Structure Matters: Effects of Semantic Relatedness and Proximity on Consumer Search and Integration Tasks

John Grishin and Douglas J. Gillan
Psychology Department
North Carolina State University
Raleigh, North Carolina 27695-7650 USA

Information displays should be clear and easily understood. This research examined whether principles developed by Kosslyn (1989) and Carswell and Wickens (1987) for charts, graphs, and object displays could be extended, or adapted, to another type of display, the food item package. We hypothesized that a food package on which label items had been arranged according to their similarity, or semantic relatedness, would facilitate better user performance than a package on which label items had been arranged in other ways. Participants rated the semantic relatedness of 12 label items found on a common food item package. Using multi-dimensional scaling (MDS) outputs from the ratings, we created three versions of a consumer cough drop package: 1) Similarity version—label elements that received higher similarity ratings were depicted closer together than elements with lower similarity ratings, 2) Dissimilarity version—elements that received higher similarity ratings were depicted farther apart than elements with lower similarity ratings, 3) Random version—rating values were randomly assigned to the pairs of elements. We tested user performance on search tasks and integrative tasks on each of the three versions. We hypothesized that the Similarity version would produce the best user performance and the Dissimilarity version would produce the worst. Results only partially supported the hypotheses. On the search tasks, the best performance was achieved on the Similarity and Dissimilarity versions, and the worst on the Random version. On the integrative tasks, the version made no difference in performance. Possible reasons for these results are discussed. Similar results by Fitts and Deininger (1954) and Morin and Grant (1955) suggest that performance on tasks are superior when the relationships are in an ordered structure, rather than randomly assigned, possibly because ordered structures make possible the development of search strategies, whereas random arrangements do not.

INTRODUCTION

An information display should convey information in a way that is both clear and easy to understand. Kosslyn (1989) noted “most people have had the experience of opening a well-known national news magazine and puzzling over a chart or graph, trying to figure out what it is about and what it is supposed to be telling the reader” (p. 185). To address this problem, Kosslyn developed a set of acceptability principles. Charts and graphs that were designed according to these principles would be “unambiguous and easily apprehended” (p. 185). Among these acceptability principles was the principle of perceptual organization—that is, elements on a display should be arranged according to “well-known principles” (p.195) that determine how those elements will be grouped together into perceptual units, including the Gestalt laws of good continuity, proximity, good form, and the established principles of dimensional structure that determine how and when dimension of visual structure such as size and height are grouped perceptually.

Another useful principle for designing clear and easily understood displays is Carswell and Wickens’ (1987) proximity compatibility principle. The proximity compatibility principle states that, “to the extent that information sources must be integrated, there will be a benefit to presenting those dimensions in an integrated (i.e. objectlike) format. That is, high display proximity helps in tasks with high mental proximity” (Wickens, 1995). This means that, two pieces of information with close mental proximity should be placed close together on a display, as long as doing so doesn’t reduce them to clutter (Wickens & Andre, 1990).

The current research proposes to extend the general approach used in Kosslyn’s method of analysis and Carswell and Wickens’ proximity compatibility principle to another common type of display, the food product package. Given the potential safety issues associated with ambiguous or unclear food labeling, the food product package is a natural candidate for a design that conveys information clearly and is easily understood.

This research tested whether a package label that was designed in such a way that label syntax—specifically the spatial proximity of elements on the label—corresponded to semantic relatedness of elements on the label could facilitate better user performance on tasks involving information search.

Research questions and hypotheses

We used multi-dimensional scaling (MDS) to determine the mental proximity of elements on 2-D food product package. We asked participants to rate the psychological distance, or similarity, of twelve (12) items commonly found on food item packages and then performed MDS analyses with these similarity ratings to generate three geometric configurations. In the first configuration, items with high similarity were depicted closer together. In the second configuration, the principle that was used to generate the Similarity solution was reversed, that is, highly related items were depicted farther apart. The third configuration was generated by randomly assigning the ratings to the pairs of items. We named the configurations “Similarity, Dissimilarity, and Random, respectively, and created three versions of a common food item package based on them. Then we tested the three layouts.

Hypothesis 1 (H1): User performance on search tasks will be better on layouts where arrangement of elements (syntax) mapped to semantic relatedness than on layouts where elements were arranged randomly, or in reverse of semantic relatedness.

Hypothesis 2 (H2): User performance on integration tasks would be better on layouts where arrangement of
elements (syntax) mapped to semantic relatedness than on layouts where elements were arranged randomly, or in reverse of semantic relatedness. The hypotheses are summarized below:

H0:  Mapped Layout = Random layout = Reversed layout
H1, 2:  Mapped Layout > Random layout > Reversed layout

**METHOD**

There were two phases for the current research entitled Phase 1: Creating the layouts and Phase 2: Testing the layouts. Each is detailed in the Materials and Procedure section below.

**Materials and Procedure**

**Phase 1: Creating the layouts:** Before we could create the layouts, we had to determine the elements that we would include in them, and we had to develop a method for determining the degree of semantic relatedness between those elements.

**Operationalizing semantic relatedness**

Step 1 Identifying the elements to include on the layouts – We conducted an informal census of common food items in local supermarkets and determined that the elements in Table 1 were common to most food item packages.

<table>
<thead>
<tr>
<th>Element name</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batch number</td>
<td>batch_num</td>
</tr>
<tr>
<td>Choking hazard</td>
<td>choke_haz</td>
</tr>
<tr>
<td>Company logo</td>
<td>co_logo</td>
</tr>
<tr>
<td>Manufacturer contact information</td>
<td>cont_info</td>
</tr>
<tr>
<td>Excluded items (e.g. sugar free, gluten free)</td>
<td>excl_items</td>
</tr>
<tr>
<td>Expiration date</td>
<td>exp_date</td>
</tr>
<tr>
<td>Flavor</td>
<td>flavor</td>
</tr>
<tr>
<td>Ingredients</td>
<td>ingred</td>
</tr>
<tr>
<td>Nutrition facts</td>
<td>nutr_facts</td>
</tr>
<tr>
<td>Opening instructions</td>
<td>open_instr</td>
</tr>
<tr>
<td>Product logo</td>
<td>prod_logo</td>
</tr>
<tr>
<td>Universal product code</td>
<td>upc_code</td>
</tr>
</tbody>
</table>

Table 1 Twelve elements common to food item packages

Step 2 Establishing the degree of semantic relatedness between elements – We established the degree of semantic relatedness between the twelve elements by asking participants to compare elements and rate their similarity. Using the survey application, Qualtrics, we developed a computer program that displayed the twelve elements to participants, two at a time. Between the twelve elements, there were sixty-six (66) possible unique pairings. Participants were shown each of the 66 pairings, one at a time, and asked to rate the similarity between the elements in the pair on a scale of zero to ten. Zero (0) indicated that the elements were least similar, and ten (10) indicated that the elements were most similar.

**Participants.** We used the Internet polling site, Mechanical Turk, to recruit one hundred (100) participants. Participants were paid $0.55 each for their participation. Of the 100 participants, ten did not provide ratings for all 66 pairs of elements, so we did not include their ratings in our analysis. Additionally, four of the respondents gave only two rating values, or fewer, for each of the 66 pairings, suggesting that they might have misunderstood the assignment. For this reason, we excluded their ratings from our analysis. This left us with similarity ratings from eighty-six (86) participants. We used these data to calculate the mean similarity rating for each of the 66 pairs of elements.

Step 3 MDS analysis of similarity ratings – We analyzed the results of the similarity ratings in Table 2 with the multidimensional scaling (MDS) tool in SPSS. MDS is a set of mathematical techniques that enables a researcher to uncover the “hidden structure” or theoretical meaning of data (Kruskal & Wish, 1976). We applied MDS to the ratings of psychological distance between stimulus elements. The MDS tool uses these ratings as inputs for an algorithm that generates a spatial representation consisting of a geometric configuration of points. Each point on the configuration corresponds to one stimulus element. The geometric configuration reflects the proximity values between elements. With the similarity ratings as inputs, we used SPSS to generate three MDS solutions. The first MDS had the twelve elements arranged so that elements with high similarity ratings were closer together than elements with low similarity ratings. We named this representation Similarity. In the second MDS solution, the twelve elements were arranged so that elements with high similarity ratings were farther apart than elements with low similarity ratings. We named this representation Dissimilarity. In the third MDS solution, a random number generator was used to ensure that the elements on the representation were not arranged in a planned order. We named this representation Random.

**Create layouts using the MDS solutions**

We chose a Hall’s lozenge package as a representative example of a typical consumer goods package. Figure 1 shows the Hall’s package layouts based on the three MDS solutions.
Figure 1. The three Halls package layouts arranged in accordance with the three MDS solutions. On Similarity, elements that received higher similarity ratings are depicted closer together than elements with lower similarity ratings. On Dissimilarity, elements that received higher similarity ratings are depicted farther apart than elements with lower similarity ratings. On Random, the rating values have been randomly assigned to the pairs of elements.

Phase 2: Testing the layouts: We developed two tasks to test the layouts: 1) Search 2) Integration.

1) Search for semantically related/unrelated items – We used the application programming interface (API) for survey software, Qualtrics, to develop an interactive search task to test whether there was any difference between the three layouts—Similarity, Dissimilarity, Random—in participants’ response times in searching for elements on the layout. Participants completed a familiarization section in which he/she was shown the 12 elements from Table 1. Then the participant was shown instructions informing them which element they should search for on the trial that was about to begin (e.g. EXPIRATION DATE), and that clicking on a gray box would display the layout and begin the trial.

The expiration date in the preceding instructions was the “orientation” element. When the participant clicked the gray box, a clock started and one of the 3 layouts—Similarity, Dissimilarity, Random—appeared. The participant located the orientation element from the preceding instructions and clicked on it. The clock stopped and the elapsed time was recorded. All elements on the layout were clickable, as was the background, so it was possible for the participant to click on the wrong element. When the participant clicked on the orientation element, a black screen obscured the layout and instructions appeared directing the participant which layout element (e.g. KRAFT FOODS) to search for when the next screen appeared. Clicking the OK button made the next screen appear.

The box containing these instructions and the OK button were centered directly on top of the orientation element, and the black screen obscured the rest of the layout so that no part of the layout was visible until the participant clicked the OK button. The KRAFT FOODS logo was the “target” element. When the participant clicked the OK button, the layout reappeared, the clock resumed, and the participant searched for the target element. The participant located the target element and clicked on it. The clock stopped, and the elapsed time was recorded. A gray box appeared congratulating the participant on his/her completion of the practice session and that they should click the red next button to begin the actual sessions. When the participant clicked the red next button, they were presented with sessions very similar to the practice session. Each of the 12 elements on the layout was used as an orientation element in two searches for a total of 24 searches. The 12 orientation elements were presented in random order to each participant using a randomization algorithm in the Qualtrics API. The target elements in each search were chosen for the degree of their relatedness to the orientation element, one having a high relatedness rating to the orientation item, the other having a low relatedness rating to the orientation item. This meant that two searches were conducted from each element on the layout—for an element that was highly related to the orientation element and for an element that was less highly related to the orientation element. Participants were randomly assigned to one of the three layouts—Similarity, Dissimilarity, Random—at the beginning of the trial using a randomization algorithm in the Qualtrics API. Once the layout was assigned, it did not change, so participants who received the Similarity layout, for example, performed all 24 search sessions on the Similarity layout.

2) Integration task – We developed three (3) integration tasks. Each integration task required that the participant retrieve and integrate information from a combination of elements on the layout to complete the task successfully. We named the integration tasks 1) Kraft, 2) Doctor, and 3) Sugar. Two measures were recorded from each integration task: 1) whether the participant correctly completed the task. 2) The amount of time that it took to complete the task. The order of presentation of these integration tasks was counterbalanced to prevent order effects. However, we did not counterbalance the search task portion of the experiment with the integration portion of the experiment. In order to prevent the search task results becoming
contaminated by participants’ familiarity with the layout, we decided that all participants would perform the 30 search tasks before they performed the integration tasks.

Participants. We used the Internet polling site, Mechanical Turk, to solicit 90 participants using the same screening criteria as in Phase 1. Of the 90 participants, two missed the target items in the search section more than 50% of the time, so we excluded their data from our analyses. This left us with data from eighty-eight (88) participants.

RESULTS

1) Search task – We compared the three layout conditions to see whether layout condition affected participants’ response times on the search task. The three layout conditions yielded similar overall mean search times (Similarity = 3.16 sec; Dissimilarity = 3.13 sec; Random = 3.16 sec) and response times did not vary significantly; \( F(2, 255) = 2.20, p = .11, \eta^2 = .007 \). All three conditions showed very high accuracy (Similarity=96%, Dissimilarity=97%, Random=96%, Total accuracy = 96%). The differences in accuracy were not significant by a Chi Square test \( (df=2, \ Chi\ Square=1.1, p > .05) \).

We also conducted an analysis of a multiple regression model \( \text{LOG10RT} = a + b1(\text{Layout}) + b2(\text{Semantic Relatedness}) + b3(\text{Layout*Semantic Relatedness}) \) to determine the interaction between the two independent variables, layout and semantic relatedness, on participants’ response times on the search task. The values for Semantic Relatedness in the regression model were supplied by participant ratings shown in Tables 2 and 3. Results showed that there was no significant main effect of layout or semantic relatedness on response times. However, there was a significant interaction between layout and semantic relatedness \( F(8, 1993) = 2.49, p = .01 \). Results of the regression analyses are presented in Table 2.

Table 2 Results of multiple regression model \( \text{LOG10RT} = a + b1(\text{Layout}) + b2(\text{Semantic Relatedness}) + b3(\text{Layout*Semantic Relatedness}) \)

<table>
<thead>
<tr>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic relatedness (RU_A)</td>
<td>4</td>
<td>0.864</td>
</tr>
<tr>
<td>Layout (Similarity, Dissimilarity, Random)</td>
<td>2</td>
<td>2.039</td>
</tr>
<tr>
<td>RU_A * Layout</td>
<td>8</td>
<td>2.485</td>
</tr>
</tbody>
</table>

The interaction between layout and semantic relatedness shown in Table 2 occasioned additional regression analyses to determine the effect of semantic relatedness (shown as RU_A in the x axis label in Figure 2) on participants’ response times (shown as LOG10 of Response Time on the y axis label in Figure 2) on the search task. As Figure 2 shows, search times tended to decrease as semantic relatedness increased in both the Similarity and Dissimilarity conditions. In contrast, search times increased slightly as semantic relatedness increased in the Random condition.

![Figure 2](image-url) Search times (LOG10RT) as a function of semantic relatedness (RU_A). 7a. Similarity condition—elements with high semantic relatedness (higher RU_A) were depicted closer together than elements with low semantic relatedness. 7b. Dissimilarity condition—elements with high semantic relatedness were depicted farther apart than elements with low semantic relatedness. 7c. Random condition—elements were depicted in a random arrangement with no systematic relationship between semantic relatedness and placement on the layout.

2) Integration tasks – A 3 x 3 (3 layouts x 3 tasks) analysis of variance revealed no significant differences between layouts in response times on the integration tasks, \( F(2, 255) = .386, p = .68 \). All three conditions showed high accuracy (Similarity=84%, Dissimilarity=92%, Random=93%, Total accuracy = 89%), although slightly lower than for the search task. The differences between conditions in the integration task were not significant by a Chi Square test \( (df=2, \ Chi\ Square=4.7, p > .05) \).

Additionally, there was no significant interaction between layout and question in response time, \( F(4, 255) = .197, p = .94 \). This analysis included both incorrect and correct answers. An additional 3 x 3 analysis of variance that included only correct answers to the integration tasks also revealed no significant differences between layouts in
response times on the integration tasks, \( F(2, 255) = .478, p = .62 \), and there was no significant interaction between layout and question, \( F(4, 255) = .288, p = .89 \).

**DISCUSSION**

The results showed that, with the Similarity and Dissimilarity layouts, the two layouts in which the arrangement of elements was mapped to semantic relatedness or in reverse of it, search task performance improved as the semantic relatedness of the items on the layouts increased. However, on the Random layout, search task performance did not change as semantic relatedness increased. These results only partially support H1. H1 suggests that performance on search tasks will be better on layouts where the arrangement of elements (syntax) is mapped to semantic relatedness than on layouts where elements are arranged randomly, or in reverse of semantic relatedness. Consequently, a layout with highly related elements placed far apart, as in the Dissimilarity condition, should produce the worst search task performance, but the data show that this was not the case.

Results such as the above are not unprecedented. Fitts and Deininger (1954) found that performance on perceptual-motor tasks was influenced by the degree of correspondence between the stimulus and response sets as measured by the degree to which S-R relationships corresponded to “population stereotypes” (p. 491). Fitts and Deininger varied the degree of S-R correspondence by creating three correspondence conditions: 1) maximum, 2) mirrored, and 3) random correspondence. In the maximum correspondence condition, stimuli were matched with responses to provide “maximum agreement with population stereotypes” (p. 486). This resembles the Similarity condition in the present research. In the mirrored correspondence condition, “S-R mating consisted of reversing the left-right relations in the maximum correspondence” (p. 485) such that these S-R relationships were now in reverse of those in the maximum correspondence condition. This is analogous to the Dissimilarity condition in the present research. In the random correspondence condition, stimuli were randomly assigned to responses as they were in the Random condition of the present research. Fitts and Deininger found that performance in the maximum and mirrored conditions was superior to performance in the random condition, suggesting that performance on perceptual-motor tasks is superior when the S-R relationships are in some kind of ordered structure rather than randomly assigned. Morin and Grant (1955) had similar findings. Participants made key-pressing responses to light stimuli under nine degrees of correspondence between light and keys represented by the following values: +1.00, +0.86, +0.57, +0.29, 0.00, -0.29, -0.57, -0.86, and -1.00. The value +1.00 indicated direct correspondence between the light stimuli and the key press responses, much like Fitts and Deininger’s (1954) maximum correspondence condition and the Similarity condition in the present study. The value of -1.00 indicated that the display-control relationships between light stimuli and key presses was the reverse of that in the +1.00 condition, much like Fitts and Deininger’s mirrored condition and the Dissimilarity condition in the current study. The value 0.00 indicated no ordered relationship at all between the light stimuli and key press responses, much like the random conditions in the other studies. Like the researchers in the present study, Morin and Grant, predicted that “performance should be best for a correlation of +1.00 and show progressively more degradation as the correlation approaches -1.00” (p. 40), suggesting a relationship similar to the one implied by H1 in the present research: +1.00 > 0.00 > -1.00. They found instead that “… a complete reversal of display-control relationships (\( \tau = -1.00 \)) yields better performance than do the more irregular rearrangements (\(-0.86 \leq \tau \leq -0.86\))” leading them to conclude that a subject “responds readily to orderliness, direct or reversed (high positive or negative values of \( \tau \))” (p. 45). Morin and Grant’s findings, like those in Fitts and Deininger’s and the present research, again suggest that performance is superior when the S-R relationships are in some kind of ordered structure rather than randomly assigned.

Hypothesis 2 (H2) asserts that user performance on integration tasks would be better on layouts where arrangement of elements (syntax) was mapped to semantic relatedness than on layouts where elements were arranged randomly, or in reverse of semantic relatedness. Since each of the integration tasks contained a search component and, as already noted, search times improved significantly as a function of increasing semantic relatedness on two of the three layouts, it might initially seem surprising that the results did not support H2. However, given the small interaction between layout and semantic distance and given that the overall mean response times on the integration tasks did not differ significantly by layout (Figure 9), it is probable that the effect of Layout on the integration tasks was simply too small for detection by this experiment. It is also possible that practice effects may have affected response times. Participants performed the integration tasks after they had performed the search tasks. Participants were randomly assigned either to the Similarity, Dissimilarity, or Random layout. They performed 24 search tasks on that layout followed by the integration tasks. By the time they performed the integration tasks, participants were well familiar with the layout and the general location of elements on it.

This research could be useful in designing user interfaces that facilitate rapid search. The main finding of this research—that an ordered mapping to population stereotypes, direct or reversed, improves search task performance—might be applied to other kinds of product labels and other complex displays. The most obvious and specific application would be in the further design of food item packaging.

**REFERENCES**


