communicating exact data (Meyer, 2000), line graphs for trends (Bryant & Tversky, 1999), and bar graphs for identifying maxima (Meyer, Shinar, & Leiser, 1997) or simple contrasts (Bryant & Tversky, 1999). Based on our results, we can conclude that CCW-tree maps are particularly effective for the judgment of small differences and especially in more complex graphs involving many indicators.

10.1 | Pie charts: Pros and cons

Disadvantages of pie charts include the difficulty to compare surface areas (or the length of the curved circle segment) of pie slices (see Carswell, 1992; Cleveland & McGill, 1984; Kosslyn, 1994), which are unusually complex geometrical figures involving both curved and straight lines. Second, it has been argued that pie slice comparisons may involve mental rotation. Finally, pie charts come with a disadvantage regarding the extraction of the referents related to the slices due to font alignment issues. Specifically, it appears difficult to implement an easy-to-read, upright direct labeling of slices, which is why pie slices are often labeled outside the actual pie region. Experiment 3 revealed that the latter option comes at the cost of increased oculomotor effort in terms of a substantial increase of both horizontal and vertical fixation dispersion, suggesting that placing labels outside of the pie slices cannot be compensated for by means of parafoveal processing.

These drawbacks of pie charts are countered by some advantages. These include anchoring benefits (e.g., Chandrasekaran & Lele, 2010; Simkin & Hastie, 1987; Spence & Lewandowsky, 1991), which, however, can also be provided in tree maps (see Experiment 1). Second, pie charts are known to provide benefits for proportion judgments at least when compared to nonstacked bar graphs under some conditions (Hollands & Spence, 2001). However, our present results did not suggest an advantage of pie charts over tree maps, probably due to the fact that the organization of tree map data is similar to pie charts in that all elements are displayed as part of a whole, simple geometric figure (here: a rectangle).

A third advantage of pie charts refers to the possibility of size ordering (e.g., clockwise), a feature that is not in the same way possible with tree maps. Therefore, information about the relative position of a pie segment can principally be used as a cue for comparison judgments. However, it is interesting to note that despite the fact that our pie chart stimuli were size ordered, participants obviously did not use this cue to avoid errors in the comparison tasks (Experiments 2 and 3). Thus, the potential advantage of size ordering in pie charts did not transfer to corresponding performance benefits in our study, where no explicit instructions regarding size ordering were given.

A final advantage of pie charts refers to their greater familiarity, that is, the fact that participants have more experience with pie charts than with other forms of frequency graphs. Expectations with respect to graph layout are known to affect performance. Specifically, the availability of fitting schemes to read graphs (Carpenter & Shah, 1998; Pinker, 1990) should have provided some advantage for the more frequent type of graph in our study, namely, the pie charts. Thus, it is especially interesting to notice that the CCW-tree maps nevertheless were generally associated with better performance in our study.

10.2 | Comparison to previous studies

Although some research suggested that bar graphs are superior to pie charts (Cleveland & McGill, 1984; Eells, 1926), other studies reported advantages for pie charts, for example, when combinations of indicators are compared (Spence & Lewandowsky, 1991), or for relative proportion judgments (Hollands & Spence, 2001). Further research suggested that proportion judgments were as accurate for pie charts as for bar graphs (Simkin & Hastie, 1987). Our own results suggest that some of the potential benefits of pie charts can largely be retained while avoiding many of the disadvantages, namely, by using CCW-tree maps.

10.3 | Mechanisms of integration processes during the comprehension of pie charts and tree maps

In the following, we will further focus on the mechanisms underlying integration processes especially in the comparison tasks in Experiments 2 and 3, where the data suggested strong evidence for more difficult integration processes for pie charts than for CCW-tree maps. According to Ratwani et al. (2008), visual integration processes are based on similarity regarding perceptual (form, color), semantic (e.g., semantically related areas), or spatial (e.g., proximity) features (see also Liu & Wickens, 1992). In pie charts, similarity is established via the overall similarity of the pie segments, whereas in the CCW-tree maps, it refers to the rectangular shapes and their spatial features. When these visual clusters are related to the referents, further integration is assumed to occur via a comparison–contrast mechanism, which focuses on relevant differences between clusters. A similar mechanism is proposed by Carpenter and Shah (1998), who postulated an interpretation phase of graph comprehension during which arithmetic operations on encoded values and the comparison of spatial relations of indicators are assumed to take place (see also Gillian &

Lewis, 1994). Crucially, they assume that viewers are more likely to focus on the relationship of those data, which are perceptually grouped (e.g., by proximity or by being connected with a line). This grouping appears to be processed in a more efficient manner in CCW-tree maps than in pie charts.

One crucial advantage of our present CCW-tree maps is that they utilize only one dimension (height) to represent relevant information, whereas the other (width) is held constant (similar to bar width in typical bar graphs, see, e.g., Cleveland & McGill, 1984; Simkin & Hastie, 1987; Spence, 1990; Friel et al., 2001). This design feature might thus represent a very important advantage when compared to the original tree maps developed by Shneiderman (1992), where rectangle width was variable and thus two spatial dimensions needed to be taken into account for comparison purposes. Our main assumption that one-dimensional coding of the relevant information (straight line length of constant spatial orientation) is the main reason for the CCW-tree map advantage is further corroborated by Experiment 4, where a type of stacked bar graph (associated with the same benefits) also performed better than pie charts, although at the cost of low subjective preference. Although Simkin and Hastie (1987) actually reported very low graph reading performance for stacked bar graphs, it is important to note that they used multiple (instead of single) stacked bar graphs, where it is difficult to compare shares across several bars. This difference likely accounts for the divergent observations regarding performance.

10.4 | Limitations and further research

Task performance involving graphs is largely dependent on the reader's mathematical ability (Friel et al., 2001). In this study, we tested an academic sample (mainly psychology students), which is certainly not representative for a general audience of graph readers (e.g., in school or readers of magazines). Thus, further research is needed to assess the impact of previous knowledge and expertise on the present effects.

One potential limitation of the first three experiments of the current study refers to the fact that we used data sets that could be subdivided into two 50% shares in order to achieve an overall rectangular design. As suggested in Experiment 4, it might be necessary to design tree maps (of equal column length) that deviate from a perfect rectangle in that one column might be slightly longer than the other. Although this may be considered a drawback on aesthetic grounds, the data from Experiment 4 showed that this layout did not counteract the overall readability of the tree maps (which obviously benefit from using constant column width). If the main aesthetic goal for designing a CCW-tree map still is to come up with a perfectly rectangular shape, it could also be possible to include one share (e.g., at the bottom of the graph) that extends across both columns (resulting in a hexagon, see Figure 7D). However, this option would likely make it difficult to compare the size of this specific share to other shares in the graph.

Another limitation is related to the manipulation of the number of indicators. The chosen three values (5, 7, and 9) certainly do not allow us to generalize our results to graphs with either a smaller or larger set of indicators. However, our data clearly suggest that this variable plays an important role for performance, and thus, it appears rewarding to study this variable more extensively in the future.

Finally, one might ask why one should not recommend the of use simple (nonstacked) bar graphs instead of CCW-tree maps. Normal bar graphs should have an advantage over both pie charts and CCW-tree maps specifically for comparison processes, because readers can utilize a common scale during comparison (e.g., the y-axis when bars are aligned along the horizontal x-axis; e.g., Cleveland & McGill, 1985). However, such an alignment may potentially have unknown side effects in the context of other graph reading tasks (e.g., proportion judgments), but further research is certainly necessary to address these open issues empirically.

10.5 | Summary and outlook

In this study, we addressed cognitive processes underlying graph comprehension along with the aim to optimize the display of frequency data in a theory-driven manner. By this means, our study continues a research tradition in which graph design is evaluated empirically (e.g., Carpenter & Shah, 1998; Carswell et al., 1991; Fischer, 2000; Huestegge & Philipp, 2011; Körner et al., 2014; Peebles & Cheng, 2003; Ratwani et al., 2008; Shah & Carpenter, 1995; Siegrist, 1996; Spence, 1990; Zacks et al., 1998), which is a necessary precondition to understanding the underlying mechanisms of specific graph advantages in particular and graph comprehension in general. In this context, it was especially informative to study eye movement control during frequency graph comprehension, which turned out to be a viable and sensitive tool to understand the attentional mechanisms underlying differences in overall performance (in terms of RTs and accuracy).

This study is not the first one which proposed an alternative for pie charts (e.g., see Cleveland & McGill, 1984; Spence & Lewandowsky, 1991; Gillian & Callahan, 2000). However, none of the alternatively proposed visualization methods were widely accepted, and eventually, pie charts remained as the most common method for displaying shares and percentages, probably due to their space effectiveness and availability in typical graph software. A reasonable strategy to promote the use of better visualization methods therefore seems to be the development of software plug-ins. The general algorithm framework for the generation of CCW-tree maps is thus provided in the appendix.

Although the findings of this study do not provide unanimous evidence for the superiority of the CCW-tree maps under all conditions and task demands, the use of CCW-tree maps generally appears to be a reasonable option to improve graph comprehension in many contexts by building on an already well-established layout option (tree maps). It should also be noted that even small performance differences found under controlled lab situations might well scale up in the real world, where cognitive resources are usually restricted due to multiple task demands. Thus, even small effects found in the lab may have a strong impact in the real world (where one single error regarding graph reading may have serious consequences in some critical situations). However, this conjecture certainly demands explicit empirical testing in the future. In sum, graph designers should be encouraged to more frequently use CCW-tree maps for displaying percentages and shares instead of pie charts.

ORCID

Lynn Huestegge http://orcid.org/0000-0002-1323-7336

REFERENCES

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

Bruls, M., Huizing, K., & van Wijk, J. J. (2000). Squarified treemaps. In Proceedings of the joint Eurographics and IEEE TCVG Symposium on Visualization (pp. 33–42). Berlin: Springer-Verlag.

Bryant, D. J., & Tversky, B. (1999). Mental representations of spatial relations from diagrams and models. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *25*, 137–156.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, *4*(2), 75–100.

Carswell, C. M. (1992). Choosing specifiers: An evaluation of the basic tasks model of graphical perception. *Human Factors*, *34*(5), 535–554.

Carswell, C. M., Frankenberger, S., & Bernhard, D. (1991). Graphing in depth: Perspectives on the use of three-dimensional graphs to represent lower dimensional data. *Behaviour and Information Technology*, *10*(6), 459–474.

Chandrasekaran, B., & Lele, O. (2010). Mapping descriptive models of graph comprehension into requirements for a computational architecture: Need for supporting imaginary operations. In A. K. Goel, M. Jamnik, & N. H. Narayanan (Eds.), *Diagrams 2010* (pp. 235–242). Berlin: Springer-Verlag.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, *79*(387), 531–554.

Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science*, *229*(4716), 531–554.

Cleveland, W. S., & McGill, R. (1986). An experiment in graphical perception. *International Journal of Man–Machine Studies*, *25*(5), 491–500.

Cleveland, W. S., & McGill, R. (1987). Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society. Series A*, *150*(3), 192–229.

Eells, W. C. (1926). The relative merits of circles and bars for representing component parts. *Journal of the American Statistical Association*, *21*(154), 119–132.

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making*, *27*(5), 672–680.

Fischer, M. H. (2000). Do irrelevant depth cues affect the comprehension of bar graphs? *Applied Cognitive Psychology*, *14*(2), 151–162.

Fischer, M. H., Dewulf, N., & Hill, R. L. (2005). Designing bar graphs: Orientation matters. *Applied Cognitive Psychology*, *19*(7), 953–962.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, *32*(2), 124–158.

Gattis, M., & Holyoak, K. J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *22*(1), 231–239.

Gillan, D. J., Wickens, C. D., Hollands, J. G., & Carswell, C. M. (1998). Guidelines for presenting quantitative data in HFES publications. *Human Factors*, *40*(1), 28–41.

Gillian, D. J., & Callahan, A. B. (2000). A componential model of human interaction with graphs: VI. Cognitive engineering of pie graphs. *Human Factors*, *42*(4), 566–591.

Gillian, D. J., & Lewis, R. (1994). A componential model of human interaction with graphs: 1. Linear regression modeling. *Human Factors*, *36*(3), 419–440.

Harris, R. L. (1999). *Information graphics: A comprehensive illustrated reference*. Oxford: Oxford University Press.

Hollands, J. G., & Spence, I. (2001). The discrimination of graphical elements. *Applied Cognitive Psychology*, *15*, 413–431.

Huestegge, L., & Philipp, A. M. (2011). Effects of spatial compatibility on integration processes in graph comprehension. *Attention, Perception & Psychophysics*, *73*(6), 1903–1915.

Huestegge, L., & Radach, R. (2012). Visual and memory search in complex environments: Determinants of eye movements and search performance. *Ergonomics*, *55*(9), 1009–1027.

Huestegge, L., Radach, R., Kunert, H.-J., & Heller, D. (2002). Visual search in long-term cannabis users with early age of onset. *Progress in Brain Research*, *140*, 377–394.

Körner, C., Höfler, M., Tröbinger, B., & Glichrist, I. D. (2014). Eye movements indicate the temporal organisation of information processing in graph comprehension. *Applied Cognitive Psychology*, *28*(3), 360–373.

Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, *3*(3), 185–225.

Kosslyn, S. M. (1994). *Elements of graph design*. New York: W.H. Freeman.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*(1), 65–99.

Liu, Y., & Wickens, C. D. (1992). Use of computer graphics and cluster analysis in aiding relational judgment. *Human Factors*, *34*(2), 165–178.

Lohse, G. L. (1993). A cognitive model for understanding graphical perception. *Human-Computer Interaction*, *8*(4), 353–388.

Lohse, J. (1991). A cognitive model for the perception and understanding of graphs. In Proceedings of the SIGHI Conference on Human Factors in Computing Systems: Reaching through technology (pp. 137–144). New York: Association for Computing Machinery.

Meyer, J. (2000). Performance with tables and graphs: Effects of training and a visual search model. *Ergonomics*, *43*(11), 1840–1865.

Meyer, J., Shinar, D., & Leiser, D. (1997). Multiple factors that determine performance with tables and graphs. *Human Factors*, *39*(2), 268–286.

Peebles, D. (2008). The effect of emergent features on judgments of quantity in configural and separable displays. *Journal of Experimental Psychology Applied*, *14*(2), 85–100.

Peebles, D, & Cheng, P. C.-H. (2001). Extending task analytic models of graph-based reasoning: A cognitive model of problem solving with Cartesian graphs in ACT-R/PM. In Proceedings of the Fourth International Conference on Cognitive Modeling (pp. 169–174). Mahwah: Erlbaum.

Peebles, D., & Cheng, P. C.-H. (2003). Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors*, *45*(1), 28–46.

Pinker, S. (1990). A theory of graph comprehension. In R. Feedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73–126). Marwah: Erlbaum Hillsdale.

Press, G. (2013, September 05). A very short history of big data. Forbes. com. Retrieved from http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/

Ratwani, R. M., Boehm-Davis, D. A., & Trafton, J. G. (2008). Thinking graphically: Connecting vision and cognition during graph comprehension. *Journal of Experimental Psychology Applied*, *14*(1), 36–49.

Renshaw, J. A., Finlay, J. E., Tyfa, D., & Ward, R. D. (2004). Understanding visual influence in graph design through temporal and spatial eye movement characteristics. *Interacting with Computers*, *16*(3), 557–578.

Riechelmann, E., & Huestegge, L. (2018). Spatial legend compatibility within versus between graphs in multiple graph comprehension. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-018-1484-0

Shah, P. (1995). *Cognitive processes in graph comprehension (Unpublished doctoral dissertation)*. Pittsburgh: Carnegie Mellon University.

Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology. General*, *124*(1), 43–61.

Shneiderman, B. (1992). Tree visualization with tree maps: A 2-d space-filling approach. *ACM Transactions on Graphics*, *11*(1), 92–99.

Siegrist, M. (1996). The use or misuse of three-dimensional graphs to represent lower-dimensional data. *Behaviour and Information Technology*, *15*(2), 96–100.

Simkin, D., & Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association*, *82*(398), 454–465.

Sparrow, J. A. (1989). Graphical displays in information systems: Some data properties influencing the effectiveness of alternative formats. *Behaviour and Information Technology*, *8*(1), 43–56.

Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 683–692.

Spence, I., & Lewandowsky, S. (1991). Displaying proportions and percentages. *Applied Cognitive Psychology*, *5*(1), 61–77.

Trafton, J. G., Marshall, S., Mintz, S., & Trickett, S. B. (2002). Extracting explicit and implicit information from complex visualizations. In M. Hegarty, B. Meyer, & H. Narayanan (Eds.), *Diagramatic representation and inference* (pp. 206–220). Berlin: Springer-Verlag.

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire: Graphics Press.

Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, *22*(2), 219–240.

Vessey, I. (1994). The effect of information presentation on decision making: A cost–benefit analysis. *Information Management*, *27*(2), 103–119.

Washburne, J. N. (1927). An experimental study of various graphic, tabular and textual methods of presenting quantitative material. *Journal of Educational Psychology*, *18*(6), 465–476.

Wattenberg, M. (1999). Visualizing the stock market. In Proceedings of ACM CHI, 188–189.

Wickens, C. D., & Andre, A. D. (1990). Proximity compatibility and information display: Effects of color, space and object display on information integration. *Human Factors*, *32*(1), 61–78.

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and its relevance to display design. *Human Factors*, *37*(3), 473–494.

Zacks, J., Levy, E., Tversky, B., & Schiano, D. J. (1998). Reading bar graphs: Effects of depth cues and graphical context. *Journal of Experimental Psychology Applied*, *4*(2), 119–138.

Zacks, J., & Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory & Cognition*, *27*(6), 1073–1097.

How to cite this article: Huestegge L, Pötzsch TH. Integration processes during frequency graph comprehension: Performance and eye movements while processing tree maps versus pie charts. *Appl Cognit Psychol*. 2018;1–17. https://doi.org/10.1002/acp.3396

APPENDIX A.

General algorithm for the generation of CCW-tree maps

Any programming language may be used for implementation, but it may be beneficial to use a combination of HTML5 and JavaScript as this provides the possibility to use canvas technology (e.g., d3.js, chart.js, Google Chart Tools, Plotly, and other chart engines are based on canvas technology). The specifics of the visual style may be configured within the script. Runtime optimization for this general algorithm may be researched.

1. Input data are collected from the input mask into an array [a1].
2. Input data are sorted from largest to smallest and saved into another array [a2].
3. The number of necessary indicators is counted and saved into a variable [v1].
4. The variable [v2] is generated to save the current progress and set to 0.
5. A two-column grid is generated.
6. Variables [v3] and [v4] are generated to save the cumulated shares in the two columns.
7. A while-loop is implemented checking if [v2] < [v1] and drawing the graph.
   a. Value [v2] is drawn from [a2] and locally stored as [v5].
   b. If [v5] ≤ 0.5 and [v3] + [v5] ≤ 0.5, then the area of the indicator is calculated and visually added to the first column and v[3] = [v3] + [v5].
   c. Else if [v5] ≤ 0.5 and [v4] + [v5] ≤ 0.5, then the area of the indicator is calculated and visually added to the second column and v[4] = [v4] + [v5].
   d. Else, [v5] − [v3] is locally stored as [v6] and visually added to the first column. [v3] is set to 0.5. [v5] − [v6] is locally stored as [v7] and visually added to the second column.
   e. Labeling may be added.
   f. [v2] = [v2] + 1.