Discrete Versus Solid: Representing Quantity Using Linear, Area, and Volume Glyphs

MEZGEB TESFAYESUS MIHTSENTU and COLIN WARE, University of New Hampshire

It is common in infographics for quantities to be represented by stacks of discrete blocks. For example, a magazine illustration showing automobile production in different countries might use stacks of blocks with each block representing a thousand cars. This is unlike what is done to represent quantity in the charts used by statisticians, or for quantitative glyphs used in maps. In these cases, solid bars or solid area glyphs such as circles are commonly used to represent quantity. This raises the question of whether breaking bars, area, or volume glyphs into discrete blocks can improve the rapid estimation of quantity. We report on a study where participants compared quantities represented using bar, area, and volume glyphs in both solid and discrete variants. The discrete variants used up to 4, 4×4, and $4 \times 4 \times 4$ blocks or 10, 10×10, and $10 \times 10 \times 10$ blocks for bar, area, and volume, respectively. The results show that people are significantly more accurate in estimating quantities using the discrete versions, but they take somewhat longer. For both areas and volumes, the accuracy gains were considerable.

Categories and Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Evaluation/methodology*; H.1.2 [**Models and Principles**]: User/Machine Systems—*Human information processing*

General Terms: Human Factors

Additional Key Words and Phrases: Glyphs, representing quantity, subitizing, visualization

ACM Reference Format:

Mezgeb Tesfayesus Mihtsentu and Colin Ware. 2015. Discrete versus solid: Representing quantity using linear, area, and volume glyphs. ACM Trans. Appl. Percept. 12, 3, Article 12 (July 2015), 10 pages.
DOI: http://dx.doi.org/10.1145/2767129

1. INTRODUCTION

Symbols that show one or more quantities through their shape, size, color, or other attributes are called *glyphs*; size is probably the attribute most commonly used. Examples of graphics that represent quantity by means of size are bar charts, pie charts, and circle glyphs on maps and other charts. In some cases, quantity is represented by *linear* extent, as in a bar chart glyph. In other cases, *area* is used to represent quantity, as in the case of circle glyphs used to represent the population of cities and placed on a map. We might also consider the use of *volume* glyphs to represent quantity. One reason for using areas and volumes is that a greater range of variation can be represented. Figure 1 compares

This work was supported by NOAA grant NA05NOS4001153.
Authors' addresses: M. T. Mihtsentu, University of New Hampshire, Durham, NH 03824; email: mezgebth@gmail.com; C. Ware, Jere A. Chase Ocean Engineering Lab, 24 Colovos Road, Durham, NH 03824; email: cware@ccom.unh.edu.

DOI: http://dx.doi.org/10.1145/2767129

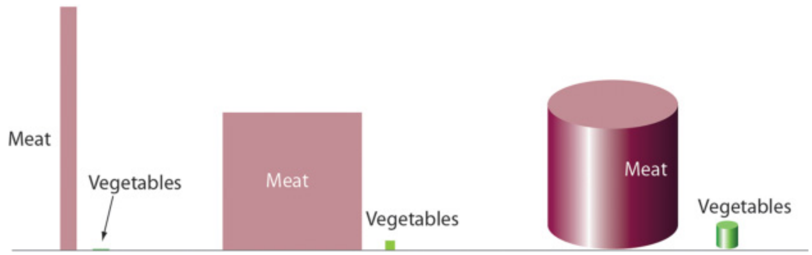

Fig. 1. U.S. federal agricultural subsidies for meat versus vegetables. Quantities 300 and 1 are represented using linear, area, and volume glyphs from Ware [2013].

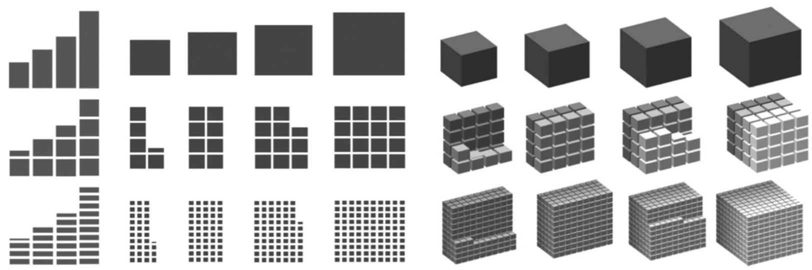

Fig. 2. Ways of representing quantities using solid filled glyphs and discretized variants. From left to right are linear, area, and volume glyphs, respectively. From top to bottom are solid glyphs followed by two variants: discrete 4 and discrete 10.

pairs of objects, one of which is only 1/300th of the size of the other. With a linear scale, the small object is almost invisible, whereas it has significant size for the volume object.

Using concrete objects for quantity representation has a long history of use in infographics. A major example is the International System of Typographic Picture Education (Isotype) of Neurath [1936]. Isotypes support the creation of variations on bar charts using multiple stacked iconic symbols, where each symbol represents an amount and pictorially evokes what is being represented. An example is a grid of small soldier icons, each representing a thousand soldiers killed in a conflict. Countries with greater losses will have more soldier icons to represent them. More recently, Chevalier et al. [2013] discuss what they call *concrete scales*, such as stacks of $100 bills to represent large amounts of money, and suggest guidelines for their use—for example, ways of coming up with reference objects (the Empire State Building has the right scale to be compared to the national debt expressed in $100 bills). Our purpose here is to investigate representations that, although still concrete, lie between the pictographic symbols of Isotype and the abstract-filled rectangles and circles, commonly used in visualizations.

In summer 2012, we began experimenting with schemes for dividing glyphs into discrete blocks, as shown in Figure 2. Independently, Few [2013] suggested that breaking glyphs hierarchically into square blocks would allow more accurate quantity estimation and recommended the technique for use with map displays. These representations raise the possibility that using discrete representations may allow more accurate quantity judgments, and this is the subject of the experiments reported here. But first we discuss the relevant research literature.

2. JUDGING QUANTITES

Experimental results show that humans are not as accurate in judging relative areas as they are in making linear comparisons, and they are much worse in estimating volumes [Croxton and Stein 1932; Clarke 1959; Ekman et al. 1961]. Croxton and Stein [1932] compared bars, circles, squares, and perspective drawings of cubes as the basis for representing magnitudes. They found bars superior to all and areas to be superior to cubes. Clarke [1959] evaluated the adequacy of judgment of quantities in reading glyphs placed on maps. He assessed the accuracy, error, and variability of the estimate of size of glyphs that were 1D (lines), 2D (squares and circles), and 3D (spheres and cubes). Like Croxton and Stein [1932], his findings show that humans are better at making linear judgments and are less accurate when making area or volume judgments.

Several studies have examined the functional relationship between judged size and actual size [Ekman and Junge 1961; Ekman et al. 1961; Teghtsoonian 1965]. Baird et al. [1970] reviewed the literature on judgments of lengths, areas, and volumes. They found that length judgments varied linearly with actual length. The results for area judgments were mixed; some authors found a power relation between actual area and judged area. For example, Teghtsoonian [1965] found the following power relation:

$$Judged\ area = area^{\gamma}, \tag{1}$$

where γ varied between 0.76 and 0.81 for different shapes. However, others found a more linear relationship. For volume judgments, a variety of studies found exponents ranging between 0.59 and 0.79. Ekman et al. [1961] carried out a perceptual study of volumetric symbols used on maps including cubes and spheres. They found perceived size to be proportional to volume raised to the power 2/3 and noted that this is consistent with subjects making their judgments based on image area rather than volume. Cleveland and McGill [1984] concluded from a review of the literature that linear glyphs are perceived most accurately, with areas and volume progressively more inaccurate.

2.1 Anchoring

Some studies of size judgments have found higher accuracy at particular ratios. Especially when one object is exactly half the size of another one, judgments can be more accurate [Baird et al. 1970]. This effect is called *anchoring*. Simkin and Hastie [1987] incorporated the concept of anchoring into a process-based theory of visual comparisons. They proposed a theory of comparisons based on a set of *visual subroutines* (anchoring, scanning, projection, and superimposition) as follows. *Anchoring* is a visual routine by which the observer mentally divides an object into a number of segments, used as anchors. Those anchors provide a rough scale, used in comparisons. For example, the observer may divide the object into four, providing anchor points at 25%, 50%, and 75% of the total size. Then, by applying a *scanning routine*, the observer choses the closest anchor and "sweeps" over the object to make the final estimation of the size. *Projection* is a routine where the observer mentally draws a line from one point of the compared object to the reference object. The observer then estimates the relative size by scanning from the point of projection. If simple projection is not possible, the observer may use an alternative method called a *superimposition routine*, mentally placing the estimated object on top of another object, to make a relative size estimate.

2.2 Subitizing

People, and even some animals, are able to count small numbers of objects "at a glance," and *subitizing* is the term used for this rapid, apparently effortless counting process [Kaufman et al. 1949; Peterson and Simon 2000; Clements 1999]. Estimates of the maximum number of items that can be counted through subitizing vary between three and five [Freeman 1912; Trick and Pylyshyn 1994; Gallistel and Gelman 1991, 1992]. When there are more than the subitizable number of items, they must be estimated, which is prone to error, or counted, which is slow. Subitizing may be useful in estimating quantities represented by discretized glyphs, but only if they are subdivided into five or fewer parts.

2.3 Grouping

Grouping of objects may help in subitizing larger numbers of objects. Freeman [1912] inferred from several experiments that humans have the ability to perceive and correctly judge grouped objects, which are too large to subitize exactly. Freeman also argued that a person might be able to compare two groups of objects without actually counting the number of objects in either group. Beckwith and Restle [1966] found that quantity estimation of objects grouped in regular rows and columns was more accurate than estimation of randomly placed objects. This suggests that estimating quantity can be a somewhat hierarchical process. First, the number of groups may be rapidly subitized and then the quantities within each group may be subitized or otherwise estimated.

3. DESIGN CONCEPT

Our key design concept is to take advantage of human subitizing and anchoring skills in accurately estimating quantities. If glyphs are divided into a small number of parts, subitizing may allow us to effortlessly count the parts. These parts may then be visual anchor points, both from their midpoints and their ends, providing a basis for more accurate judgments. For the area glyphs, we propose that dividing the glyph into a 4×4 matrix may allow for grouping into columns, subitizing the number of columns, then subitizing within a column. Additionally, the number of complete columns may provide anchor points yielding more accurate judgments near 25%, 50%, and 75%.

We were also interested in assessing whether the idea could be extended to three dimensions, using cubes as elements. Although the cognitive subroutines to estimate quantities in this case would be admittedly complex, it would be interesting to know if dividing a volume into rows, columns, and stacks would lead to improvements over the very poor judgments that have been obtained with volume glyphs.

To evaluate the preceding concept, three different versions of linear, area, and volume glyphs were constructed. These are shown in Figure 2. In each case, two different discrete representations were used to assess the role of subitizing in correctly judging quantities. For the linear comparison, a maximum (100%) quantity was subdivided into 4 or 10 rectangular segments. In almost all cases, the quantity could not be represented by an integer number of rectangles and the top rectangle was scaled to represent the remainder. For the area comparison, a maximum (100%) quantity was subdivided into 4×4 or 10×10 segments, represented as rectangles with columns filled from left to right. For the volume comparison, a maximum (100%) quantity was subdivided into $4 \times 4 \times 4$ segments or $10 \times 10 \times 10$ segments, represented as boxes. First, the back bottom row was filled, next the back plane was filled from bottom to top, and then the planes were filled back to front. To increase the differentiation of planes, they were colored a darker grey at the back than the front. Our hypothesis for the first experiment was that discretization of all three types of glyphs would improve accuracy; however, because of subitizing and anchoring, we expected the schemes based on 4 segments to be better than the schemes based on 10 segments. We also hypothesized that in the case of the discrete variants with subdivisions into 4, 4×4, and $4 \times 4 \times 4$ parts, accuracy would be greater at anchoring points (25%, 50%, and 75% of scale) than between anchoring points (12.5%, 37.5%, 62.5%, and 87.5%).

4. EVALUATION

We carried out two experiments to evaluate the alternative representations. In the first experiment, a reference glyph was always present on the left of the display, and targets of various smaller sizes were shown on the right. The purpose was to determine nonlinear scaling effects and any anchoring effects that might exist. In the second experiment, participants compared two randomly sized glyphs to provide a more realistic use case.

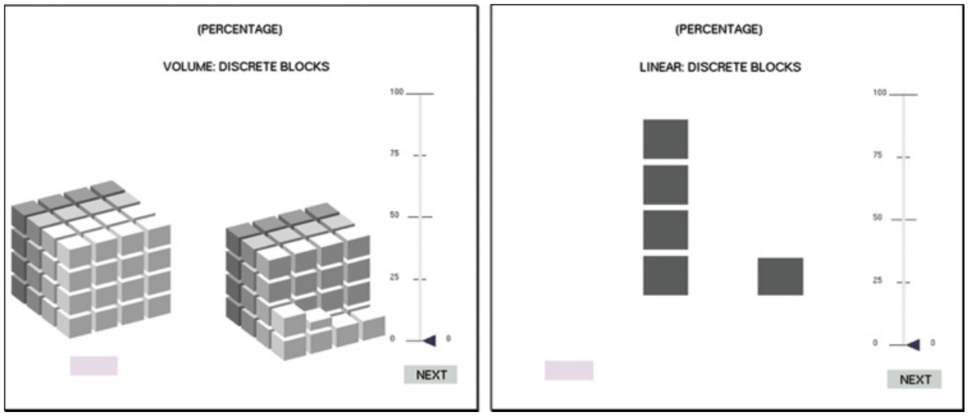

Fig. 3. The interface for the first experiment showing two of the conditions. On the left is the discrete 4 volume condition, and on the right is the discrete 4 linear condition. The participant estimated the size of the right figure to the left figure (100%) in percentage and used the vertical linear slider and the Next button to enter the estimate.

4.1 Experiment 1: Percentage of the Maximum

Participants were asked to judge the size of the glyph on the right-hand side relative to a standard reference glyph on the left-hand side. The left-hand (reference) object always represented 100%. Figure 3 shows two examples.

4.1.1 *Task.* In each trial, a pair of glyphs from one of the nine conditions (detailed later) was displayed for 3 seconds. The size of the right-hand object was randomly determined and varied between 1% and 99% of the left-hand object. Participants used a vertical linear slider bar on the right-hand side of the screen to enter their settings. The participants clicked a Next button provided to proceed to the next trial. The time from the beginning of a given trial's display to the time the Next button was clicked was saved as the response time.

4.1.2 *Subjects.* There were 15 participants, all of whom were undergraduate students paid to participate.

4.1.3 *Conditions.* Included in the experiment were three types of glyph (linear, area, and volume) with three variants on each as follows:

—*Linear*: Solid, discrete 4, discrete 10
—*Area*: Solid 2D, 4 × 4, discrete 10 × 10
—*Solid*: Solid 3D, discrete 4 × 4 × 4, discrete 10 × 10 × 10

This 3 × 3 design yielded nine conditions. Before each experiment, there was a short training session with 9 trials, one for each of the nine conditions. The nine conditions were presented in a random order in blocks of trials. In a trial block, a participant received three trials before proceeding to the next condition. The entire set was repeated four times, yielding 12 trials per condition and a total of 108 estimates from each participant. Estimated sizes and response times were saved together with the actual sizes and the condition.

4.1.4 *Results from Experiment 1.* The average error and time to respond results are summarized in Figure 4. We begin with errors. The linear condition produced the lowest errors overall, followed by area, followed by volume. A repeated measures ANOVA using the Greenhouse-Geisser sphericity

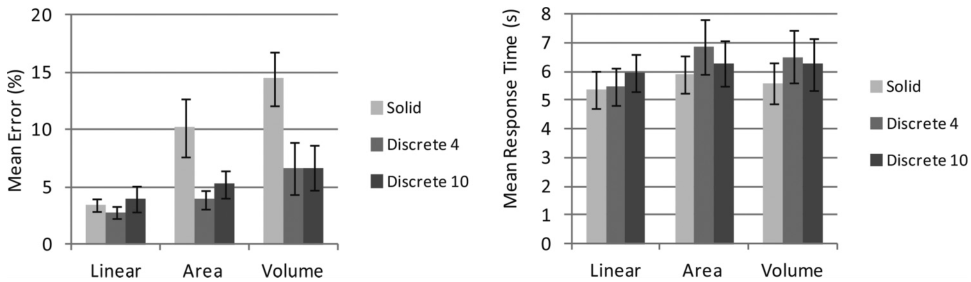

Fig. 4. Left: Average errors for the nine conditions of Experiment 1. Right: Average response times. Bars represent 95% confidence intervals.

correction showed these effects to be highly significant: main effects of the dimension of glyph (linear, area, volume) ($F[1.86, 24.2] = 61.1$; $p < .001$); main effects for type (solid, discrete 4, discrete 10) ($F[1.525, 19.8] = 62.0$; $p < .001$). There was also a highly significant interaction effect ($F[2.46, 31.9] = 10.9$; $p < .0001$). To examine this interaction, separate ANOVAs were run using the Greenhouse-Geisser sphericity correction with Bonferroni comparisons for each of linear, area, and solid glyphs. For the *linear* glyphs, there was no effect of type. For the *area* displays, there was a significant main effect of type ($F[1.57, 20.33] = 29.7$; $p < .001$), and the Bonferroni comparisons showed the discrete 4 and discrete 10 versions to be significantly better than the solid version ($p < .001$). For the *volume* displays, there was a significant main effect of type ($F[1.54, 20.05] = 23.96$; $p < .001$), and the Bonferroni comparisons showed the discrete 4 and discrete 10 versions to be significantly better than the solid version ($p < .001$).

The average times to respond are plotted in Figure 4. Response times were somewhat slower with the discretized glyphs. A repeated measures ANOVA using the Greenhouse-Geisser sphericity correction showed significant effects of glyph type (linear, area, volume) ($F[1.5, 19.6] = 14.8$; $p < .005$) and whether the glyphs were solid, discrete 4, or discrete 10 ([F1.76, 22.9] = 11.6; $p < .005$). There was no a significant interaction.

The relationship between estimated and actual sizes was plotted, and the results are shown in Figure 5. As previous researchers found, there is a nonlinearity in the solid versions. In contrast, the discrete versions appear to be more linear.

To determine if there were anchoring effects, the size errors were plotted against glyph size. To accomplish this, the responses were averaged into bins of width 12.5% based on the actual value (as opposed to the estimated value). Bins were centered on 25%, 50%, and 75%, as well as intervals in between (12.5%, 37.5%, 62.5%, and 87.5%). The results are plotted in Figure 6. To compare the aggregated errors in anchor bins with aggregated errors in nonanchor bins for all nine conditions, t-tests were used. The results were significant for the linear 4 condition ($p < .001$), the area 4 condition ($p < .001$), and the volume solid condition ($p < 005$). However, in the volume solid condition, observation of the curve suggests that this cannot reasonably be taken as evidence for anchoring since the curve has an inverted U shape. Presumably, the significant effect came from the fact that errors were much lower for the 12.5% and 87.5% bins.

4.2 Experiment 2: Ratio of Two Glyphs

The first experiment was somewhat artificial, because in information graphics, people do not typically make comparisons with a standard reference glyph but instead make comparisons between glyphs. In the second experiment, participants were asked to make comparisons between pairs of glyphs, each of which had a randomly determined size, including cases where the estimated object (right-hand side)

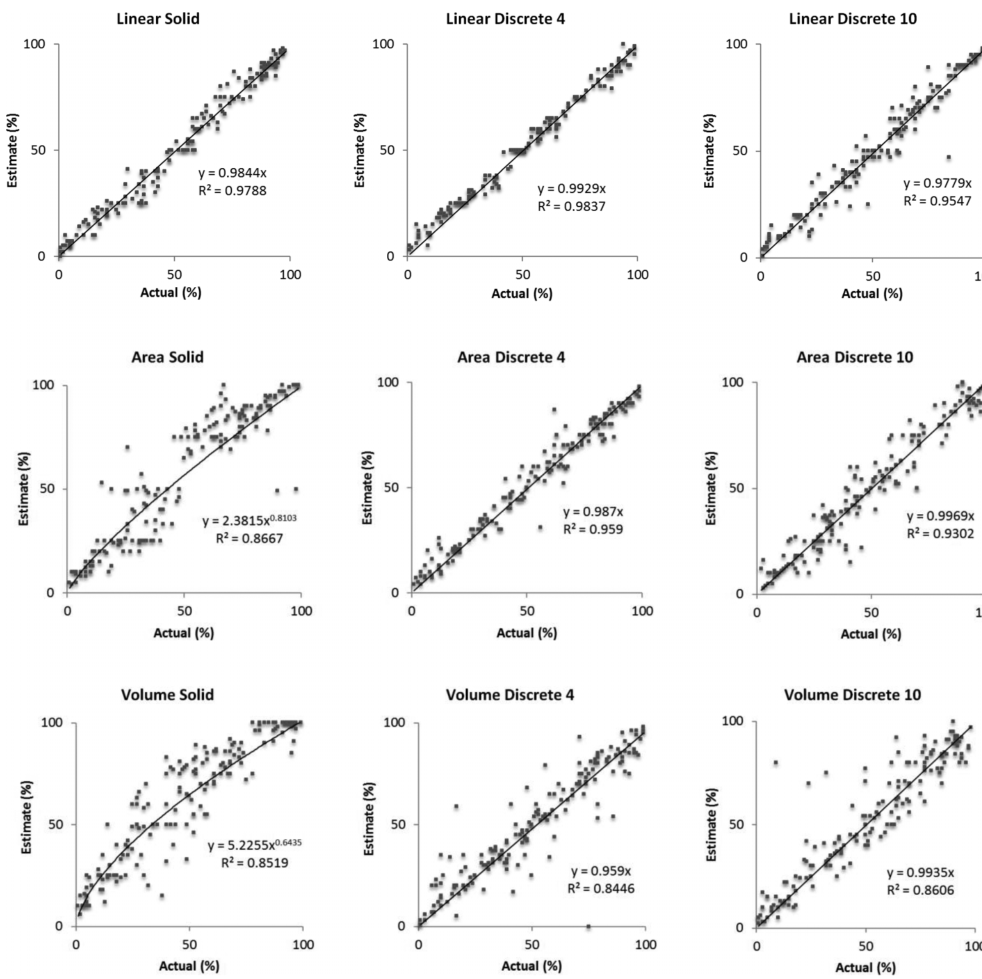

Fig. 5. The relationships between actual and estimated size for all nine conditions.

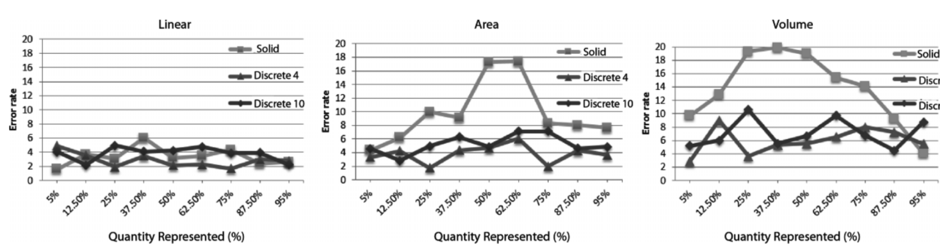

Fig. 6. Errors plotted as a function of quantity represented.

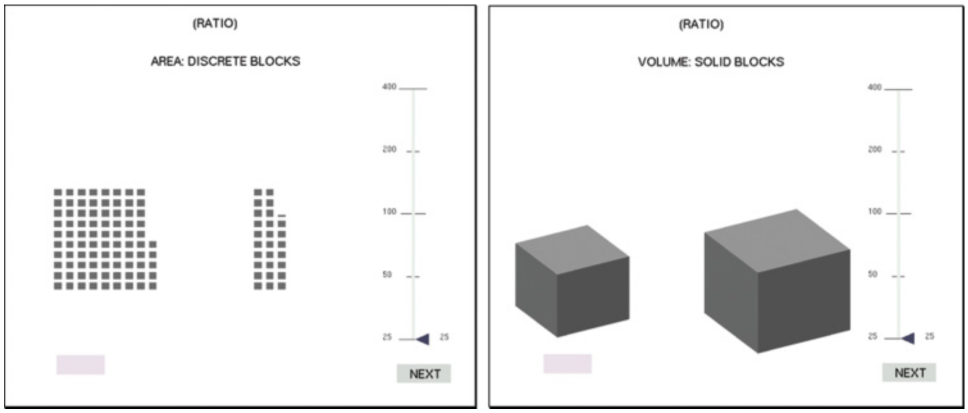

Fig. 7. In both images, participants were asked to estimate the size of the right figure to the left (reference) figure. To the right of the figures are the vertical logarithmically increasing slider and the Next button.

was smaller than the reference object (left-hand side) and cases where the estimated object (right-hand side) was larger than the reference object (left-hand side).

4.2.1 *Task.* In each trial, a pair of objects from one of the conditions was displayed for 3 seconds. Participants were asked to estimate the ratio of the size of the estimated (right-side) object to that of the reference (left-side) object. The estimate was entered using a logarithmically scaled vertical slider ranging from 25% to 400%. The participants clicked a Next button provided to proceed to the next trial. The time from the display of a given trial to the time the Next button was clicked was saved as the response time. Figure 7 shows two example screens.

4.2.2 *Trials.* The nine conditions were presented in a random order in blocks of trials. In a trial block, a participant received 3 trials before proceeding to the next condition. The entire set was repeated four times, yielding 12 trials per condition and a total of 108 responses from each participant.

4.2.3 *Results from Experiment 2.* Because the task was to estimate ratios, we converted the data to logs for the analysis. We also corrected a number of response errors. Participants sometimes did not follow the instruction to always compare the right-hand side object with the left-hand side object. Therefore, where the correct answer should have been 30%, they responded 333%. Since it is implausible that they actually saw something that was much smaller as if it were much larger, we interpreted this as a response error and corrected it with a simple rule:

if ((actual ratio > 110% and estimate size < 90%) OR (actual ratio < 90% and estimate size > 110%)) invert the estimated ratio.

The average error and time to respond results are summarized in Figure 8. A repeated measures ANOVA using the Greenhouse-Geisser sphericity correction showed main effects of the dimension of glyph (linear, area, volume) ($F[1.67, 21.8] = 6.37$; $p < .005$) and main effects for type (solid, discrete 4, discrete 10) ($F[1.91,24.9] = 10.53$; $p < .001$). There was also a highly significant interaction effect ($F[3.30, 42.9] = 9.11$; $p < .0001$). To examine this interaction, separate ANOVAs were run using the Greenhouse-Geisser sphericity correction with Bonferroni comparisons for each of linear, area, and solid glyphs. The Bonferroni comparisons revealed no significant differences for either the linear or area glyphs. There was a highly significant effect for the volume glyphs. For volume glyphs, discrete 4 and discrete 10 versions were significantly more accurate than the solid version ($p < .001$).

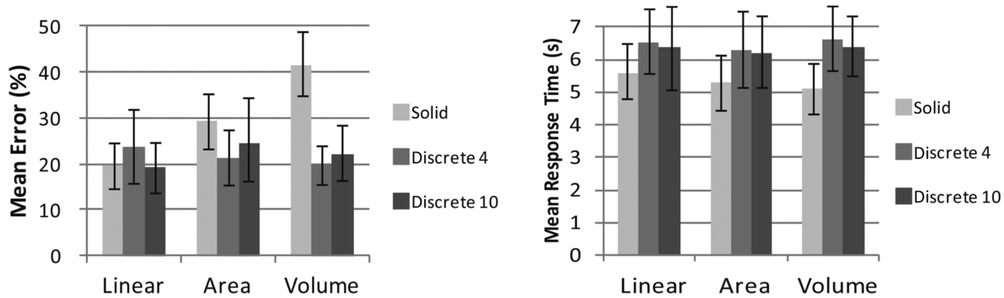

Fig. 8. Left: Average errors for the nine conditions of Experiment 2. Right: Average response times. Bars represent 95% confidence intervals.

A repeated measures ANOVA using the Greenhouse-Geisser sphericity correction of the timing data showed a main effect of glyph dimension ($F[1.43, 18.6] = 32$; $p < .001$). Both area and volume glyphs produced significantly slower responses than solid glyphs according to Bonferroni comparisons.

5. DISCUSSION

Our results suggest that breaking area and volume glyphs into discrete parts can result in significantly more accurate quantity estimation. For the first experiment, the mean error rates for discrete 4×4 area and discrete $4 \times 4 \times 4$ volume glyphs were less than half of the error rates obtained with the solid glyphs. In addition, there was a more linear relationship between the actual sizes and the estimated sizes for both the area and the volume glyphs. For the second experiment, the results suggest that area and volume glyphs could be made as effective as linear glyphs by breaking them into discrete parts. However, both experiments also showed an additional cost in terms of slightly longer response times.

The hypothesis that the discrete 4 glyphs would be better than the discrete 10 glyphs because of subitizing was not supported, as there were no significant differences between them.

There was some evidence for an anchoring effect for the linear 4 and area 4×4 conditions where the errors were somewhat smaller at 25%, 50%, and 75%. This is not altogether surprising, as there were clear visual boundaries at these points. There were also reduced errors near the end points in some conditions, especially the solid volume glyph. This can be accounted for by noting that the solid volume glyph is the one for which the errors are largest overall and sizes near zero and 100% will also be anchored by the end points of the scale.

The main disadvantage of discretized glyphs is that they cannot be as compact as solid glyphs. Clearly, there will be size thresholds below which the individual subcomponents cannot be resolved, and at some point, a discretized glyph will appear very much like a solid glyph. However, the advent of extremely high resolution displays means that pixel size will be less of a factor and possibly quite small discretized glyphs will be usable. Finding the smallest size for which discretized glyphs are effective would be a useful topic for future research.

It should be emphasized that we are not recommending that bar charts should be constructed with discretized bars. We included linear glyphs for comparison, and indeed our results show no significant advantages for discretizing linear glyphs. Even if discretization were to improve performance in a bar graph, a properly constructed scale could perform the same function, as one of the roles of the discretization is to provide a scale (albeit arbitrary).

To summarize, we suggest the use of either 3D or 3D discretized glyphs in infographics where glyph size is less of an issue because the illustrations are usually quite large. Objects broken into discrete parts provide a kind of visual rhetoric; a thousand small cubes looks like a lot, in a way that a single solid cube, rectangle, or bar does not. In infographics, this can be useful because it adds to the visual impact. Discretized glyphs may also be useful as symbols on maps, but in this case, the availability of compact glyphs is likely to be more of an issue.

REFERENCES

J. C. Baird, D. Romer, and T. Stein. 1970. Test of a cognitive theory of psychophysics: Size discrimination. *Perceptual and Motor Skills* 30, 2, 495–501.

M. Beckwith and F. Restle. 1966. Process of enumeration. *Psychological Review* 73, 5, 437.

F. Chevalier, R. Vuillemot, and G. Gali. 2013. Using concrete scales: A practical framework for effective visual depiction of complex measures. *IEEE Transactions on Visualization and Computer Graphics* 19, 12, 2426–2435.

J. I. Clarke. 1959. Statistical map-reading. *Geography* 44, 2, 96–104.

D. H. Clements. 1999. Subitizing: What is it? Why teach it? *Teaching Children Mathematics* 5, 400–405.

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

F. E. Croxton and H. Stein. 1932. Graphic comparisons by bars, squares, circles, and cubes. *Journal of the American Statistical Association* 27, 177, 54–60.

G. Ekman and K. Junge. 1961. Psychophysical relations in visual perception of length, area and volume. *Scandinavian Journal of Psychology* 2, 1, 1–10.

G. Ekman, R. Lindman, and W. William-Olsson. 1961. A psychophysical study of cartographic symbols. *Perceptual and Motor Skills* 13, 3, 355–368.

S. Few. 2013. Building insight with bricks: A new more perceptible method for encoding quantitative values in geospatial displays. *Perceptual Edge Visual Business Intelligence Newsletter*. April/May/June issue.

F. N. Freeman. 1912. Grouped objects as a concrete basis for the number idea. *Elementary School Teacher* 12, 7, 306–314.

C. R. Gallistel and R. Gelman. 1991. Subitizing: The preverbal counting process. *Essays in Honor of George Mandler*, 65–81.

C. R. Gallistel and R. Gelman. 1992. Preverbal and verbal counting and computation. *Cognition* 44, 1, 43–74.

E. L. Kaufman, M. W. Lord, T. W. Reese, and J. Volkmann. 1949. The discrimination of visual number. *American Journal of Psychology* 62, 4, 498–525.

O. Neurath. 1936. *International Picture Language; the First Rules of Isotype: With Isotype Pictures*. Trubner & Company.

S. A. Peterson and T. J. Simon. 2000. Computational evidence for the subitizing phenomenon as an emergent property of the human cognitive architecture. *Cognitive Science* 24, 1, 93–122.

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

M. Teghtsoonian. 1965. The judgment of size. *American Journal of Psychology* 78, 3, 392–402.

L. M. Trick and Z. W. Pylyshyn. 1994. Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review* 101, 1, 80.

C. Ware. 2013. *Information Visualization: Perception for Design* (3rd ed.). Elsevier.

Received April 2015; revised April 2015; accepted April 2015