Reinforcing Visual Grouping Cues to Communicate Complex Informational Structure

Juhee Bae and Benjamin Watson *Member, IEEE*

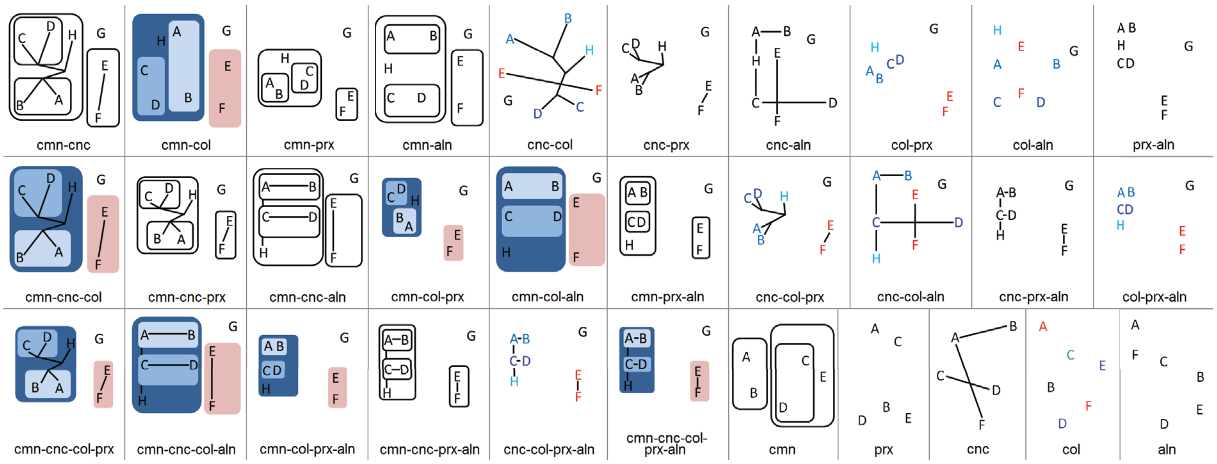

Fig. 1. The five grouping cues used in isolation and in reinforcing combination in our second study. These visuals are simplified for illustration, experimental visuals were more complex. (cmn=common region, cnc=connectedness, col=color similarity, prx=proximity, and aln=alignment)

Abstract— In his book Multimedia Learning [7], Richard Mayer asserts that viewers learn best from imagery that provides them with cues to help them organize new information into the correct knowledge structures. Designers have long been exploiting the Gestalt laws of visual grouping to deliver viewers those cues using visual hierarchy, often communicating structures much more complex than the simple organizations studied in psychological research. Unfortunately, designers are largely practical in their work, and have not paused to build a complex theory of structural communication. If we are to build a tool to help novices create effective and well structured visuals, we need a better understanding of how to create them. Our work takes a first step toward addressing this lack, studying how five of the many grouping cues (*proximity*, *color similarity*, *common region*, *connectivity*, and *alignment*) can be effectively combined to communicate structured text and imagery from real world examples. To measure the effectiveness of this structural communication, we applied a digital version of card sorting, a method widely used in anthropology and cognitive science to extract cognitive structures. We then used tree edit distance to measure the difference between perceived and communicated structures. Our most significant findings are: 1) with careful design, complex structure can be communicated clearly; 2) communicating complex structure is best done with multiple reinforcing grouping cues; 3) *common region* (use of containers such as boxes) is particularly effective at communicating structure; and 4) *alignment* is a weak structural communicator.

Index Terms—Visual grouping; visual hierarchy; gestalt principles; perception; visual communication

1 Introduction

How can people tell that the spots behind those leaves are not unrelated shadows, but a leopard? Psychologists call this ability visual grouping, and have studied it for almost a century. For nearly as long, designers have been exploiting our visual grouping ability to add meaningful structure to their graphics. But almost from its beginning, design practice has also outstripped psychological research, with visuals using combinations of grouping cues to communicate such complex structures that even today, psychologists have not researched them in any great detail.

As the means of visual communication become much more available, and today's information demands make using visual communication well more important than ever, tools for creating visuals are not keeping pace. While users have many tools that make it easy to create visuals, few help them build visuals that communicate clearly.

Communication is much more successful when the information it contains is well organized. How can users organize information visually? Gestalt psychology and its principles of visual grouping [4] give them a starting point, but complex information demands complex structure, and as we make clear below, research on combining Gestalt principles has not yet answered all the questions visual communication raises. If we are to create better tools for visual communication, we need a better understanding of visual communication itself. Our research examines several different combinations of visual grouping cues, and how effectively they communicate complex, hierarchically structured information.

2 Related Work

Our work is inspired by Richard Mayer and his theory of multimedia learning [7], or learning from words and pictures. A crucial assumption of that theory is that humans actively process visuals to construct a coherent mental representation of their experiences. This cognitive

- *Juhee Bae is with North Carolina State University. E-mail: jbae3@ncsu.edu.*
- *Benjamin Watson is with North Carolina State University. E-mail: bwatson@ncsu.edu.*

Manuscript received 31 Mar. 2014; accepted 1 Aug. 2014. Date of publication 11 Aug. 2014; date of current version 9 Nov. 2014. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

Digital Object Identifier 10.1109/TVCG.2014.2346998

processing involves paying attention to select relevant information, organizing that information into a meaningful knowledge structure, and integrating that structure with existing knowledge. Many knowledge structures are possible, but five basic structures are processes describing a cause-and-effect chain, comparisons of elements along several dimensions, generalizations outlining main ideas and supporting details, enumerations compiling sets of related items, and classifications analyzing a domain into subsets. The corresponding structures are flow charts, matrices, trees, lists, and hierarchies. Well designed multimedia, Mayer says, should have a coherent structure, and provide the learner with guidance that helps them build the correct structure.

With their theory of visual hierarchy, graphic designers strive to provide that guidance to viewers. Indeed their work regularly goes beyond the frontiers of perceptual psychology. Yet with their traditional focus on the practical, their theory does not often provide the detailed guidance needed by novice designers.

2.1 Gestalt grouping

In the early 20th century, the Gestalt school of psychology developed several laws describing how people infer structure from what they see, a process they called visual grouping [4]. They include similarity (similar items are grouped), proximity (neighbors are grouped), common fate (similar movements are grouped), symmetry (reflected items are grouped), good continuation (or alignment, with segments of the same curve grouped), and closure (segments of the same container are grouped). Relatively recently, Palmer and Rock [10, 8] introduced connectivity (linked items are grouped) and common region (items in the same container are grouped).

In the natural world, the grouping cues described by Gestalt laws exist in abundance, and often conflict with different cues suggesting competing organizations of our environment. How do people combine multiple cues and resolve those conflicts? We organize the research addressing these questions into three sets: work studying grouping cue combinations including *proximity*, *connectedness*, and *common region*.

Research shows that proximity often dominates similarity [1, 9]. Yet domination depends on cue strength, with proximity weakening as inter-item distance increases, and similarity weakening with reduced viewing time. Kubovy and van den Berg [5, 14] review a great deal of the literature on the simultaneous presence of these two cues and suggest that in fact they operate additively: they reinforce or interfere with one another in direct proportion to their strength.

Han's work [2] showed that reinforcing similarity cues with agreeing connectedness cues made groups much easier to perceive, while combining proximity with connectedness had little effect. Palmer and Beck [9] found that conflicting similarity and connectedness cues made groups harder to perceive. Interestingly however, when they reinforced similarity with connectedness, groups became slightly *more difficult* to perceive. To explain this, Palmer and Beck introduced the notion of *intrinsic* grouping cues, which depend on properties of the grouped items themselves and include similarity (shape and color) and proximity (position); as well as *extrinsic* cues, which depend on other items and include connectedness (linking contours) and common region (surrounding contours). Reliance on extrinsic grouping cues therefore implied additional items and more complex visuals, making groups a bit more difficult to perceive.

Palmer and Beck [9] also found that conflicting similarity and common region cues made groups much more difficult to perceive, while reinforcing cues made groups slightly more difficult to perceive. Recently, Luna and Montoro investigated combinations of common region with proximity or similarity [6]. Conflicting cues made grouping more difficult, while reinforcing cues made grouping easier. Their results supported an additive model for combinations with extrinsic grouping cues, much like the model for intrinsic cues proposed by Kubovy and van den Berg [5].

2.2 Grouping in visualization

Although not extensive, there is work exploring applications of visual grouping in the field of visualization. In his book *Visual Thinking for Design*, Ware discusses widespread use of Gestalt grouping to communicate semantic structure, but he does not discuss the complex combination of grouping cues [15]. Ziemkiewicz and Kosara [18] experimentally verify this mapping of the visual to the semantic, and suggest that it may also be influenced by viewers' physically based interpretation (e.g. gravity) of what they see.

Wattenberg and Fisher built an analysis engine that extracts grouping from imagery [16]. They point out that the structure of visualization should match the structure of the data and convey its intention. Rosenholz et al. [12] improved on Wattenberg and Fisher's analysis engine by generalizing it to sense color and orientation, and making it simpler to extend the tool to detect other perceptual features as well. The tool responded to Gestalt proximity, similarity, and good continuation stimuli much like people would, and was able to convincingly extract structure from text, textures, interface, information graphics, and visualizations. Later, Rosenholz et al. [11] studied the use of the tool in the design process, finding that it could be helpful. However, as Wattenberg and Fisher point out, these algorithmic models need experimental validation.

While the above research is certainly useful, it does not provide a complete solution to the problem of helping novice designers synthesize good visuals. For example, how should grouping cues be combined to communicate complex informational structures containing groups within subgroups? Existing research does not provide a clear answer.

3 Studying disjoint combinations of grouping cues

To begin answering our questions, we performed some experiments examining how Gestalt grouping cues can be combined to communicate complex informational structures. Our first experiment examined a disjoint approach, with each cue communicating different subsets of the structure.

3.1 Methods

We limited the complexity of the information structures we attempted to communicate to hierarchies containing two grouping levels: groups (outer) and groups-within-groups (inner). In these hierarchies, any leaf is an item to be organized, any internal node directly descended from the root is the parent of an *outer group*, and any internal node within an *outer group* is the parent of an *inner group*. With a disjoint approach, one grouping cue communicated the *outer groups*, the other, the *inner groups*. For example, in Figure 4, *common region* communicates *outer groups* (and not inner) while color communicates the *inner groups* (and not outer).

We then used a two-factor 5 *outer grouping cues*$\times$5 *inner grouping cues* between-subjects design, with both independent variables *outer* and *inner groupings* using one of five grouping cues: *common region*, *connectedness*, *color similarity*, *proximity*, and *alignment*. Figure 2 illustrates these five grouping cues in disjoint combination. 250 people recruited through Amazon Mechanical Turk (117 male, 133 female, aged from 18 to 73), participated in the experiment online. Heer and Bostock [3] found that Amazon Mechanical Turk can produce reliable results in graphical perception experiments, as long as concerns of overlap, completion and quality are addressed. Our study addressed these concerns by recording when Turkers participated in multiple experiments, rewarding them only when tasks were completed, and requiring them to perform verifiable pre-qualification tasks. All Turkers had normal or corrected-normal vision and passed a color-blindness test. They were paid $1.50 for their effort and 10¢ as bonus for every correct answer.

3.1.1 Apparatus and Stimuli

Our online test environment displays a visual on the left and hierarchical dialogue on the right (Figure 3). At the beginning of each trial, we fill the dialogue with elements representing the items in the visual (usually text). Participants can input the hierarchical structure they perceive by creating one or more grouping folders, and moving the elements in the dialogue using their mouse. Participants can also remove empty folders and expand or collapse the hierarchy as desired.

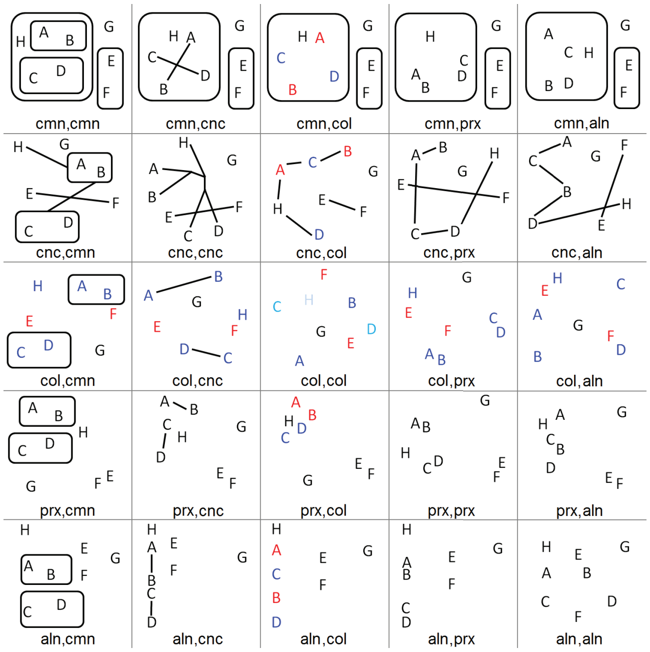

Fig. 2. The five grouping cues used in disjoint combination in our first study (disjoint combination indicated in ‘outer grouping cue, inner grouping cue’). These visuals are simplified for illustration, experimental visuals were more complex. (cmn=common region, cnc=connectedness, col=color similarity, prx=proximity, and aln=alignment)

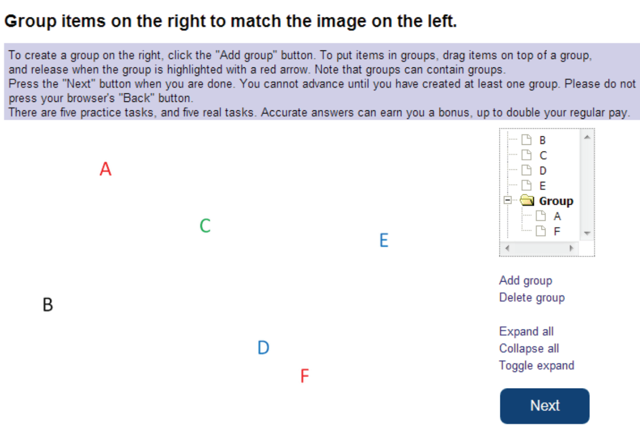

Fig. 3. Grouping letters with same colors (A and F) in a practice task in our first study

Because participants used their own equipment, we did not have direct control of their display, but we told participants an 800×600 pixel monitor was a minimum.

We drew our experimental information content and visual stimuli from five real world examples: a warning sign, a toothpaste product label, and three captured websites: a Google search's results, a Facebook wall, and an Amazon product category. Two experimenters viewed these examples, recorded the structure they perceived in independently, and then discussed and resolved their differences to reach agreement on the information structures depicted in the examples. For each example, we then extracted no more than 20 items (limiting task complexity), and without changing the items themselves, created 25 new stimuli depicting the portion of the information structure containing those items (Figure 4).

The five practice stimuli viewed by all participants (last five stimulus of Figure 1) illustrate how we applied Gestalt cues to communicate visual structure. Here the items were simple letters, while the same grouping cue was applied at both the *outer* and *inner* levels. Each stimulus used a different grouping cue. For example, the *common region* practice stimulus places A and B in one outer group and C, D, and E in another; with C and D in an inner group.

To maintain internal validity, we attempted as much as possible to apply each cue in isolation. Because some cues are naturally related, this could lead to some awkward-looking stimuli. *Alignment* may be the most telling case: without *proximity*, not all items will be close together, making grouping much more difficult to perceive. Although this reduced external validity, we felt it a useful trade-off in this initial experiment.

3.1.2 Procedure

Participants began the experiment by performing five practice trials. If they successfully completed three of those trials, they performed five recorded trials, each depicting a different real world example. These were ordered for each participant using two 5×5 Latin squares. We determined which five of 25 grouping treatments a participant saw — and by varying their order, limited any fatigue or transfer effects — by assigning them to one row in each Latin square, with one Latin square ordering the cues applied for the *outer grouping*, and the other ordering the cues applied for the *inner grouping*. In this way, every participant saw every grouping cue used for *outer grouping* and for *inner grouping*. With 125 participants performing five recorded trials each, we collected 625 user described structures, or 25 trials per disjoint combination of grouping cues.

We first obtained informed consent from the participants, and presented them with written instructions describing their grouping task. In particular, we ensured that participants understood that groups could contain groups, and that a trial was not complete until they created at least one group. They then performed five color deficiency tests. If they completed all five correctly, they performed five practice trials, with the correct answer being displayed after each. If they performed at least three practice trials correctly, participants were asked to provide their age and gender. They then performed five recorded tasks, without any feedback. On average, participants needed 30 minutes to complete the experiment, with about 20 minutes dedicated to the recorded tasks.

As dependent measures for each trial, we collected the informational structure input by the participant — our focus in this paper. We measured the correctness of that structure by obtaining the tree edit distance [17] between the depicted and the participant's structures, which counts the operations (e.g. add, delete) needed to transform one tree to another.

3.2 Hypotheses

How should two grouping cues be combined to communicate a two-level structure (containing groups with subgroups) successfully? With

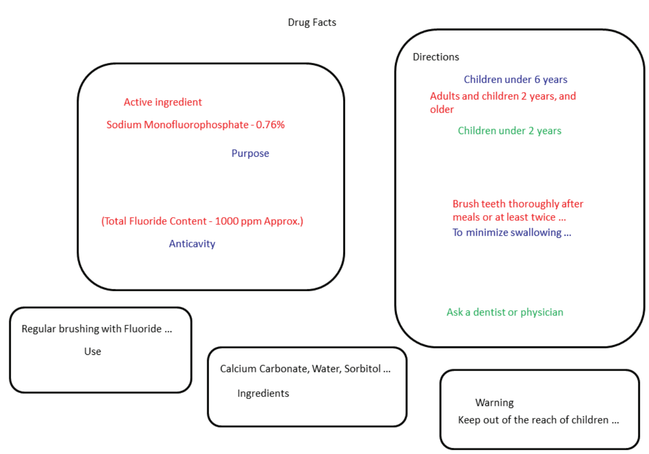

Fig. 4. Toothpaste label stimuli with common region (containment) communicating the outer groups, and color similarity in the inner groups

disjoint grouping cues (each cue communicating different groups), how does a viewer know that one grouping cue applies at the outer level, and the other at the inner? We believe this cue-to-level mapping is a crucial component of successful structural communication with disjoint grouping cues, and two of our three hypotheses focus on it.

Hypothesis 1: Using the same grouping cue at both inner and outer levels will usually communicate poorly. Communicating two grouping levels is difficult with only one cue. With intrinsic cues (*proximity*, *alignment* and *similarity*), spatial and color resolution is limited. Extrinsic cues can add grouping elements (containers or links) as needed, but used in isolation as in this study, *connectedness* can only communicate the cue-to-level mapping topologically. Because the presence of a container within a container indicates this mapping so clearly, *common region* may be the lone exception to this rule.

Hypothesis 2: When using two different grouping cues, communication is clearer when the outer cue dominates the inner cue. Since the global knowledge structure indicated by outer groups is more important than the detail indicated by inner groups, it should be communicated as clearly as possible. Since Gestalt research shows that *proximity*, *connectedness* and *common region* often (certainly not always) dominate similarity (e.g. [5]), we expected that treatments using *proximity*, *connectedness* or *common region* on the outer level would be among the most successful.

Hypothesis 3: Alignment is best used to communicate inner groups. Alignment is inherently one-dimensional, limiting its communicative potential. Used in isolation (for example, without *proximity*), it is even more difficult to understand since grouped items may not be adjacent.

3.3 Results and Discussion

Gender and age had no significant main effects, so we do not discuss these further below.

inner \ *outer*	**cmn**	**cnc**	**col**	**prx**	**aln**
cmn	5.16	11.16	13.08	16.66	20.14
cnc	9.12	17.3	13.6	10.86	20.74
col	10.78	17.02	17.24	13.84	19.86
prx	12.26	14.42	17.72	16.38	21.04
aln	13.6	16.1	15.78	19.02	22.88

Table 1. Tree edit distance for each disjoint combination of grouping cues, with lower distances indicating viewers understand structure better. (cmn=common region, cnc=connectedness, col=color similarity, prx=proximity, and aln=alignment)

Table 1 shows average perceived structure correctness for each treatment. We analyzed the effects of the *inner* and *outer grouping cue* variables on perceived structure correctness with a two-factor ANOVA. Table 2 shows the analysis results. Both main effects were significant, indicating that the cues used to communicate outer and inner grouping changed the effectiveness of structural communication. The effects' interaction was also significant, indicating that the specific combination of inner and outer grouping cues used also affected communication effectiveness. For example, the *common region-common region* treatment was most effective at communicating informational structure, and in pairwise comparisons was significantly more effective than any intrinsic-intrinsic treatment. The *alignment-alignment* treatment was least effective in communicating structure, and in pairwise comparisons was significantly less effective than any treatment using *common region* to communicate outer grouping, while its effectiveness did not significantly differ from any other treatment using *alignment* to communicate outer groups.

We found good support for *hypothesis 1*. In pairwise comparisons, none of the *connectedness-connectedness*, *color-color*, *proximity-proximity* or *alignment-alignment* treatments differed from one another significantly, and as a group they were among the least effective at communicating knowledge structure. As we have outlined above, as the treatment that communicated best overall, the *common region-common region* combination was a glaring exception.

Results provided mixed support for our *hypothesis 2*. Since Gestalt research shows that *common region*, *connectedness* and *proximity* often dominate other cues, we expected that treatments using these cues to communicate outer grouping would be most effective. This held true with one surprising exception: treatments using color to communicate outer grouping were just as effective as treatments using *connectedness* and *proximity*. In particular, the *color-common region* and *color-connectedness* combinations communicated well, were not significantly different from one another, and were only marginally different from the most effective treatment, the *common region-common region* combination. In fact, *common region* and *connectedness* cues were the best communicators of inner level grouping overall.

Analysis also revealed strong support for *hypothesis 3*. *Alignment* was the worst communicator of both inner and outer grouping, though it was a better communicator of inner than outer grouping, with the *common region-alignment* combination particularly effective. Clearly *alignment* is difficult to use without reinforcement by other cues.

Overall, participants did well understanding fairly sparse depictions of structure, with inner and outer grouping each encoded by exactly one cue. With one combination of cues (*common region-common region*), they approached a perfect understanding of the information structure.

Two of the three results from our study of disjoint cues remind us of the limits of using grouping cues in isolation. Indeed in graphic design practice, most groups are communicated with multiple, reinforcing cues. As we performed the first study, we made two observations that argue further for reinforcing cues. First, many cues depend on *proximity*: both *proximity* and *alignment* are positional, placing items within *common region*'s containers requires some *proximity*, and *connectedness* is not very effective if the items it links are widely scattered. Second, some cues are "persistent": they can rarely be completely removed from a visual. Consider *proximity* and *alignment*: when items are distant or misaligned, rather than perceiving nothing about their grouping, viewers will understand that the items are *not* related. In contrast, when *common region*, *connectedness* or *color similarity* are not used, viewers make no inferences about grouping.

4 Studying Reinforcing Combinations of Grouping Cues

Therefore in our second study we examined combinations of reinforcing cues, with all cues each communicating the entire knowledge structure. We improved external validity by studying the interaction of these *cue combinations* with visuals that consist primarily of either images or text, in low or high densities.

4.1 Methods

Our methods in this study had many similarities to the previous study. We focus on the differences in method in this section.

We again examined the communication of two-level hierarchical information structures. With the reinforcing approach, one to five cues all communicated the entire structure, including both inner and outer groups. Figures 1 and 2 contrast the reinforcing and disjoint approaches.

independent variable	**ANOVA of edit distance**	
	F	***p***
outer group	F(4,1225)=31.824	p<0.001
inner group	F(4,1225)=6.149	p<0.001
outer×inner	F(16,1225)=2.037	p=0.009

Table 2. Significant main effects on accuracy of structural communication using disjoint combinations of grouping cues, as measured by tree edit distance.

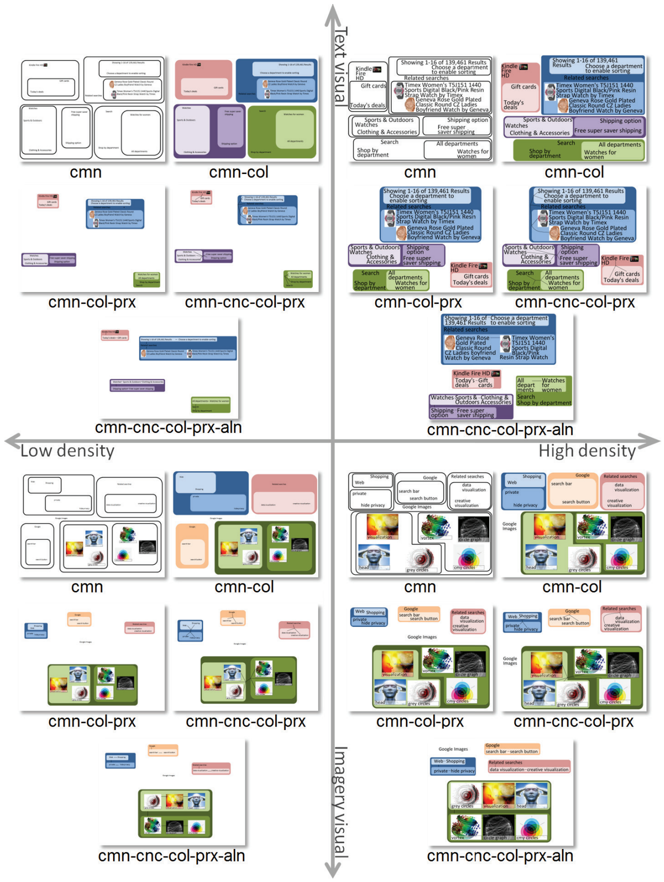

Fig. 5. Some text and imagery visuals in low and high density. Content here shows an Amazon example for the text-dominant visual and Google Images for the image-dominant visual. (cmn=common region, cnc=connectedness, col=color similarity, prx=proximity, aln=alignment)

We used a three-factor 31 *cue combinations* (between)×2 *visual densities* (within)×2 *visual types* (within) mixed design, with the reinforcing *cue combinations* involving the same five grouping cues used in the previous study (2^5 - 1 = 31). We did not include a completely unstructured combination because it would be unrealistic, we could not obtain an accuracy measure for it, we did not want to frustrate the participants, and were confident that significant differences would be found without it.

Visual density indicates the percentage of stimuli area that was occupied by content. For low density, this was 10%, for high, 40% (quite crowded). *Visual type* represents the dominant content of the stimuli: text or imagery. For image-dominant visuals, 85% of pixels describing content were dedicated to images, for text-dominant visuals, 93% of pixels describing content were dedicated to text, where content is visual elements not including grouping elements such as boxes or links. The same web pages used in the previous study (Google search results, a Facebook wall, and an Amazon product category) were our text dominant stimuli, while a Netflix home page, a Pinterest board, and a Google Image search result were our image dominant stimuli.

372 people recruited through Amazon Mechanical Turk (178 male, 194 female, aged from 18 to 61), participated in the experiment online. All had normal or corrected-normal vision and passed a color-blindness test.

4.1.1 Apparatus and Stimuli

Our online test environment was very similar to that in the previous stimuli, with two differences: six practice stimuli were either image or text dominant, and participants could also input "how much they liked" the stimuli from 1 — indicating "strongly dislike" to 5 — indicating "strongly like".

For each, we extracted no more than 17 items of content (limiting task complexity), and without changing the items themselves, created 31 new stimuli, one per *cue combination*, depicting the portion of the information structure containing those items (one per *cue combination*, for example, Figure 1).

We constructed low and high density versions of each example by scaling the visual elements (text or imagery) to fill the required percentage of stimuli area. Each stimulus used a different *cue combination*, *visual density*, and *visual type*. Figure 1 illustrates the 31 *cue combinations* with simple letter stimuli, used here only for illustration. The first 26 *cue combinations* depict the same structure: G is not grouped, E and F form an outer group, and all other letters form an outer group with two inner groups.

Practice stimuli contained small collections of plants, animals or places organized into simple two level classification structures. Half were text dominant, half image dominant; half high, half low density. Two used two grouping cues, the other two used three, and the rest used four grouping cues.

4.1.2 Procedure

Participants began the experiment by performing five color tests and then six practice trials. If they successfully completed four practice trials, they performed six recorded trials. Within those six trials, each participant saw all six content examples. Half were low and half were high density; half were text- and half were image-dominant. We determined which *cue combinations* a participant saw by randomly ordering the 31 combinations and repeating the order across participants. Since 31 is not divisible by 6, the order shifted by one trial on every sixth participant and every *cue combination* appeared equal number of times in each trial. *Visual density* and *visual type* were ordered so that each level appeared equally in each trial. With 372 participants performing six trials each, we collected 2232 user described structures, or 18 trials per *cue combination*×*visual density*×*visual type* triple.

Only seven of 372 participants also took part in our first study. On average, participants needed 33 minutes to complete the experiment, with about 22 minutes dedicated to the recorded tasks. As dependent measures for each trial, in addition to input structure, we recorded trial time, and the participant's rating of the stimulus's appeal.

4.2 Hypotheses

How should grouping cues be combined to communicate informational structure successfully? We expected that our study would indicate the following:

Hypothesis 1: Using reinforcing cues improves communication clarity (reduces edit distance). As the number of cues communicating a knowledge structure increases, viewers will perceive a structure more accurately as measured by tree distance, and more easily as measured by time.

Hypothesis 2: As visual density declines, communication clarity will improve. Lower *visual density* leaves more room for structural grouping to operate.

Hypothesis 3: Proximity and alignment will affect structural communication even in their absence. As "persistent" cues, we expected their absence to communicate not a lack of structure, but a different structure. We therefore anticipated that these cues would have significant and particularly strong effects and interactions.

Hypothesis 4: Many cues rely on reinforcing proximity. The dependence of *alignment*, *connectedness* and *common region* on *proximity* leads us to expect strong interactions between *proximity* and each of these other cues.

Hypothesis 5: Text dominant visuals rely more heavily on spatial arrangement than image dominant visuals. Reading requires a very regular spatial processing of text, so we expected a matching interaction between *visual type* and *proximity* and/or *alignment*.

4.3 Results

We performed three analyses. The first per *cue combination* analysis treated each combination as an experimental level. The second reinforcement analysis grouped combinations by the number of cues

cmn	prx	col	cnc	aln	cue combination	edit distance	trial time	preference
on	on	on	on	on	cmn cnc col prx aln	1.31	191.88	3.19
on	on	on	on	off	cmn cnc col prx	2.51	225.65	3.44
on	on	on	off	on	cmn col prx aln	1.65	195.67	3.40
on	on	on	off	off	cmn col prx	1.97	187.07	3.49
on	on	off	on	on	cmn cnc prx aln	1.42	192.60	2.88
on	on	off	on	off	cmn cnc prx	2.44	167.72	3.26
on	on	off	off	on	cmn prx aln	2.46	163.96	3.22
on	on	off	off	off	cmn prx	1.72	178.89	3.07
on	off	on	on	on	cmn cnc col aln	2.6	204.31	3.44
on	off	on	on	off	cmn cnc col	2.46	184.65	3.08
on	off	on	off	on	cmn col aln	2.25	201.29	3.17
on	off	on	off	off	cmn col	1.97	169.42	3.25
on	off	off	on	on	cmn cnc aln	2.83	217.49	3.32
on	off	off	on	off	cmn cnc	1.81	194.24	3.03
on	off	off	off	on	cmn aln	2.78	172.39	3.13
on	off	off	off	off	cmn	2.93	194.13	2.76
off	on	on	on	on	cnc col prx aln	6.99	224.94	2.85
off	on	on	on	off	cnc col prx	3.61	216.72	3.32
off	on	on	off	on	col prx aln	10.31	208.86	2.19
off	on	on	off	off	col prx	9.97	246.28	2.64
off	on	off	on	on	cnc prx aln	12.79	203.43	2.60
off	on	off	on	off	cnc prx	7.42	196.19	2.78
off	on	off	off	on	prx aln	16.15	238.71	2.21
off	on	off	off	off	prx	15.68	265.31	1.99
off	off	on	on	on	cnc col aln	13.61	289.32	2.36
off	off	on	on	off	cnc col	10.88	322.07	2.01
off	off	on	off	on	col aln	20.06	262.64	2.13
off	off	on	off	off	col	20.99	265.10	1.93
off	off	off	on	on	cnc aln	19.99	330.15	1.94
off	off	off	on	off	cnc	16.31	305.49	1.64
off	off	off	off	on	aln	23.06	282.24	1.90
off	off	off	off	off	*N/A*	*N/A*	*N/A*	*N/A*

Fig. 6. Tree edit distance, trial time (in seconds), and preference means for the per cue combination analysis. Darker red, yellow, and blue indicate more accurate, quicker, and preferred forms of communication. (cmn=common region, cnc=connectedness, col=color similarity, prx=proximity, aln=alignment)

they contained, and treated each of those as a level. The third per cue analysis treated each cue as a different experimental variable, with each cue being present or absent. We discuss all main effects, all two-way interactions and some three-way interactions. Gender had no effect on any of our three measures and older people spent more time (F(38,2193)=6.873, p<0.001) completing tasks. Because of the limited nature of these effects, we do not discuss gender and age below.

4.3.1 Per cue combination analysis

Figure 6 shows average perceived structure correctness, trial time, and preference for each *cue combination*. We analyzed the effects of *cue combination*, *visual density*, and *visual type* with a three-factor (31×2×2) ANOVA. We detail significant effects on distance and preferences in Table 3.

Cue combination, *visual density*, and *visual type* all had significant effects on distance (Table 3). Reinforcement universally lowered the distance and improved understanding of the structure. On the other hand, denser (6.95 for low and 8.72 for high density) and image-dominant visuals increased distance, reducing structural understanding (7.38 for text- and 8.29 for image-dominant visuals).

The interaction between *cue combination* and *visual density* was significant by the distance measure (Table 3). Pairwise comparisons showed that when visuals were dense, viewers had a particularly hard time understanding the depicted structures when grouping cues were less reinforced (fewer cues were used).

Cue combination had a significant effect on the time measure (F(30,2108)=6.897, p<0.001), with means for each combination in Figure 6. Reinforcement generally reduced time.

In addition, *cue combination*, *visual density*, and *visual type* all had significant effect on preferences (means Figure 6, ANOVA Table 3). Participants preferred reinforced cues. There was slight preference for lower density (2.88 for low and 2.64 for high density) and a stronger preference for imagery (2.93 for imagery visuals and 2.59 for text). (Even though viewers understood structure in text-dominant visuals better.) The interaction between *visual density* and *visual type* significantly impacted preference. Preferences for lower density were much stronger with text-dominant visuals.

4.3.2 Reinforcement analysis

Our reinforcement analysis grouped *cue combinations* by the number of cues they employed. This created an experimental variable with five levels. Data was unbalanced across this new variable, so we used a compensating (5×2×2) ANOVA. With this analysis, we can measure the effects of reinforcement more directly than in the per *cue combination* analysis.

Reinforcement, *visual density*, and *visual type* all had significant effect on distance (Table 4). The means for each *cue combination* are in Figure 7. The effects of *visual density* and *type* mirrored those in the first study: text-dominant visuals were easier to understand, but image-dominant were preferred; while low density visuals were easier to understand, and also preferred. The new reinforcement variable had the effects one would predict from our per combination study: increasing reinforcement improved structural understanding, and was preferred. It also lowered task times (F(4,2225)=10.49, p<0.001).

4.3.3 Per grouping cue analysis

Our per grouping cue analysis grouped trials by whether or not a certain cue was employed. It created five experimental cue variables with two levels each (used or not). Data was unbalanced across each of these variables since we did not have an all cues absent combination, so we used a compensating (2×2×2×2 ×2×2×2) ANOVA. With this analysis, we can measure the effects of each cue more directly than in the per combination analysis.

independent variable	**ANOVA of edit distance**	
	F	***p***
cue combination	F(30,2108)=75.023	p<0.001
visual density	F(1,2108)=35.945	p<0.001
visual type	F(1,2108)=9.607	p=0.002
cue combination×visual density	F(30,2108)=1.712	p=0.01
independent variable	**ANOVA of preference**	
	F	***p***
cue combination	F(30,2108)=23.152	p<0.001
visual density	F(1,2108)=31.227	p<0.001
visual type	F(1,2108)=63.872	p<0.001
visual density×visual type	F(1,2108)=11.526	p=0.001

Table 3. Significant effects on edit distance and preferences for the per cue combination analysis.

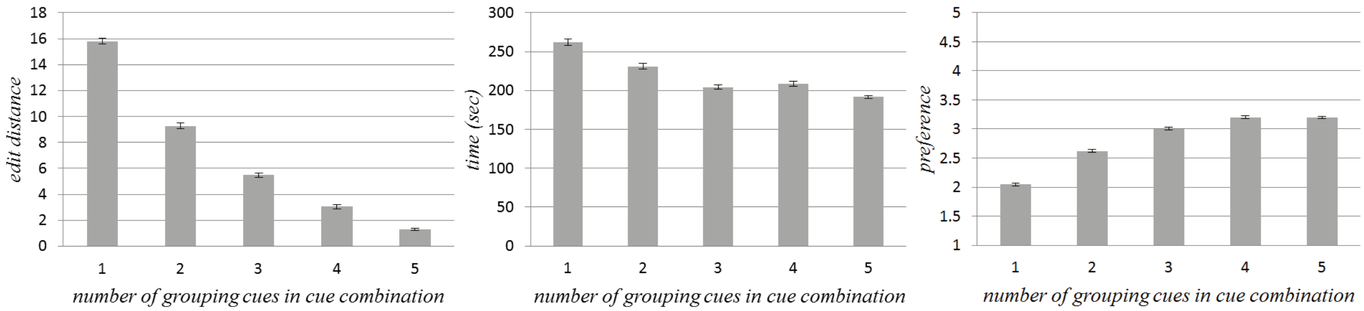

Fig. 7. Distance and trial time decline (i.e. understanding improves) as reinforcement increases. Preference grows with more reinforcement.

independent variable	ANOVA of edit distance	
	F	*p*
reinforcement	F(4,2225)=124.15	p<0.001
visual density	F(1,2225)=21.79	p<0.001
visual type	F(1,2225)=5.82	p=0.016
independent variable	**ANOVA of preference**	
	F	*p*
reinforcement	F(4,2225)=69.36	p<0.001
visual density	F(1,2225)=26.86	p<0.001
visual type	F(1,2225)=54.95	p<0.001

Table 4. Significant effects on edit distance and preferences for the reinforcement analysis.

cue	edit distance mean		time mean (sec)		preference mean	
	with	*without*	*with*	*without*	*with*	*without*
cmn	2.19	13.85	190.1	257.16	3.2	2.3
cnc	6.81	8.93	229.18	215.46	2.82	2.7
col	7.07	8.65	224.74	220.19	2.87	2.65
prx	6.15	9.63	206.49	239.66	2.91	2.61
aln	8.76	6.84	223.74	221.26	2.75	2.78

Table 5. Distance, trial time, and preference means for the per grouping cue analysis. (cmn=common region, cnc=connectedness, col=color similarity, prx=proximity, aln=alignment)

We detail means in Table 5, and main effects and interactions on the distance, time, and preferences in Tables 6, 7, and 8 respectively.

Effects on Distance All variables had significant main effects on structural distance (Table 6). With the exception of *alignment*, each grouping cue —*common region*, *connectedness*, *color similarity*, and *proximity*— reduced distance and improved structural understanding. However, *alignment* increased distance. As in the previous per cue combination and reinforcement analyses, low *density* and text-dominant visuals reduced distance.

All variables except *visual type* interacted with *common region*. *Connectedness*, *color similarity*, *alignment* harmed structural communication without *common region*, but had no effect with *common region*. High *visual density* harmed communication without *common region*, but had less harm with *common region*. There were also two other interactions with *alignment*. *Connectedness* improved structural communication without *alignment*, but with *alignment*, improvement was more modest. *Visual type* had no effect without *alignment*, but with *alignment*, it was harder to communicate structure with text-dominant visuals.

In three way interactions, the powerful effects of *common region* masked several two-way interactions. The two-way effects we discuss below were only present when the *common region* cue was not used. This includes the *connectedness* and *alignment* interaction we discussed in the previous paragraph. Without *proximity*, *alignment* had no effect. With *proximity*, *alignment* harmed structural communication. *Proximity* improved communication especially well when *visual density* was low. *Proximity* improved communication more with text-dominant visuals. In image-dominant visuals, *density* had no effect, but with text-dominant visuals, high *density* harmed communication.

Effects on Time *Common region* and *proximity* had significant main effects, with each reducing time (Table 7).

The two-way interaction of *common region* and *proximity* mirrored the same interaction's effect on distance. With *common region*, *proximity* had no effect on time. Without *common region*, *proximity* reduced time. *Common region* and *visual density* also interacted. With *common region*, using lower density visuals reduce time. Without *common region*, *visual density* had no effect. Finally, without *connectedness*, *proximity* had no effect. With it, *proximity* decreased time.

Once again, the powerful effects of *common region* masked some two-way interactions. The two-way *connectedness-proximity* interaction we discussed in the preceding paragraph was only present without *common region*. The masked *proximity-visual type* interaction's

number of cues	independent variable	ANOVA of edit distance	
		F	*p*
single cue	cmn	F(1,2208)=1401.63	p<0.0001
	cnc	F(1,2208)=71.01	p<0.001
	col	F(1,2208)=46.10	p<0.001
	prx	F(1,2208)=158.10	p<0.001
	aln	F(1,2208)=14.91	p=0.0001
	visual density	F(1,2208)=32.16	p<0.001
	visual type	F(1,2208)=7.98	p=0.005
two cues	cmn×cnc	F(1,2168)=98.22	p<0.0001
	cmn×col	F(1,2168)=58.19	p<0.001
	cmn×prx	F(1,2168)=162.25	p<0.001
	cmn×aln	F(1,2168)=10.27	p=0.001
	cnc×aln	F(1,2168)=8.38	p=0.004
	cmn×density	F(1,2208)=6.56	p=0.01
	aln×type	F(1,2208)=5.73	p=0.02
three cues	cmn×cnc×aln	F(1,2128)=15.59	p<0.001
	cmn×prx×aln	F(1,2128)=4.02	p=0.045
	cmn×prx×density	F(1,2168)=13.6	p=0.0002
	cmn×prx×type	F(1,2168)=8.58	p=0.003
	cmn×density×type	F(1,2208)=6.84	p=0.009

Table 6. Significant main effects and interaction of grouping cue, visual density, and visual type on the edit distance for the per grouping cue analysis. (cmn=common region, cnc=connectedness, col=color similarity, prx=proximity, aln=alignment)

number of cues	independent variable	ANOVA of time	
		F	*p*
single cue	cmn	F(1,2208)=110.69	p<0.001
	prx	F(1,2208)=29.53	p<0.001
two cues	cmn×prx	F(1,2168)=26.30	p<0.001
	cnc×prx	F(1,2168)=6.16	p=0.01
	cmn×density	F(1,2208)=4.72	p=0.03
three cues	cmn×cnc×prx	F(1,2128)=4.11	p=0.04
	cmn×prx×type	F(1,2168)=4.36	p=0.03

Table 7. Significant main effects and interaction of grouping cue, visual density, and visual type on the time for the per grouping cue analysis. (cmn=common region, cnc=connectedness, col=color similarity, prx=proximity, aln=alignment)

number of cues	independent variable	ANOVA of preference	
		F	*p*
single cue	cmn	F(1,2208)=448.23	p<0.001
	cnc	F(1,2208)=16.27	p<0.001
	col	F(1,2208)=38.20	p<0.001
	prx	F(1,2208)=63.52	p<0.001
	visual density	F(1,2208)=30.33	p<0.001
	visual type	F(1,2208)=63.32	p<0.001
two cues	cmn×cnc	F(1,2168)=19.64	p<0.001
	cmn×prx	F(1,2168)=44.08	p<0.001
	col×aln	F(1,2168)=6.90	p=0.01
	prx×aln	F(1,2168)=30.86	p<0.001
	cmn×type	F(1,2208)=6.91	p=0.01
three cues	cmn×cnc×prx	F(1,2128)=14.89	p<0.001
	cmn×prx×density	F(1,2168)=5.88	p=0.01
	cmn×density×type	F(1,2208)=4.47	p=0.03

Table 8. Significant main effects and interaction of grouping cue, visual density, and visual type on the preference for the per grouping cue analysis. (cmn=common region, cnc=connectedness, col=color similarity, prx=proximity, aln=alignment)

effects on time mirrored those on distance, with *proximity* lowering times more for text-dominant visuals.

Effects on Preference All grouping cues except *alignment* had significant effect on preference (Table 8), with the use of a grouping cue being preferred. Low *visual density* and image-dominant visuals were preferred.

There were five significant two-way interactions. With *common region*, viewers preferred the use of *connectedness*, but without *common region*, they had no preference. The relationship of *common region* to *proximity* mirrored the distance and time interactions. With *common region*, viewers had no preference, but without *common region*, they preferred the use of *proximity*. With *common region*, viewers preferred image-dominant visuals, but without it, their preference was stronger. With *alignment*, viewers had a slight preference for the use of color, but without it, their preference for color was stronger. Similarly, with *alignment*, viewers had a slight preference for the use of *proximity*, but without it, their preference for *proximity* was stronger.

Common region masked or partially masked some significant two-way interactions. With *common region*, *proximity* and *connectedness* did not interact. Without *common region*, when *proximity* was used, *connectedness* increased preferences strongly, while when *proximity* was not used, adding *connectedness* had no effect. With *common region*, *proximity* and *density* had significant but fairly weak interacting effects. When *proximity* was used, *visual density* did not affect viewer preferences, while when *proximity* was not used, viewers preferred lower density visuals. Without *common region*, *proximity* and *density* had stronger interacting effects. When *proximity* was used, viewers preferred lower density visuals, while when *proximity* was not used, viewers had no preference. With *common region*, *visual density* and *visual type* had significant but relatively weak interacting effects. When *visual density* was low, preferences were uniformly high. When *visual density* was high, viewers preferred image-dominant visuals. Without *common region*, no matter what the visual density, viewers preferred image-dominant visuals.

4.4 Discussion

Hypothesis 1: Using reinforcing cues improves communication clarity. We found strong evidence in support of this hypothesis, with the accuracy of the knowledge structures participants perceived improving rapidly as reinforcement increased. In addition, viewers preferred reinforced structure, and were able to understand and input them more rapidly. The levels of structural accuracy viewers reached in this reinforcement study were much higher than in our first disjoint study, which leads us to speculate that any possibility of competition between grouping cues should be avoided.

Hypothesis 2: As visual density declines, communication clarity will improve. We also found strong support for this second hypothesis. In all three of our analyses, viewers preferred low density visuals, and were able to perceive knowledge structures with greater accuracy in low density settings. When *common region* was not used, lower density visuals improved structural clarity, and increased *proximity*'s affectedness and appeal. These interactions with *proximity* provide some evidence that lower density visuals improve clarity by giving some grouping cues the visual space they need to be effective.

Hypothesis 3: Proximity and alignment will affect structural communication even in their absence. We found some support for this hypothesis as it applies to *proximity*. *Proximity* had the second strongest effect overall on both perceived structural accuracy and preferences. *Proximity*'s interaction with *connectedness* demonstrated its effect in absence: without *proximity*, structures communicated by *connectedness* took much longer to understand. However, *proximity* did not interact with grouping by color, and its interactions with *common region* and *connectedness* limited its effects: *proximity* had little effect when *common region* cues were present, and when *connectedness* cues were not present.

We found no support for *hypothesis 3* as it might apply to *alignment*. In fact, our application of *alignment* was problematic: rather than improving structural understanding, it harmed it. In order to follow the overall design of our reinforcement study, we were forced to develop a method for communicating two-level structures without the use of any reinforcing cues. The obvious solution to communication of multi-level structure with *alignment* is outlining, however this technique relies heavily on *proximity*. Our alignment-only solution, relying on horizontal location to communicate outer groups and vertical location to communicate inner groups, was not effective, and did not permit us to examine this portion of *hypothesis 3*. In fact we are not at all certain that an effective alignment-only method for communicating hierarchical structure exists. It may be that for communicating complex structures, *alignment* is best used in support of other grouping cues.

Hypothesis 4: many cues rely on reinforcing proximity. We found limited support for this hypothesis, which was based on the observation that every cue we used except color was spatial in nature. Without *proximity*, they should therefore be less effective. An interaction with *connectedness* supported the hypothesis: without *proximity*, viewers understood knowledge structures much more slowly. However, as we have just discussed, our use of *alignment* was problematic, and *common region* – by far the most powerful grouping cue – was equally effective, with or without *proximity*.

Hypothesis 5: text dominant visuals rely more heavily on spatial arrangement than image-dominant visuals. We found strong support for this hypothesis. When *common region* was not in use, *proximity* cues improved the understanding of text-dominated visuals, and enabled viewers to reach that understanding more quickly. Only text-dominated visuals were sensitive to the problems introduced by our *alignment* solution.

5 General Discussion

Disjoint grouping and dominance   Perception research shows, and our own research largely confirms, that when cues are disjoint (communicating different structures), they often interfere with one another in complex ways. Were the dominance of one cue over another easy to predict, designers might make good use of disjoint grouping combinations for aesthetic effect, but such grouping is rare in professional design practice. Our own results suggest that disjoint grouping combinations should be avoided, with the possible exception of disjoint combinations with *common region*.

Visual understanding and preferences   As Figure 6 shows, visual grouping's effects on viewer preferences closely mirrored its effects on viewer understanding, with clearer communication almost always being preferred. This was also true of *visual density*. *Visual type* was the primary exception: while viewers understood text-dominant visuals more clearly, they preferred imagery. This apparent contradiction will be familiar to experienced designers. Does this mean that good understandability implies good visual appeal, at least with respect to visual grouping? Our results are good evidence of correlation; further research is needed to demonstrate any causality. In HCI, research on the relationship between usability and aesthetics has found evidence of an asymmetry in causality, with good usability improving aesthetics, but not the reverse [13]. It seems quite possible that understandability will have a stronger relationship to visual appeal than usability.

6 Lessons for designers

In this section we distill our results into principles that can guide novice designers as they create visuals. We have one general lesson:

Reinforcement   Using multiple cues, each communicating the same structure, is a reliable way for novice designers to add structural clarity to their visuals. In our research, this sort of reinforcement improved communication accuracy and speed, and was preferred by viewers.

6.1 Lessons by grouping cues

Common region   *Common region* is a very powerful grouping cue. In our experiments, using it resulted in the most accurate, fastest, and most preferred structural communication. Moreover, *common region* can compensate for the lack of other reinforcing grouping cues and eliminate the harmful effects on structural communication of high *visual density*. Finally, for responsive design on displays of varying size, *common region* is particularly useful because its property of containment is independent of position and size.

Connectedness   *Connectedness* is a more challenging cue to use. In our studies, viewers took the most time interpreting the structures communicated by it and only had a stronger dislike for *alignment*. *Connectedness* should be reinforced with *proximity*. In our results, using *connectedness* slowed structural communication and was not preferred unless it was reinforced with *proximity*. This recommendation might be explained by the visual language required for successful communication through graphs. An unorganized graph is not a clear form of structural communication: successful communication requires systematic layout, including aligning nodes, minimizing link distance, and use of internal nodes.

Color similarity   Color is an effective communicator of simple structures. For more complex structures, it should be reinforced strongly. Because of its non-spatial nature, it rarely interacts with other grouping cues. This might explain its effectiveness in our first experiment as a communicator of disjoint outer groups and should make it useful for responsive designs.

Proximity   In our experiments, *proximity* was the second most effective grouping cue, as measured by distance, time, and preference. Its effectiveness can be improved by reinforcing *connectedness*. *Proximity* is particularly effective in text-dominant and low density visuals.

Alignment   Whenever possible, *alignment* should be used with strong reinforcement. In our experiments, *alignment* was the least effective grouping cue as measured by distance and preferences, and among all grouping cues, only *alignment* reduced the accuracy of structural communication. It may be that *alignment* is better used for non-structural communication such as sequence. Good *alignment* is especially important with text-dominant visuals.

6.2 Lessons by content

Visual density   Structural communication is clearer and preferred with low-density visuals. If designers must use high-density visuals, *common region* can make communication effective but reliance on *proximity* should be reduced. Lowering density is particularly important when content is text-dominant.

Visual type   This will not surprise designers, but viewers prefer image-dominant visuals. However, it may be surprising to learn that structural communication is more difficult in visuals dominated by imagery. *Proximity* is an especially powerful grouping cue with text-dominant visuals.

7 Conclusion and Future Work

We begin our conclusion with a discussion of some limitations of this study. First, we did not vary the complexity of the structures or information participants viewed. We anticipate that the effectiveness of many grouping cues would improve as complexity drops; since we believe our levels of complexity were already somewhat challenging. Next, as designers know and Mayer's work implies, effective visual communication is about much more than communication of hierarchical structure. In particular, non-hierarchical structures are common, sequence may be just as important as hierarchical relationships, and we did not address emphases or importance. Third, ours is not truly an end-to-end study of visual communication and understanding. For example, we did not examine knowledge transfer as defined by Mayer and other learning science researchers, using test questions testing what viewers learned. Finally, although we asked viewers whether they liked the stimuli they viewed, we did not rigorously study the aesthetics of our stimuli. We believe strongly that the aesthetic quality of a visual is a crucial contributor to its effectiveness.

Despite these limitations, we believe our work makes an important contribution to the science needed for more general communicative applications of visualization. Our research is a unique synthesis of concepts from learning science (active learning and structural organization), design (visual hierarchy), and psychology (Gestalt cues and measurement of cognitive structures) to answer pressing questions about the use of visualization technology in broader communicative domains. Our research frames these questions in a testable form, finds useful answers to them, and discusses their meaning in a general applied context.

This work has made clear the importance of future research that could address the research limitations raised above, including communication of sequence and emphasis, as well as the relation of effective communication to aesthetics.

Acknowledgments

The authors wish to thank Agnes Gall, Anne McLaughlin, and Douglas J. Gillan for their valuable feedback, and the Turkers who participated in our studies.

References

[1] M. Ben-AV and D. Sagi. Perceptual grouping by similarity and proximity: Experimental results can be predicted by intensity autocorrelations. *Vision research*, 35:853–866, 1995.

[2] S. Han, G. W. Humphreys, and L. Chen. Uniform connectedness and classical Gestalt principles of perceptual grouping. *Perception & Psychophysics*, 61:661–674, 1999.

[3] J. Heer and M. Bostock. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. *ACM Human Factors in Computing Systems (CHI)*, pages 203–212, 2010.

[4] K. Koffa. *Principle of Gestalt Psychology*. London, Routledge & Kegan Paul Ltd., 1935.

[5] M. Kubovy and M. Van den Berg. The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological Review*, 115(1):131–154, 2008.

[6] D. Luna and P. R. Montoro. Interactions between intrinsic principles of similarity and proximity and extrinsic principle of common region in visual perception. *Perception-London*, 40(12):1467–1477, 2011.

[7] R. Mayer. *Multimedia Learning*. Cambridge University Press, 2009.

[8] S. Palmer. Common region: A new principle of perceptual grouping. *Cognitive Psychology*, 24:436–447, 1992.

[9] S. Palmer and D. Beck. The repetition discrimination task: An objective method for studying perceptual grouping. *Perception and Psychophysics*, 69(1):68–78, 2007.

[10] S. Palmer and I. Rock. Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic bulletin & review*, 1(1):29–55, 1994.

[11] R. Rosenholtz, A. Dorai, and R. Freeman. Do Predictions of Visual Perception Aid Design? *ACM Transactions on Applied Perception*, 8(2), 2011.

[12] R. Rosenholtz, N. R. Twarog, N. Schinkel-Bielefeld, and M. Wattenberg. An intuitive model of perceptual grouping for HCI design. *In Proceedings of SIGCHI 2009*, pages 1331–1340, 2009.

[13] A. N. Tuch, S. P. Roth, K. Hornbæk, K. Opwis, and J. A. Bargas-Avila. Is beautiful really usable? toward understanding the relation between usability, aesthetics, and affect in hci. *Computers in Human Behavior*, 28(5):1596–1607, 2012.

[14] J. Wagemans, J. H. Elder, M. Kubovy, S. E. Palmer, M. A. Peterson, M. Singh, and R. von der Heydt. A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138:1172–1217, 2012.

[15] C. Ware. *Visual thinking for design*. Morgan Kaufmann, 2008.

[16] M. Wattenberg and D. Fisher. A model of multi-scale perceptual organization in information graphics. *Infovis 2003*, 2003.

[17] K. Zhang and D. Shasha. Simple fast algorithms for the editing distance between trees and related problems. *SIAM Journal of Computing*, 18:1245–1262, 1989.

[18] C. Ziemkiewicz and R. Kosara. Laws of attraction: From perceptual forces to conceptual similarity. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1009–1016, 2010.