Laws of Attraction: From Perceived Forces to Conceptual Similarity

Caroline Ziemkiewicz, *Member, IEEE*, Robert Kosara, *Member, IEEE*

Abstract—Many of the pressing questions in information visualization deal with how exactly a user reads a collection of visual marks as information about relationships between entities. Previous research has suggested that people see parts of a visualization as objects, and may metaphorically interpret apparent physical relationships between these objects as suggestive of data relationships. We explored this hypothesis in detail in a series of user experiments. Inspired by the concept of implied dynamics in psychology, we first studied whether perceived gravity acting on a mark in a scatterplot can lead to errors in a participant's recall of the mark's position. The results of this study suggested that such position errors exist, but may be more strongly influenced by attraction between marks. We hypothesized that such apparent attraction may be influenced by elements used to suggest relationship between objects, such as connecting lines, grouping elements, and visual similarity. We further studied what visual elements are most likely to cause this attraction effect, and whether the elements that best predicted attraction errors were also those which suggested conceptual relationships most strongly. Our findings show a correlation between attraction errors and intuitions about relatedness, pointing towards a possible mechanism by which the perception of visual marks becomes an interpretation of data relationships.

Index Terms—Perceptual cognition, visualization models, laboratory studies, cognition theory.

1 Introduction

The central idea of information visualization (infovis) is that people can derive conceptual relationships and patterns from the layout of marks in a visual representation. However, the process by which this derivation happens remains somewhat mysterious. The reading of individual data values is relatively straightforward, requiring the user only to decode a known mapping of data to visual properties. Understanding how those visual attributes combine and interact to suggest relationships among data points is more complex.

In visualization, many standard visual elements are used to suggest relationships of various kinds: lines connect, outlines group, and so forth. Infovis researchers know to some extent which of these are most useful and what kinds of relationships they seem to naturally suggest. And yet we know relatively little about why these particular associations between image and conceptual relationship are so strong, and indeed, why some are stronger cues than others.

In essence, the question is why certain visual structures reliably suggest certain information structures to a viewer. Why do outlines group items? Why do line graphs show a trend while bar graphs show separate groups even when this interpretation doesn't fit the data, as shown in a study by Zacks and Tversky [16]? Is this simply a matter of convention, or are there more basic mechanisms at work? Preattentive processing can explain why color or shape similarity causes marks to be seen as belonging together, but what about more complex cues such as the connecting lines of a network graph?

Answering these questions is of theoretical interest to infovis researchers, but also has very practical consequences. Understanding why a particular set of marks suggests the relationship it does means being able to predict what a novel visual representation will indicate, and how to refine that representation towards a specific goal. It means being able to predict how a user might read a pattern, and being able to use that information for evaluation or to adapt the representation to highlight or analyze such patterns. It means being able to compare two visualizations not just in terms of response times or error rates, but in terms of what each says about its data.

We propose that a possible mechanism by which some visual relationship cues acquire meaning is by prompting mental simulation of forces by a viewer. That is, a visual cue such as a connecting line between two marks may suggest to the viewer two items being pulled together, leading her to see the items as related. Researchers of implied dynamics in psychology have demonstrated that such mental simulations of forces can lead to position errors in the recall of static scenes, a finding which we use to test this theory in a series of user experiments. We argue that our findings show that visual cues which suggest relationships between data items also increase this perceived attraction between the related marks.

This paper contributes the first steps towards a model of how visual representations are interpreted as conceptual relationships based on implied dynamics, and presents a series of experiments that support this claim in the case of perceived attraction. These findings can be used to form the basis for future research on how visual dynamics are read as conceptual structure.

2 Related Work

Previous research by the authors [17] found significant effects of visual structure on semantic responses to simple charts that were said to depict financial data about a company. When prompted, participants tended to describe these semantic interpretations in terms of apparent forces acting within and on the charts. For example, borders were said to block movement between pieces, which suggested communication difficulties between departments to some participants. Rectangular charts were less likely to "roll away" than pie or donut charts, and so suggested a more stable company.

One way to interpret this is through the perception theory of Gestalt psychology [12], which argues that viewers naturally simplify complex visual information by grouping, connecting, and finding symmetry among parts of a scene in a predictable fashion. The Gestalt psychologists introduced several laws that predict patterns of organization in the visual field, such as similarity (similar objects are seen as belonging together) and common fate (objects that appear to be moving in the same direction belong together). This classic body of theory has obvious applications to visualization, and indeed helps to explain many of the patterns that appear in information visualization.

However, while Gestalt principles describe some of the patterns people see in a visualization, they don't necessarily explain *why* those particular perceptual configurations suggest a grouping, or help us predict what other kinds of patterns might arise. To go beyond Gestalt, we need a more generalizable way to connect low-level perception to high-level patterns. Some vision researchers have introduced possible ways of understanding object and pattern recognition by analyzing low-level visual errors. For example, Burbeck and Pizer [3] argue for their model of mental object representation by showing that it predicts certain errors in the perception of distances between lines. This rich

- *Caroline Ziemkiewicz is with UNC Charlotte, E-mail: caziemki@uncc.edu.*
- *Robert Kosara is with UNC Charlotte, E-mail: rkosara@uncc.edu.*

Manuscript received 31 March 2010; accepted 1 August 2010; posted online 24 October 2010; mailed on 16 October 2010.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

(a) Horizontal distribution (b) Diagonal distribution (c) Horizontal distribution, isolated target (d) Diagonal distribution, isolated target

Fig. 1. Visualizations used to test the effect of gravity on a participant's memory for a visual mark. One of the circles in each of these graphs would flash, and the participant would try to remember its position after the graph vanished. The black arrows in these images have been added to indicate the targets for illustration purposes and were not present during the study.

body of work on visual illusions and what they say about mental representations is similar in spirit to our work and in some cases may suggest alternate explanations for the errors we find.

More directly related to our goals is the body of work described by Tversky [15], who discusses a number of ways in which biases towards symmetry, alignment, and simplicity can lead to systematic errors in memory for graphs and maps. This work also uncovered effects of top-down expectations on perceptual memory, as when participants were biased to remember a curve as more or less symmetrical based on a verbal description of the pattern it showed. It further hows that Gestalt-like simplicity biases can lead to perceptual errors and also affect what patterns we interpret in a visualization, and raises the possibility that memory distortions of other kinds can shed light on how we interpret visual representations.

However, biases towards symmetry and simplicity do not seem to explain all types of visual relationships. We argue for a more general mechanism based on implied dynamics [6], a theory in perceptual psychology that states that people simulate implied motion and physical forces when viewing static scenes. This theory is supported by experiments showing that people viewing a still photograph taken in the middle of some motion tend to remember the object or person as being slightly further along in their path of motion than they actually were [5]. This is interpreted as meaning that a viewer continues to simulate the apparent motion past the point at which it was captured, rather than remembering it as a static image.

Freyd et al. [7] found that this effect applies not only to implied motion, but also to apparent physical forces such as gravity. Viewers were shown a cartoon drawing of a potted plant on a table, then a second image in which the table either was or was not removed. They were then tested on their recall of the plant's vertical position. When the table was removed, viewers remembered the plant as being slightly further down than it actually was, but not when the table was still there. This suggests that people also simulate gravity acting in a static image. Further research by Hubbard [10] found that varying the apparent weight of these objects by changing their size could increase the degree of such gravity effects, suggesting a surprising degree of physical simulation at work in basic perception.

If these physical simulations are applied to static visualizations as well, this might explain why viewers so readily use apparent physical forces between marks as a basis for semantic interpretations of their relationships. These examples study pictorial representations such as people and plants, and it may be argued that physical forces which make sense in this context do not apply to the abstract points, blobs, and lines of typical visualizations. However, Freyd and Pantzer's finding [8] that participants remember arrows as being farther along in the direction in which they point than they actually were shows that implied dynamics can apply to more abstract representations as well. The authors also found this effect with variously shaped triangles, with the narrowness of the triangles point being a good predictor for the degree of displacement in the pointing direction. This suggests that there is more at work in this effect than merely the convention of an arrow as a pointing device. This work raises the possibility that the meaning of arrows as pointing towards something derives from the viewer's internal simulation of the arrow moving in that direction. The purpose of our current work is to determine whether such simulations influence meaning in other types of visual representations.

3 The Role of Dynamics in Visualization

Based on these earlier findings, we argue that people construct meaning in a visualization in part by simulating the apparent physical dynamics acting on marks and metaphorically interpreting the results. That is, perceived forces and conceptual relationships are closely linked in visualization use. Reading a visualization involves understanding how objects relate to each other, which objects belong together, and how the overall structure can be acted upon. We argue that mental simulations of dynamics may be used to extract this kind of structural meaning from abstract visual patterns. This view of visualization is related to theories of how Gestalt groupings are used to determine relationships in a visualization, but attempts to identify a more general mechanism by which such visual patterns acquire meaning.

To test this hypothesis, we first studied whether people see implied dynamics of the kind found by Freyd and others when viewing a visualization at all (Section 4). The results of this study suggested that implied gravity may distort a user's memory of a visualization, but also raised the possibility that simulated attraction between marks is a more salient effect. We therefore went on to perform two related studies that examined this attraction effect in depth. The first one tested whether any visual cues used to suggest relationships, such as connecting lines and outlines, cause memory distortions in the direction of the implied relationship (Section 5). In the context of implied dynamics, these memory distortions would imply that the participant mentally simulates the marks as being physically drawn towards one another. The second tested whether the strength of this simulated attraction also corresponds to the strength of the apparent conceptual relationship suggested by these visual cues (Section 6). Taken together, these studies support the hypothesis that a visual cue used to suggest a relationship between data items causes the viewer to see the related marks as actually attracted to one another. In the following sections, we describe these studies in detail and discuss their implications.

4 Visual Gravity

Our first experiment focused on the mental simulation of gravity in a visualization. We initially aimed to test whether viewers of a scatterplot-style visualization mentally simulate gravity acting on marks, and whether this simulation is affected by the layout of marks. We hypothesized that like the biases towards simplicity and symmetry suggested by Gestalt perception theory, implied dynamics could also account for systematic memory distortion errors in a visualization.

4.1 Experiment

The purpose of our first experiment was to establish whether simulated gravity affects perception of marks in a visualization and whether this effect is influenced by visual and structural properties of the visual marks. For example, the apparent weight of a mark, as determined by size and color, may affect how much visual momentum it has. By testing whether apparent gravity causes perceptual errors in a visualization, we intended to analyze whether this and other perceived forces can predict the patterns read by users.

4.1.1 Materials

The materials used in this study were a series of "bubble" scatterplot graphs similar to those used in Gapminder [14]. Examples are shown in Figure 1. We used these because they were relatively simple visualizations which allowed for natural variation in the size, color, and layout of visual marks. Each graph contained 15 circles laid out in a semi-random formation and randomly colored according to a categorical scale from ColorBrewer [2]. One of these circles was the "target" circle, and participants were told to remember its position and recall it on a blank graph afterward. The overall design was inspired by Freyd's studies of representational momentum and implied dynamics, although we had participants directly report their memory of the mark's position with a mouse click. As in the implied dynamics work, we measured recall of the point rather than direct perception because our aim was not just to demonstrate perceptual biases, but to attempt to uncover the underlying mental simulations that drive these biases. Memory errors may show that a participant is simulating the movement of visual marks past their originally perceived position.

We varied several visual factors of the target circle as well as the overall layout of the graph to test whether any perceived gravity was altered by the apparent weight of the target or its relation to the rest of the distractor points, as suggested by Hubbard [10]. We altered the color and size of the target, since we hypothesized these factors to have the most direct influence on the target's apparent weight. We also varied its position systematically with a focus on testing whether marks that are higher up in the graph area are more likely to show a gravity effect. In addition to this, we varied the target's position relative to the distractor points; the distractors were clustered together, and the target was either within this cluster or isolated from it. Finally, this cluster of distractors was either roughly horizontal or laid out along a diagonal in a roughly linear relationship. This linear relationship was always positive (that is, the diagonal stretched from the lower left corner to the upper right corner). These last two factors were meant to test whether the other marks in a graph, especially those with a strong trendline, have an effect on a target's apparent dynamics that can override or confound the simple gravity effect.

4.1.2 Participants

We recruited 45 participants via Amazon's Mechanical Turk (see Section 7.1 for considerations regarding online studies). Participants included 21 females (46.7%) and 24 males. Self-reported age ranged from 18 to 64, with a mean of 30.9. Participants received an initial payment of \$0.20 through Amazon's payment service for completing the study, and received a further bonus of \$0.02 for responses that fell within an accuracy threshold of 50 pixels, for a maximum total of \$0.84. This accuracy bonus was meant to create an incentive for participants to try and remember the target's position as closely as possible, rather than clicking randomly to get through the study quickly.

4.1.3 Procedure

Over the course of the experiment, participants saw 32 scatterplot graphs. In each trial, the target circle initially flashed twice, then the graph remained visible for 2 seconds. The graph was then replaced by a black screen at the bottom of which was an "OK" button that participants had to click to continue. This button was included to prevent participants from cheating by leaving their mouse cursor centered over the target circle after the initial graph vanished. The black screen was then replaced by a blank graph in which only the X and Y axes were visible. Participants were told to click their mouse on the location on the blank graph where they remembered seeing the flashing target. An X briefly appeared on the screen where they clicked for feedback. They then continued immediately to the subsequent trial.

This trial design was entirely within-subjects. Each participant saw all of the conditions, and there were no between-subjects variables.

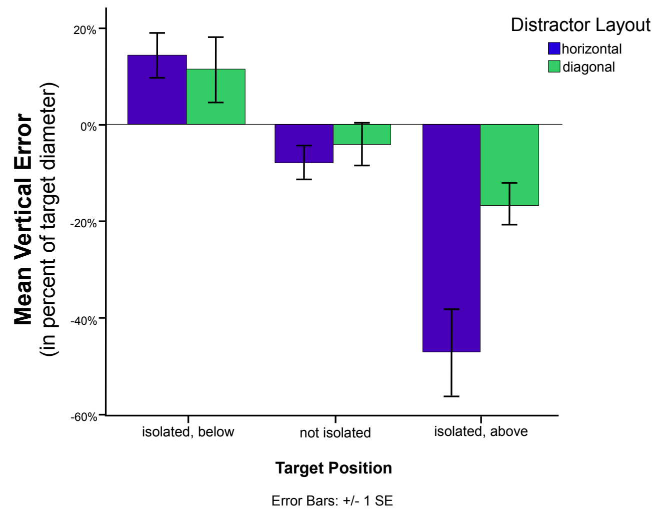


Fig. 2. The amount of vertical error in each of the target's three position conditions and the two distractor distribution conditions, in percentage of target size. A negative value indicates a downward shift. The average shift of 7% is slightly higher than those most often found in Freyd's implied dynamics work [7], and in cases where the target was above the distractor points this shift was much higher.

To reduce the impact of a learning effect, we randomized the order in which participants saw each trial condition. Participants were shown the position where they clicked, but they were not told whether they were within the correctness threshold as they went along. They were not told how many answers they got correct until after they completed the study and submitted their results, when their bonuses were granted.

Participants were not explicitly permitted breaks. However, since this study was performed online, we could not ensure that participants remained at their computer during the entire study, especially during the time when the black screen was present. We did record the start and end times of each trial along with response time, so we can determine the length of time each participant viewed the black screen. In only one case was this time longer than one minute: during one trial, a participant waited at the black screen for 408 seconds (6.8 minutes), which may indicate that this participant got up and left or performed some other task while pausing the study. We removed this outlier from our analysis, although neither this nor removing this participant altogether affects the significance of any of the subsequently reported tests. Including the outlier trial, the mean time spent on the black screen was 1.9 seconds (S.D. = 11.3). With the outlier removed, the mean black screen time was 1.6 seconds (S.D. = 2.5). Apart from the outlier, the maximum time spent on the black screen was 42.9 seconds.

We recorded the location of the participant's guess about the target's location as well as their response time. Our primary analysis measures were the accuracy of their response, measured in Euclidean distance from the actual center position of the target, with particular attention to the amount by which the target was remembered as being lower on the screen than it actually was.

4.2 Results

Because we did not control for a participant's screen resolution, the following results are presented in terms of percentage of target size or display size where appropriate. In analyzing our results, we first removed cases with extreme distance error, defined as greater than three standard deviations from the mean error, calculated as a percentage of the display width ($M = 4.73\%$, $S.D. = 9.75\%$), or greater than 33.97% of the display width. This was intended to remove any cases in which a participant was clicking randomly in order to finish the task more quickly, or where she had disregarded or failed to follow the task instructions. This resulted in the dropping of 39 trials, or 2.7% of the

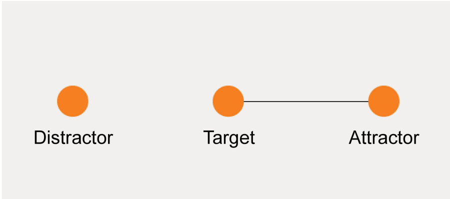
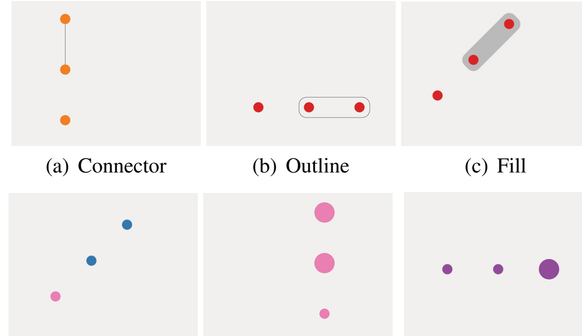


Fig. 3. The general layout of the stimuli in the attraction factors study. The attractor mark was linked with the central target mark in some fashion that we hypothesized to cause apparent attraction between the two—in this case, with a connecting line.



(a) Connector (b) Outline (c) Fill

(d) Color similarity (e) Size similarity (f) Gravity

Fig. 4. The six elements we initially hypothesized to cause perceptual attraction between marks. As shown in these examples, the layout of the row of marks was varied with respect to orientation and screen position.

total. Dropping these outliers did not affect the significance of any of the following tests.

Overall, we found a small but significant overall effect of gravity. We analyzed the measure of vertical error, or the actual position of the target subtracted from its remembered position. A negative vertical error means that the participant remembered the target as being below its actual position, while a positive vertical error means it was remembered as being above its actual position. The average amount by which the response point fell below the target point was 7% of the target diameter, which is a minor difference but nonetheless significantly greater than zero ($t(1399) = -3.1$, $p < .01$). There was no significant horizontal effect. Target diameter was chosen instead of screen size in this case to allow comparison with Freyd et al. [7], who report target height but not overall display size.

Contrary to our hypothesis, we found no effect of target color or size on this downward shift. We did, however, find a significant correlation between target height and percentage of downward shift ($r(1400) = 0.23$, $p < .001$); that is, circles that were higher up on the graph shifted further down than those closer to the bottom of the screen.

Although this effect is very small, it is worth noting that the findings in the studies of gravity shifts by Freyd et al. [7] also suggested that the amount by which an object shifts downwards in a participant's memory is quite small, with common downward errors in their experiments of 3.4% or 5.9% of the target object's height. While our experiment used a different testing method, the average amount of downward shift does seem to be similar or slightly larger. That said, it should be noted that the fact that our "OK" button was located at the bottom of the screen may have biased these results downwards in general, simply because participants were starting their mouse movement from below the graph.

However, the effects of the structural factors we varied suggest a more complex interpretation. As discussed in Section 4.1.3, we varied both the absolute position of the target and its position with respect to the clustered distractor marks. We analyzed the variable of target position by splitting the targets into three groups: those which were not isolated from the larger cluster, those which were isolated and above the cluster, and those which were isolated and below the cluster. This effect was studied with a 3x2 repeated measures ANOVA, in which factors were target position and distractor layout (horizontal or diagonal) and the dependent variable was the amount of vertical error. The main effect of target position and the interaction of target position by distractor layout both failed the assumption of sphericity, so a Greenhouse-Geisser correction is employed for these tests.

We found a significant main effect of target position, $F(1.37, 43) = 16.73$, $p < .001$, $\eta^2 = .28$. Pairwise comparisons using a Bonferroni test show that all three cases (isolated above, isolated below, and within the distractor cluster) differed significantly from one another. We found that the downwards vertical shift was most dramatic when the target was isolated above the distractor cluster, and was reversed on average when the target was isolated below the distractor cluster. That is, participants remembered the target as being higher than it actually was when it was positioned underneath the main cluster.

These findings are further clarified by the significant interaction between target position and distractor layout, $F(1.53, 43) = 5.38$, $p < .05$, $\eta^2 = .11$. This interaction is summarized in Figure 2. The strongest downwards shift is found when the target is isolated above a horizontal distractor cluster ($M = -48\%$, $S.D. = 61.6\%$), and the strongest upwards shift is found when the target is isolated below a horizontal cluster ($M = 15.5\%$, $S.D. = 41.8\%$).

Overall, these results suggest that, rather than a straightforward gravity effect pulling marks downwards, there is a tendency to remember the target as being closer to the central mass of distractors than it actually was. This tendency may even out when the distractors are laid out diagonally, since the central mass is evenly distributed across both the vertical and horizontal axes.

4.3 Discussion

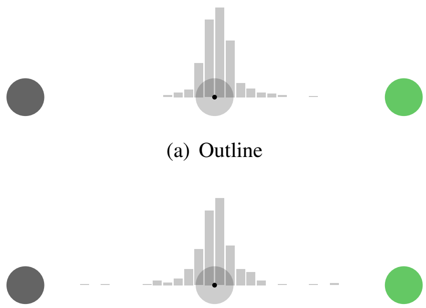
The memory distortion we found is reminiscent of Gestalt grouping principles as well as Arnheim's theories [1] about visual weight and attractions between visual shapes. Arnheim argues that such attraction is a major factor in the interpretation of composition, and can be altered in various ways by perceptions of visual weight and proportion. Another way to interpret this is that people tend to remember marks as being closer to where they would expect them to be; that is, closer to the average position. This possible explanation is supported by Tversky's other findings on visual memory distortion [15].

However, we also found a significant downward shift, arguing for the simulation of gravity alongside these other grouping principles. While the fact that we did not control for screen resolution limits our ability to evaluate our results in absolute terms, the relative magnitude of our results was comparable to findings from Freyd's experiments, lending credence to the hypothesis that the memory errors we found arose from a similar use of mental simulation.

This inital support for implied dynamics raises the possibility that the tendency to remember the target as closer to the distractor mass was caused by a sense of implied attraction between the marks, not just by Gestalt groupings. That is, mental simulation of implied dynamics could explain both the downward shift and the shift towards the distractors. This would assume, however, that participants saw the marks as attracted to each other for some reason; for example, that the target "belonged" with the other marks and so should be simulated as moving towards them. This possibility inspired the following two related studies, which examined whether visual cues used to suggest relationships can cause an attraction effect between visual marks.

5 Perceived Forces

If the dynamics model works as we have hypothesized, we should expect visual cues that indicate a relationship between items to actually cause viewers to simulate attraction between the marks they connect, as was hinted at in the previous study. The hypothesis of our second study, then, is that visual elements that imply conceptual relationships between objects represented by marks in a visualization will also cause those marks to be remembered as closer together.



(a) Outline

(b) Color

Fig. 5. Error distributions for the two of the relating elements as histograms. The green circle (right) represents the attractor, and the center circle shows the original position and size of the target circle. Histogram bars show the number of responses that fell within that distance to the attractor or distractor circle. These histograms represent all configurations of the original three circles, including vertical and diagonal orientations.

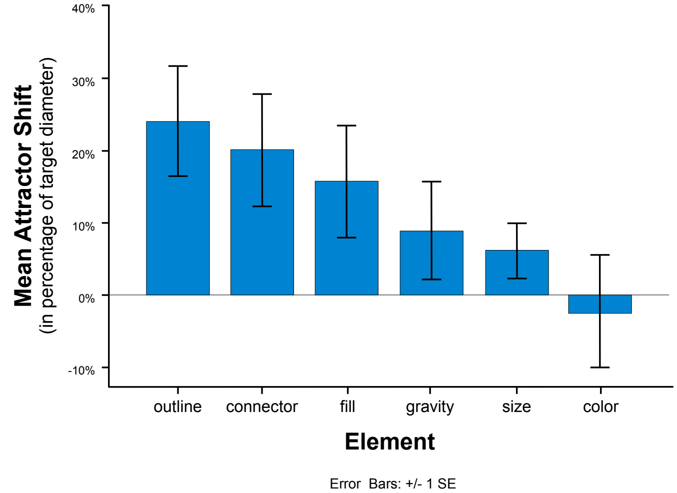


Fig. 6. The amount of attractor shift for each of the six relating elements. The deviation from zero is significant for outline, connector, and fill.

5.1 Experiment

The procedure of this experiment resembled that of the previous one, as described in Section 4.1.3. However, we simplified and focused the design to analyze the extent to which structural elements caused two marks to be remembered as closer together.

5.1.1 Materials

As before, participants saw a series of trials in which they were asked to remember the location of a target circle. However, in this case, there were only three circles in each trial, and the participant was asked to remember the location of the center circle (the target). The position of the target varied randomly between trials, but the other two circles were always positioned the same distance from the target along a straight line. This line was positioned either vertically, horizontally, or diagonally. In each trial, one of the two circles on either side of the target was the "attractor" and the other was a distractor (Figure 3).

In each case, the attractor was linked to the target with one of six elements that we hypothesized to suggest a relationship between the two marks (Figure 4). These included three external structural elements: a connecting line, an outline circling the two marks, and a fill behind the marks. There were also two cases where the target and the attractor were linked by similarity, one in which they were the same color (and the distractor was a different color) and one in which they were the same size and larger than the distractor. Finally, there was a case in which the attractor was larger than both the target and the distractor. This was meant to test whether an object with an apparently greater "mass" exerted a greater pull on the target, as suggested by Arnheim [1]. When color was not being used as a grouping elements, all three marks were the same color, which was randomly chosen.

5.1.2 Participants

We performed this experiment with 48 participants recruited online through Amazon Mechanical Turk. According to self-reported demographics data, this group included 21 females (43.8%) and 27 males (56.2%), and age ranged from 20 to 62 with an average age of 31.2. Participants were paid an initial fee of \$0.28 for their work, with an additional bonus of \$0.02 for each response that fell within an accuracy threshold of 50 pixels, for a maximum total payment of \$1.00.

5.1.3 Procedure

Each participant saw each of these six grouping elements in all six possible orientations; orientation factors included whether the line of marks was vertical, horizontal, or diagonal, as well as which side of the line the attractor was on in each case. Therefore, each participant saw 36 trials, and all participants who submitted data completed all 36 tasks. All factors were varied within subjects, and there were no between-subjects variables. Since the stimuli were much less complex than in the previous study, we showed each image for only 1 second before replacing it with the black screen. In addition, we did not include axis lines as in the previous study. Apart from these differences, the procedure was identical. As in the previous study, the order in which conditions appeared was randomized to balance out any learning effects, and participants were not told whether they answered questions correctly during the study.

As before, we did not explicitly permit breaks between tasks. The timing data in this case showed no obvious outliers in how long participants waited at the black screen before recording their response. The average wait time was 1.9 seconds (S.D. = 2.1), and the maximum wait time was 45.5 seconds.

5.2 Results

As in the previous study, we removed those responses where the overall distance error ($M = 30.59$, $S.D. = 35.89$) was greater than three standard deviations from the mean, which resulted in the removal of 2.6% of the responses. This removal did not affect the significance of any of the subsequent tests. In this experiment, our primary measure was the amount by which the target was remembered as being closer to the attractor than the distractor: that is, the distance between the response point and the center of the distractor minus the distance between the response point and the center of the distractor. We call this metric the attractor shift. Our analysis focused on whether the attractor shift was significant for any of six grouping elements.

Using a repeated measures ANOVA, we found a small but significant main effect of element type, $F(3.99, 46) = 2.59$, $p < .05$, $\eta^2 = .07$, suggesting that the elements we chose exert varying degrees of implied attraction. The strongest effect was found for outline and connecting line; color similarity had a slightly negative effect, meaning that participants remembered circles of the same color as being slightly further apart than they actually were. The error distribution for outline and color similarity are shown in Figure 5. The mean attractor shift was significantly greater than zero for the elements of outline ($t(277) = 3.2$, $p < .01$), connector ($t(282) = 2.59$, $p < .01$), and fill ($t(283) = 2.04$, $p < .05$), but not for size, gravity, or color. These results are summarized in Figure 6.

There was no significant effect of orientation, i.e., whether the layout of the marks was vertical, horizontal, or diagonal. There was also no significant interaction between orientation and visual element, meaning that the different grouping elements were not affected differently by the way in which marks were laid out.

In further contrast to our previous study, we did not find that the

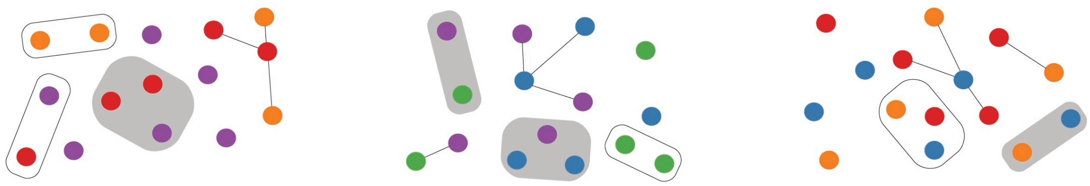


Fig. 7. Examples of the "social network graphs" shown to participants in the conceptual similarity study. These were described to participants as graphs of relationships among employees in an organization. Participants were told that the visual elements used in the graphs were significant, but these elements were not assigned any explicit meaning.

marks were remembered as being consistently lower than they actually were. In fact, we found an opposite effect in which marks were remembered as shifting upwards ($M = 3.3$, $S.D. = 16.0$), an effect which is significantly greater than zero, $t(1410) = 7.7$, $p < .01$. It is possible that the gravity effect simply vanishes in the absence of a visible X-axis, which may have served as a "ground plane" for viewers. This also supports the speculation that rather than a global gravity effect, the findings in the previous study were largely driven by attraction between the larger mass of distractors and the target.

5.3 Discussion

These results lend weight to the idea that implied dynamics between visual marks are metaphorically interpreted as statements about conceptual relationships between data elements. The strong effects of outline and connecting line are not surprising given their standard usage in infovis. The lesser effect of size and color similarity, both of which can be successfully used to group objects in visualizations, suggests nonetheless that a sense of relationship between marks can arise from factors other than the attraction we found (or perhaps that these are weaker cues to relationship). These cues may group items via Gestalt laws, rather than by simulated attraction.

In general, however, those elements we see as suggesting metaphorical "closeness"—that is, degree of relationship—seem to quite literally increase a viewer's sense of objects' physical proximity. This implies that this visual metaphor has a very literal effect on data reading at a low level. In order to test this assumption, we went on to test whether the degree of perceptual attraction is actually related to the degree of conceptual similarity between the marks linked by a given element. If color is as strong a relating element as a connecting line, for example, this would call our interpretation into question.

6 Conceptual Similarity

Our final step in this series of experiments was to test whether there was a correlation between the strength of a visual cue's implied attraction and the semantic strength of the relationship it indicates. That is, does the amount of attractor shift predict the degree of conceptual linkage between marks grouped by a given visual element? If so, this would suggest that perceived attraction really does lead to a greater sense of similarity between items.

6.1 Experiment

In this study, we took a sampling of the visual elements we found to create varying amounts of attractor shift and tested participants on the degree to which each of them implied a relationship between people in a social network graph. We hypothesized that the amount of relationship implied by a visual element would tend to correspond to the strength of the attractor shift it exerted in the previous study.

6.1.1 Materials

We presented participants with a series of images described as social network graphs of employees in an organization (Figure 7). Each graph contained fifteen circles, each of which was said to represent one employee. The marks were randomly scattered and colored, and small groups of marks throughout the graph were linked using the three direct grouping elements from the previous study: a connecting line, an outline, and a grey background fill. We did not include size differences in this study, in order to simplify the visual representation and because of concerns that the difference in diameter may confound perceptions of distance in a manner unrelated to the factors we are trying to study.

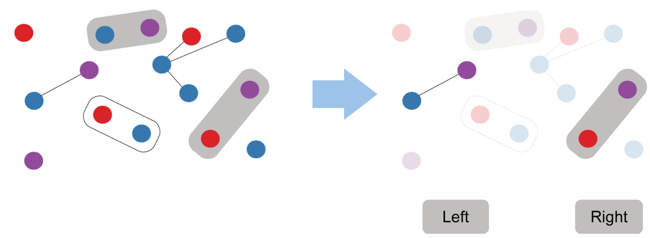


Fig. 8. A summary of the trial procedure in the conceptual similarity study. Participants first saw the image on the left for two seconds, then the image on the right with the distractor marks faded out to highlight the two target pairs. In this case, the comparison is between a connector (left) and a fill (right).

Each graph included two target pairs of marks. Each pair was either grouped using one of the four visual elements (connector, outline, fill, or color similarity) or was visually unrelated. In the case of color similarity, the grouped pair was randomly colored and not balanced across the comparison condition. We showed participants two versions of each comparison between these five pair types, for a total of 20 graphs. These graphs were designed by hand, rather than randomly generated as in the previous studies, so that each participant saw the same set of 20 images.

6.1.2 Participants

We recruited 50 participants via Mechanical Turk. Participants included 30 females (60%) and 20 males (40%), and self-reported age ranged from 18 to 67 with a mean age of 32.7. We paid participants \$0.40 for their involvement, which lasted about five minutes.

6.1.3 Procedure

Participants saw the twenty graphs in random order. Each graph was first completely visible for two seconds, and then the entire graph except for the two target pairs was faded out (Figure 8). Participants were then asked to choose which pair showed the two employees who were more related to each other. One of the target pairs was on the left side of the graph and the other was on the right, and to choose between them the participant clicked a corresponding button labeled either "Left" or "Right." The trials were designed to make it possible for participants to choose between the pairs without having to introduce further labels or highlighting visual elements that might confound our results. Participants saw each of the comparisons between elements in both orientations, to correct for any potential left-right preference. We found no significant preference for the left or right pair either overall or within subjects.

6.2 Results

In general, we found that the rankings of conceptual similarity closely resembled the rankings of attractor shift found in Section 5. Table 1

Visual Element	Attractor Shift	Choice Frequency
outline	24.1%	62.2%
connector	20.2%	64.8%
fill	15.9%	59.5%
color	-2.6%	46.8%
unrelated	N/A	16.8%

Table 1. Summary of the results from our second and third studies on conceptual similarity and perceptual attraction of marks. The *Attractor Shift* for each element is the average amount (in percentage of target diameter) by which viewers remembered marks grouped by that element as closer together than they actually were. *Choice Frequency* is the percentage of all trials including that element in which it was chosen as showing the stronger relationship between data items.

shows these attractor shifts alongside how often each grouping element was chosen as having the greater relationship, as a percentage within all comparisons that included that element. Pairs with connecting lines were most frequently chosen, while pairs with color similarity were chosen as having the greater relationship less than half the time. Within color similarity, there was no evidence that the color used made it more likely that the color similarity pair would be selected.

In order to specifically test whether attractor shift was a good predictor of perceived relationship, we took the mean attractor shifts from the previous study and recorded whether, for each trial, the choice was the visual element with the higher attractor shift. (For the unrelated case, we used the inverse of the overall mean attractor shift, or -13% of the target diameter. While this is a somewhat arbitrary value, it is only meant to capture the fact that unrelated marks should have a less powerful attraction effect than those with any grouping element, as implied by the results of our previous study.)

We found that participants chose the pair whose grouping element exhibited a higher attractor shift 67.8% of the time (S.D. = 16.1%). Individual participants' tendency to choose the higher attractor shift pair ranged from 30% to 95%. Likewise, when looking at the values for the four grouping elements, we found a significant correlation between attractor shift and choice frequency, $r(4) = .963$, $p < .05$. Since each element comparison was seen twice by each participant, we analyzed rating consistency by looking at whether participants chose the same element in both comparisons. We found that participants made the same choice in 75.6% of trials, and that no particular comparison was more likely to be rated inconsistently.

6.3 Discussion

These findings combined with those in Section 5 suggest that conceptual relationships and perceptual attraction are correlated. This can mean one of two things: either things people see as being related are remembered as being closer, or things people remember as being closer (because of some illusion or perceptual bias) are thought of as more related. It should be noted that our description of the images as social networks may have unintentionally biased participants in favor of connecting lines, which are frequently used in real-world social network graphs. However, outlines and fills are arguably less frequently used in social network graphs than color similarity, and both outperformed this cue as our model would predict.

It is notable that color similarity performed poorly against the other grouping elements; although this is a common method for showing that a collection of marks are related, and indeed has the presumed advantage of being a preattentive visual cue, it does not seem to carry the semantic weight of more direct grouping elements such as outlines and connectors. Nonetheless, it performed much better than no grouping, despite its slightly negative mean attractor shift. This suggests that color similarity does have a grouping effect that is unrelated to perceived attraction. Nonetheless, it is frequently rated as less powerful than the cues that do cause an attractor shift, suggesting that attractor shifts are associated with a stronger sense of relatedness.

While the connection between attraction and conceptual similarity is just one case in which perceived dynamics are seemingly interpreted as higher-level information, this evidence suggests that further studies of other types of dynamics may be fruitful in leading to a better understanding of how visualization cues acquire meaning.

7 General Discussion

This work shows initial evidence for the implied dynamics model of how people interpret relationships in a visualization. This model proposes that a viewer simulates the apparent forces and dynamics at work in a visualization and then metaphorically interprets those dynamics as statements about relationships and patterns within the data. While more work is needed to establish this as a general principle, we have demonstrated that in the specific example of grouping elements used to show a relationship between two marks, elements that increase simulated attraction also suggest conceptual similarity. In this section, we discuss the possible limitations of these findings as online results, their implications, and possible future research.

7.1 Limitations of Mechanical Turk

Mechanical Turk is an online job market in which people can be recruited for brief tasks and paid for their efforts. This service has become increasingly popular for use in online experiments, as a large number of relatively diverse participants can be processed very quickly [9]. Since Mechanical Turk helps correct for a number of the traditional limitations of online studies, such as the possibility of vote flooding and the lack of incentive for completion [13], it has become increasingly accepted as a user study platform among the human-computer interaction and visualization communities. It is particularly useful in studies like ours, in which there is a ground truth by which to measure results [11] and the possibility of incentivizing accurate responses through bonuses [13].

That said, some limitations remain with interpreting online studies in general. Chief among these is environmental control. In an online study, it is impossible to know whether a participant's environment is noisy or distracting or whether the participant is doing something else while performing the study. Most relevantly to a perceptual study, while some details of screen resolution and computing setup can be recorded by the study applet, we did not do so in our work, and so results are given in relative terms rather than as absolute pixel distances. There is also the possibility that participants may be "cheating" by using a ruler or other physical measurement strategy, which we cannot directly observe as in a lab experiment. While these limitations should be kept in mind while interpreting our findings, we argue that the overall consistency of our results suggests that environmental factors were not dramatic. The simplicity and short duration of our tasks may have limited the potential effects of distraction on user performance.

7.2 Predictions

Our findings suggest that apparent relationships between data items are based in part on the apparent attraction between the marks that represent them. This model of conceptual similarity can be used to make some predictions about visualization use which may guide further research and can be used to test the model.

Attractor shifts make for stronger relationship cues. Our findings in Sections 5 and 6 suggest that relating elements that cause an attractor shift are more powerful cues than those that do not. Although color similarity can be used to suggest groupings of objects, and has the advantage of being visually processed much faster than other cues, our findings suggest they are semantically less salient than visual cues that seem to physically draw marks together. Where comprehension is more important than speed, grouping marks by color may be less useful than grouping them with outlines, fills, or connecting lines.

Relating elements can be judged by attractor shift. If a visualization requires a novel visual cue to suggest a relationship, the cue can be tested for effectiveness by measuring the amount of attractor shift it causes in an experimental setup such as the one in Section 5. If the attractor shift is comparable to that found for established cues such as connecting lines and outlines, it should be naturally understandable by users. This opens the possibility of testing novel visual representations piece by piece in a controlled fashion, rather than all at once in a much more complex usability study.

Implied dynamics may explain some errors in data reading tasks. Our findings in Sections 4 and 5 suggest that small but significant errors can arise from apparent forces in a visualization. This should be considered when interpreting evaluations that depend on reading position information. It may also be possible to evaluate visualizations based on the degree to which they encourage or discourage such errors; for example, bubble-like scatterplots of the kind we used may lead to more of such errors than point scatterplots, which may not be seen as a collection of objects by the viewer.

7.3 Implications in Context

The implied dynamics theory of visualization provides a novel way to conceive of how visual representations acquire meaning. We have only demonstrated this possible mechanism in the example of similarity corresponding to attraction, but there may be other conceptual relationships that correspond to physical dynamics in this way. For example, a dependency relationship may be expressed by the implication that one mark will move if another mark is moved. Similarly, uncertainty may be expressed by apparent flexibility.

These examples can all be thought of as visual metaphors for abstract concepts. A potential theoretical impact of this work is that it shows a concrete example of a visual metaphor (*conceptual similarity is physical closeness*) in action. This argues for visual metaphors being more than an abstract trope in talking about visualization design. Rather, in this case at least, it seems that metaphorical and literal closeness are quite strongly correlated through the process of implied dynamics interpretation.

Our findings in Sections 5 and 6 also help to shed light on our initial finding of recall errors in scatterplots in Section 4 and bring an interesting perspective to the findings of Gestalt psychology. In this view, items grouped by "common fate" and other principles may be connected because these factors literally make them look closer together. The assumption in Gestalt theory is that viewers tend to be biased towards a simplified representation of complex visual scenes. It is possible that such simplification may actually happen at a basic perceptual level, and is only later interpreted by higher reasoning as indicating relationships or patterns.

7.4 Further Questions

The recall errors we found in Sections 4 and 5 raise some significant questions for our understanding of how visualizations are read. They recall other findings in perception that have had practical applications for visualization use, such as research on color theory or the errors made when people read values from area as opposed to length. In the latter case, researchers in psychophysics have actually been able to quantify the amount of this error [4], so that it can be predicted and perhaps corrected for. Further study to establish the degree of attractor shift under various conditions could potentially yield a similar basis for prediction of perceptual errors, which would be both theoretically interesting and useful for visualization designers.

Since the model for which we argue predicts that conceptual relationships should always have some basis in a perceived force that corresponds to a metaphor for that relationship, further study should focus on establishing such metaphors for other types of patterns, such as trends over time, uncertainty, and hierarchical relationships. One possible area of study is movement; by varying visual factors that suggest motion and with varying velocity and direction, can we predict the degree and nature of change that a viewer expects data to undergo over time? Many other possible directions for such research could be inspired by other physical metaphors for abstract relationships. This could expand upon our initial findings and begin to establish implied dynamics as a general principle in visualization analysis.

A final important question is that, while we have shown a correlation between perceptual attraction and conceptual similarity, our results so far do not clearly explain the direction of causality of this effect. That is, do people see marks as being related because they seem perceptually closer, or do marks seem closer because they seem related? It could be that visual illusions and biases that create illusory proximity lead to the impression of relationship between marks, or that structural elements that imply relationships encourage simulated attraction. More testing is needed to analyze exactly what is going on in this process.

8 Conclusion

In this work, we have found evidence that people simulate marks in a visualization as objects with forces acting on them, these simulations lead to perceptual errors similar to those found in studies of implied dynamics in psychology, and the amount of this error predicts the degree of conceptual relationship between the entities represented by these marks. These findings suggest that metaphorical interpretations of implied dynamics may be a general model for how visual elements are perceived as containing information about conceptual structure.

The connection between attraction and similarity points to a new way of looking at patterns in a visualization: not just in terms of overall symmetry and balance, but also in terms of explaining how marks work together to suggest complex meanings and inferences. By exploring this perspective and its implications for theory and practice, we can better understand how and why a visualization carries the information it does.

Acknowledgments

The authors wish to thank Paula Goolkasian for her guidance on statistical analysis and experimental design, and all the reviewers for their great help in shaping the final version of this paper.

References

[1] R. Arnheim. *The Power of the Center: A Study of Composition in the Visual Arts*. UC Press, 1988.

[2] C. Brewer and M. Harrower. ColorBrewer, http://www.colorbrewer.org/, 2002. Retrieved 12-4-2008.

[3] C. A. Burbeck and S. M. Pizer. Object representation by cores: Identifying and representing primitive spatial regions. *Vision Research*, 35(13):1917–1930, 1995.

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

[5] J. Freyd. The mental representation of movement when static stimuli are viewed. *Perception & Psychophysics*, 33(6):575–581, 1983.

[6] J. Freyd. Dynamic mental representations. *Psychological Review*, 94(4):427–438, 1987.

[7] J. Freyd, T. Pantzer, and J. Cheng. Representing statics as forces in equilibrium. *Journal of Experimental Psychology*, 117:395–407, 1988.

[8] J. J. Freyd and T. M. Pantzer. Static patterns moving in the mind. In S. M. Smith, T. B. Ward, and R. A. Finke, editors, *The Creative Cognition Approach*, pages 181–204. MIT Press, 1995.

[9] J. Heer and M. Bostock. Crowdsourcing graphical perception: Using Mechanical Turk to assess visualization design. In *Proceedings CHI*, pages 203–212, 2010.

[10] T. L. Hubbard. Target size and displacement along the axis of implied gravitational attraction: Effects of implied weight and evidence of representational gravity. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 23(6):1484–1493, 1997.

[11] A. Kittur, E. H. Chi, and B. Suh. Crowdsourcing user studies with Mechanical Turk. In *Proceedings CHI*, pages 453–456, 2008.

[12] W. Köhler. *Gestalt Psychology*. Liveright Publishing Company, 1947.

[13] R. Kosara and C. Ziemkiewicz. Do Mechanical Turks dream of square pie charts? In *BELIV Workshop*, pages 373–382, 2010.

[14] G. Robertson, R. Fernandez, D. Fisher, B. Lee, and J. Stasko. Effectiveness of animation in trend visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1325–1332, 2008.

[15] B. Tversky. Distortions in memory for visual displays. In S. R. Ellis and A. C. Grunwald, editors, *Pictorial Communication in Virtual and Real Environments*. CRC Press, 1991.

[16] J. Zacks and B. Tversky. Bars and lines: A study of graphic communication. *Memory & Cognition*, 27(6):1073–1079, 1999.

[17] C. Ziemkiewicz and R. Kosara. Implied dynamics in information visualization. In *Advanced Visual Interfaces*, pages 215–222, 2010.