

ABSTRACT

GRISHIN, JOHN Laying Out Information Displays Based on the Semantic Relatedness of Display Elements (Under the direction of Dr. Douglas J. Gillan)

Information displays should be clear and easily understood. Kosslyn's (1989) acceptability principles and Carswell and Wickens' (1987) proximity compatibility principle are useful "rules of thumb" for designing clear and easily understood displays. This research examined whether these principles 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 and we subjected these similarity/proximity ratings to multi-dimensional scaling (MDS) analyses. Using the MDS outputs, we created three (3) alternative versions of a common food item 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 (3) 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 this hypothesis. 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 perceptual-motor tasks is superior when the S-R 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.

© Copyright 2015 by John Grishin

All Rights Reserved

Laying Out Information Displays Based on the Semantic Relatedness of Display Elements

by
John Grishin

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Psychology

Raleigh, North Carolina

2015

APPROVED BY:

Douglas J. Gillan
Committee Chair

Anne McLaughlin

Stan Dicks

BIOGRAPHY

John Grishin has a Bachelor's Degree in Philosophy from the University of North Carolina at Chapel Hill and a Master's Degree in Technical Communication from North Carolina State University. While completing the degree requirements for his MS in Technical Communication, he became interested in the cognitive aspects of user interaction with electronic media, which led to his enrolling in the Human Factors and Applied Cognition PhD program at N.C.S.U. His chief research interests remain the usability of electronic media.

ACKNOWLEDGMENTS

The author would like to acknowledge and thank Will Walkington, Jim Creager, Michael Otteni, and Douglas J. Gillan for their indispensable roles in realizing this study. Will created the artwork for all the stimuli. Jim wrote the JavaScript program that enabled the user ratings of label elements. Michael wrote the JavaScript program that tested user performance on search tasks and integration tasks. Dr. Gillan advised and guided me through the entire process. Without their contributions, this study would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES	v
LIST OF FIGURES	vi
Introduction	1
Research questions and hypotheses	5
Method	5
Participants	5
Materials and Procedure	6
Phase 1: Creating the layouts	6
Operationalizing semantic relatedness	6
Create layouts using the MDS solutions	12
Phase II: Testing the layouts	16
Results	24
Measuring response times on the search task	24
Measuring response times on the integration tasks	26
Discussion	27
REFERENCES	36

LIST OF TABLES

Table 1 Twelve elements common to food item packages 6

Table 2 Mean Similarity Ratings for the 66 Pairs of Elements 8

Table 3 Orientation elements and their two target elements 19

Table 4 Distribution of layouts 23

Table 5 Results of multiple regression model $(\text{LOG}_{10}\text{RT}) = a + b_1(\text{Layout}) + b_2(\text{Semantic Relatedness}) + b_3(\text{Layout} * \text{Semantic Relatedness})$ 25

Table 6 Results of Pearson correlations comparing the physical distances of elements on the original product label with physical distances on the Similarity, Dissimilarity, and Random layouts 34

LIST OF FIGURES

<i>Figure 1.</i> Results of the MDS solution using the ratings from Table 1 with elements that received higher similarity ratings depicted closer together than elements with lower similarity ratings.....	9
<i>Figure 2</i> Results of the MDS solution using the ratings from Table 1 with elements that received higher similarity ratings depicted farther apart than elements with lower similarity ratings.....	10
<i>Figure 3</i> Results of the MDS solution using the ratings from Table 1, but with the rating values having been randomly assigned to the pairs of elements.	11
<i>Figure 4</i> Elements on the Halls lozenge package arranged in accordance with the Similarity MDS solution in Figure 1.	13
<i>Figure 5</i> Elements on the Halls lozenge package arranged in accordance with the Dissimilarity MDS solution in Figure 2.	14
<i>Figure 6</i> Elements on the Halls lozenge package arranged in accordance with the Random MDS solution in Figure 3.	15
<i>Figure 7</i> 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.....	26
<i>Figure 8</i> Results of ANOVA revealing no significant differences between layouts in response times on the integration tasks, $F(2, 255) = .386, p = .68$ and no significant interaction between layout and question $F(4, 255) = .197, p = .94$	27
<i>Figure 9</i> Results of ANOVA using correct answers only. There were no significant differences between layouts in response times on the integration tasks, $F(2, 255) = .478, p = .62$. and no significant interaction between layout and question, $F(4, 255) = .288, p = .89$..	27

Introduction

A Model of Layout Design Based on Mental Proximity

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). Kosslyn developed a method of analyzing such displays (quantitative graphs) at three levels, syntax, semantics, and pragmatics, as well as a set of *acceptability principles* based on this three-level analysis (pp. 185-197). 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, 1992, p. 98). 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).

Carswell and Wickens' (1987) proximity compatibility principle can be integrated into Kosslyn's (1989) system of analysis at all three levels. At the (1) syntactic level, Kosslyn focused on the organization of the elements on the display. For example, a syntactic analysis might examine the ways in which a viewer groups elements on a display perceptually into units. Therefore, an arrangement of elements in close physical proximity to facilitate their integration according to the proximity compatibility principle falls under the syntactic level of analysis. At the (2) semantic level, Kosslyn focused on the meanings of the element on a display. The semantic analysis addresses what elements on a display depict or signify, and the literal meaning that arises from the relations among those elements. Elements that are similar in meaning are said to have high mental proximity. Therefore, grouping elements with high mental proximity in close proximity on a display according to the proximity compatibility principle falls under the semantic level of analysis. Finally, Kosslyn's (3) pragmatic analysis focused on the ways in which elements on a display convey information beyond their literal meaning. Another characteristic of the proximity compatibility principle is that the integration of information dimensions into an object-like form tends to produce emergent features, like area or shape, that can help in integration task requirements (Wickens, 1992, pp. 98-99). Such emergent features fall under the pragmatic level of analysis, however the present research will focus on the application of the proximity compatibility principle only at the syntactic and semantic levels of analysis.

Although Kosslyn (1989) focused on charts and graphs, he suggested that his method for analyzing displays could be extended to other types of displays. Similarly, though Carswell and Wickens (1987) were chiefly concerned with object displays versus bar graphs, the proximity compatibility principle might be applied to other types of displays as well. In other research, Bettman, Payne, and Staelin (1986) examined the importance of spatial proximity between two related items, the costs and benefits of a product, in the design of product warning labels. Similarly, 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. The present study combined Kosslyn's (1989) method of analysis with a specific aspect of Carswell and Wickens' (1987) proximity compatibility principle—that high *display proximity* helps in tasks with high mental proximity—in the design of a consumer product package label. 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—corresponds to semantic relatedness of elements on the label can facilitate better user performance on tasks involving information search.

To arrange elements on a display in order of their semantic relatedness, it is first necessary to develop a measure of relatedness. McDonald, Stone, Liebelt, and Karat (1982) proposed a methodology for obtaining such a measure based on judgments of relatedness. In designing a fast-food keyboard, a common menu-based display, McDonald et al. obtained

estimates of relatedness “from ‘experts’ for all pairs of system actions or objects” that are found on the keyboard (McDonald, Stone, Liebelt, & Karat, 1982, p. 421). The results were subjected to scaling analyses, such as multidimensional scaling (MDS), and the results of these analyses were then used as the basis for arranging the elements on the keyboard. McDonald, Dayton, and McDonald (1988) used this methodology to demonstrate further that a general methodology based on judgments of relatedness can be used to design menu layouts. Others (Hooley, 1984; Tullis, 1985; Tullis & Bied, 1988) have used similar methodologies for a range of applications from the perceived similarity of cigarette brands to the design of a space station layout. The present research used a similar methodology to determine the semantic relatedness, or mental proximity, of elements on a food product package. We used MDS to determine the mental proximity of elements on 2-D food product package. MDS is a set of mathematical techniques that enable a researcher to uncover the “hidden structure” or theoretical meaning of data (Kruskal & Wish, 1976). The data for MDS come from participants’ ratings of the psychological distance between stimulus objects, and it uses these proximity ratings as inputs. The chief output is a spatial representation consisting of a Euclidean geometric configuration of points. Each point on the configuration corresponds to one object. The geometric configuration reflects the proximity rating values. For the current research, 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 these configurations “Similarity, Dissimilarity, and Random, respectively. We then had participants perform tasks with the three layouts and compared performance on the tasks.

Research questions and hypotheses

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

H_{1,2}: Mapped Layout > Random layout > Reversed layout

Method

Participants

There were two phases for the current research and participants for the two phases were recruited separately. The details of participant recruitment for the two phases, entitled **Phase 1: Creating the layouts** and **Phase 2: Testing the layouts**, is detailed in the Materials and Procedure section below.

Materials and Procedure

We conducted this study in two phases. In **Phase 1: Creating the layouts**, we created the layouts based on semantic relatedness of elements on the layout. In **Phase 2: Testing the layouts**, we tested the layouts for the effects of semantic relatedness on performance.

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. The elements in Table 1 were common to most food item packages.

Table 1 Twelve elements common to food item packages

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

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. We used Mechanical Turk's built-in screening tools to limit participants to those from the United States, and to those with a hit approval rate of 95% or higher. 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. The instructions to participants clearly asked for ratings on a scale of 0 to 10, and the fact that these participants used only two values between 0 and 10 for all 66 pairs of elements suggested that they might have misunderstand 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. Table 2 shows the mean similarity rating for the 66 pairs of elements.

Table 2 Mean Similarity Ratings for the 66 Pairs of Elements

	batch num	choke haz	co logo	cont info	excl items	exp date	flavor	ingred	nutr facts	open instr	prod logo	upc cod
batch_num	-	3	3	4	3	4	4	3	3	3	4	5
choke_haz	-	-	4	4	2	2	2	2	3	2	3	2
co_logo	-	-	-	3	3	3	3	3	3	3	4	3
cont_info	-	-	-	-	3	3	4	4	5	3	6	3
excl_items	-	-	-	-	-	3	5	6	5	2	4	3
exp_date	-	-	-	-	-	-	3	3	3	3	4	3
flavor	-	-	-	-	-	-	-	5	4	3	5	3
ingred	-	-	-	-	-	-	-	-	6	3	5	3
nutr_facts	-	-	-	-	-	-	-	-	-	3	5	4
open_instr	-	-	-	-	-	-	-	-	-	-	3	3
prod_logo	-	-	-	-	-	-	-	-	-	-	-	4
upc_cod	-	-	-	-	-	-	-	-	-	-	-	-

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 from Table 2 as inputs, SPSS generated the MDS solutions depicted in Figures 1-3.

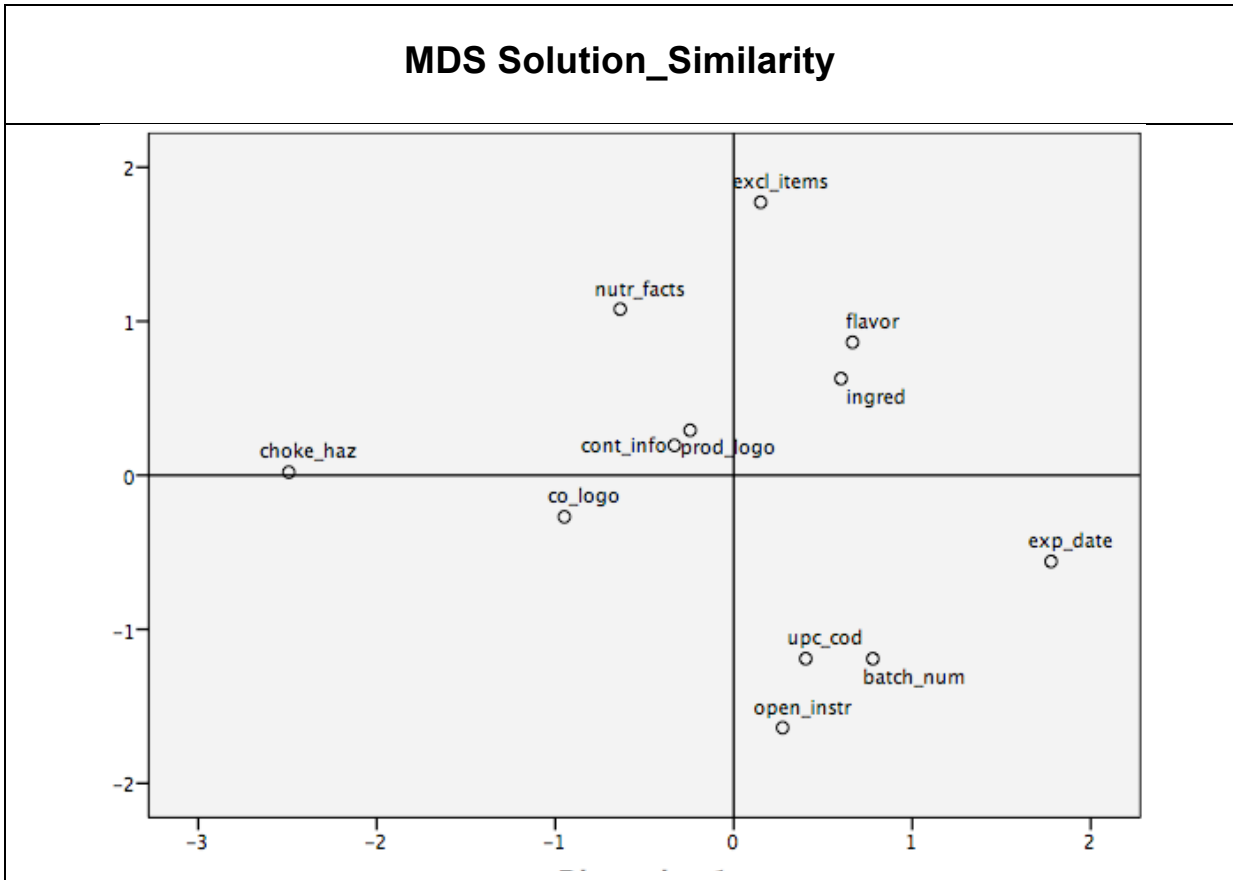


Figure 1. Results of the MDS solution using the ratings from Table 1 with elements that received higher similarity ratings depicted closer together than elements with lower similarity ratings.

In Figure 1, an MDS solution was generated with 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*.

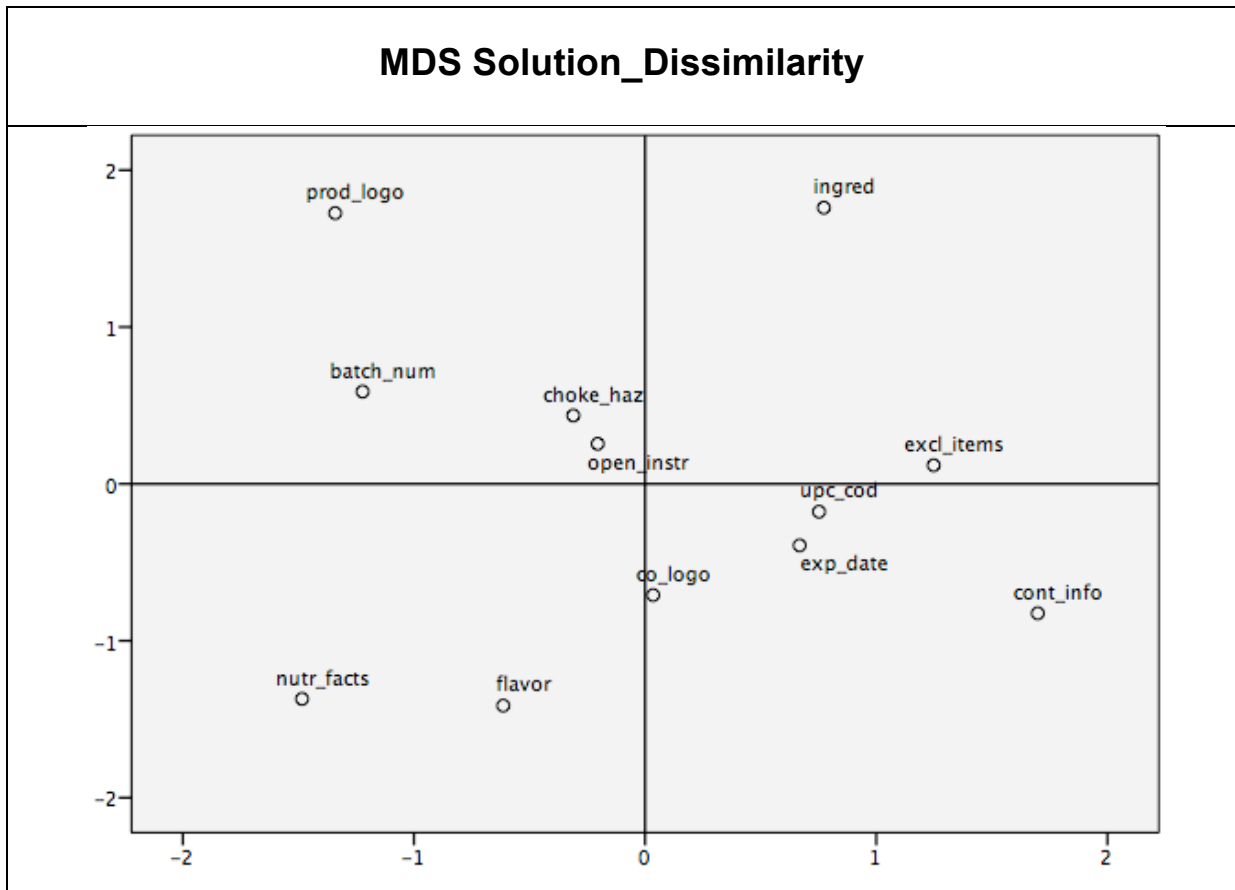


Figure 2 Results of the MDS solution using the ratings from Table 1 with elements that received higher similarity ratings depicted farther apart than elements with lower similarity ratings.

The representation in Figure 2 was generated using the reverse of the principle used to generate Similarity. Again using the ratings from Table 1, 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*.

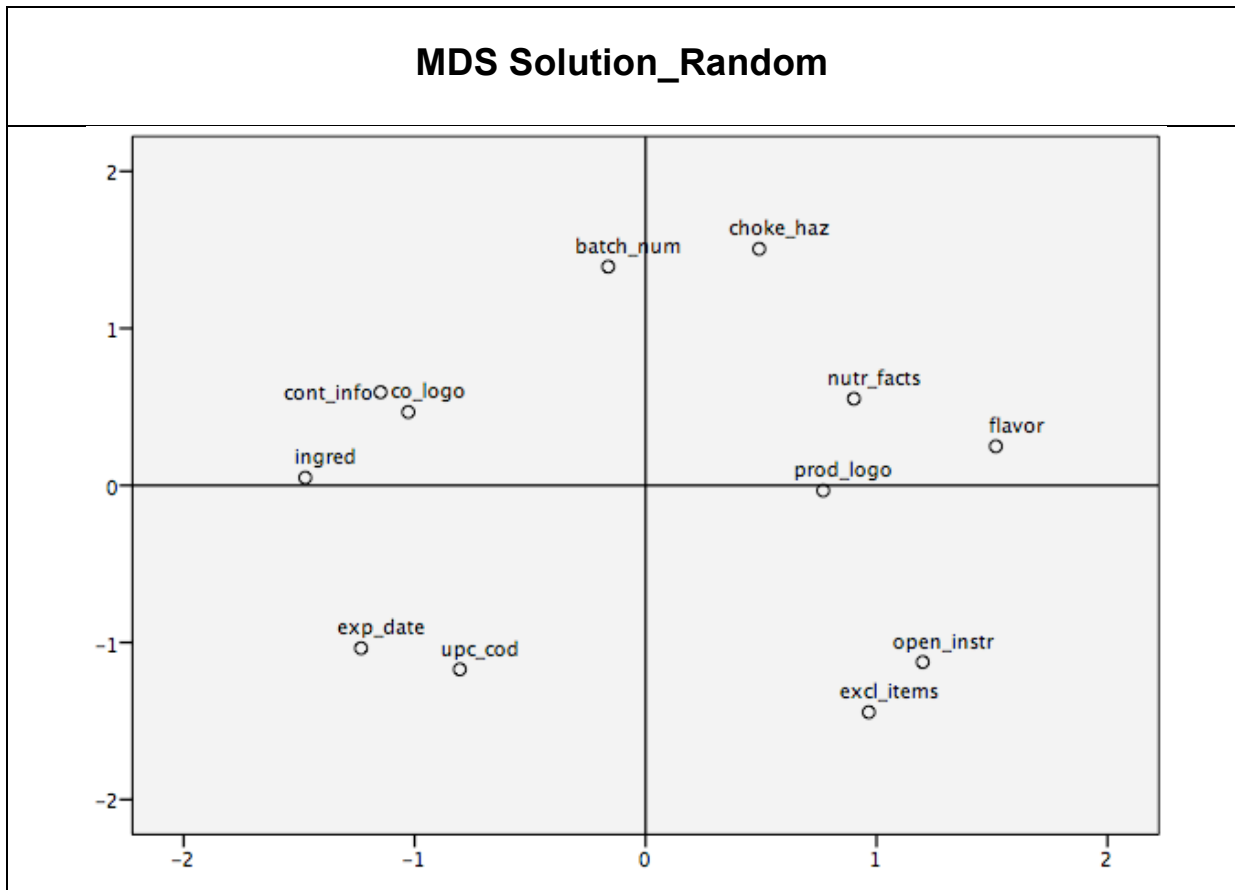


Figure 3 Results of the MDS solution using the ratings from Table 1, but with the rating values having been randomly assigned to the pairs of elements.

In Figure 3, we again used the similarity ratings from Table 1, but this time, the ratings were assigned randomly to the pairs of elements. First, we used Stat Trek's random number generator (<http://stattrek.com/statistics/random-number-generator.aspx>) to assign a random order to the pairs of elements, and then filled in the values from Table 1 in the randomly generated order. This assured that the elements on the representation were not arranged in a planned order. We named this representation *Random*. The Random MDS representation served as a baseline for later performance testing.

Create layouts using the MDS solutions

We chose a Hall's lozenge package as a representative example of a typical consumer goods package. Figures 4-6 show the Hall's package layouts based on the three MDS solutions.

1. Layout 1 Similarity; Sem→Syn (1.0): We arranged elements on the layout according to the level of semantic relatedness as depicted in the Similarity MDS solution. Elements of high similarity were placed closer together and elements of lower similarity were placed farther apart.

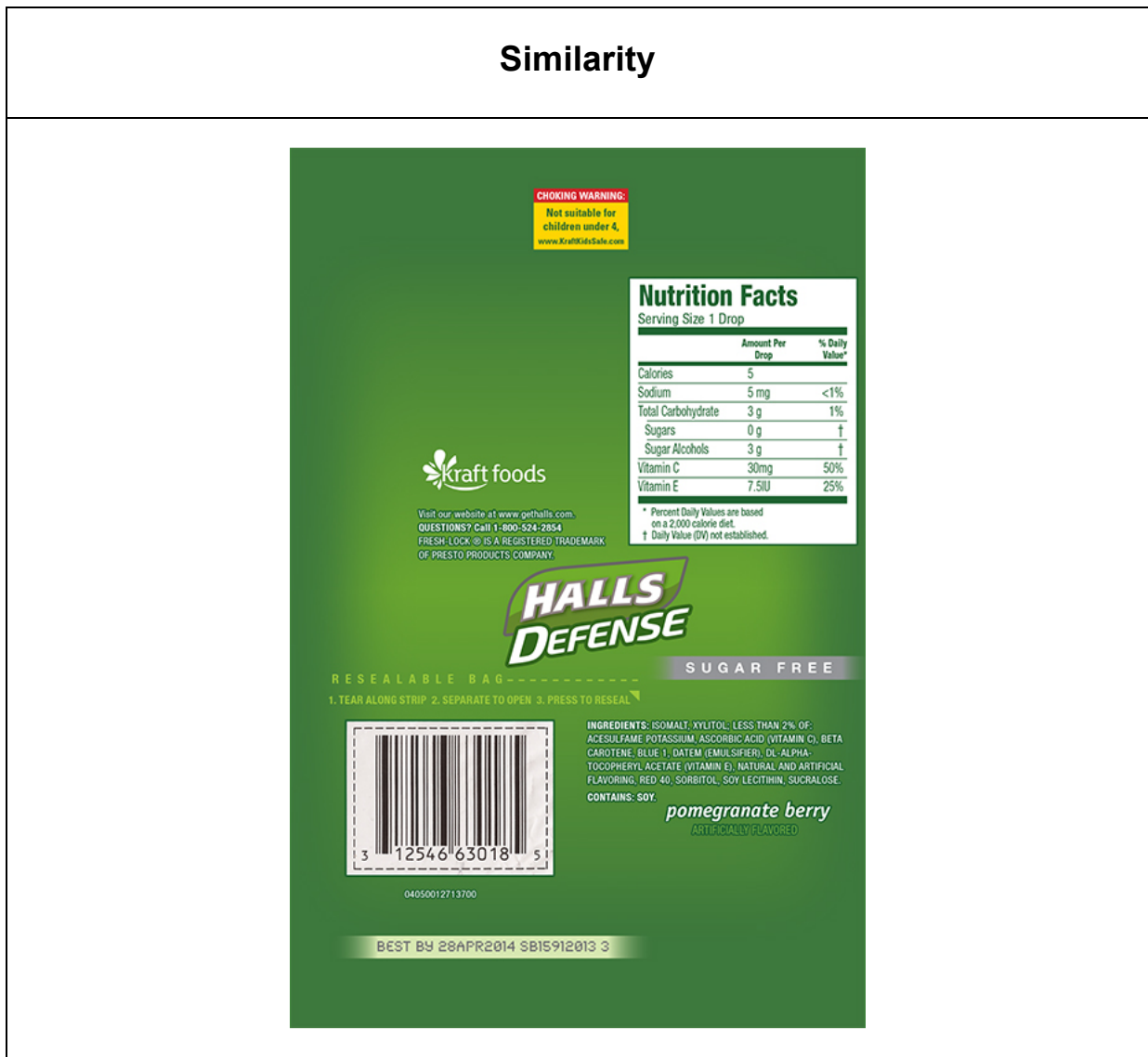


Figure 4 Elements on the Halls lozenge package arranged in accordance with the Similarity MDS solution in Figure 1.

2. Layout 2 Dissimilarity; Sem→Syn (-1.0): We arranged elements in reverse of the level of semantic relatedness as depicted in the Similarity MDS solution. Elements of high similarity were placed farther apart than elements of lower similarity.

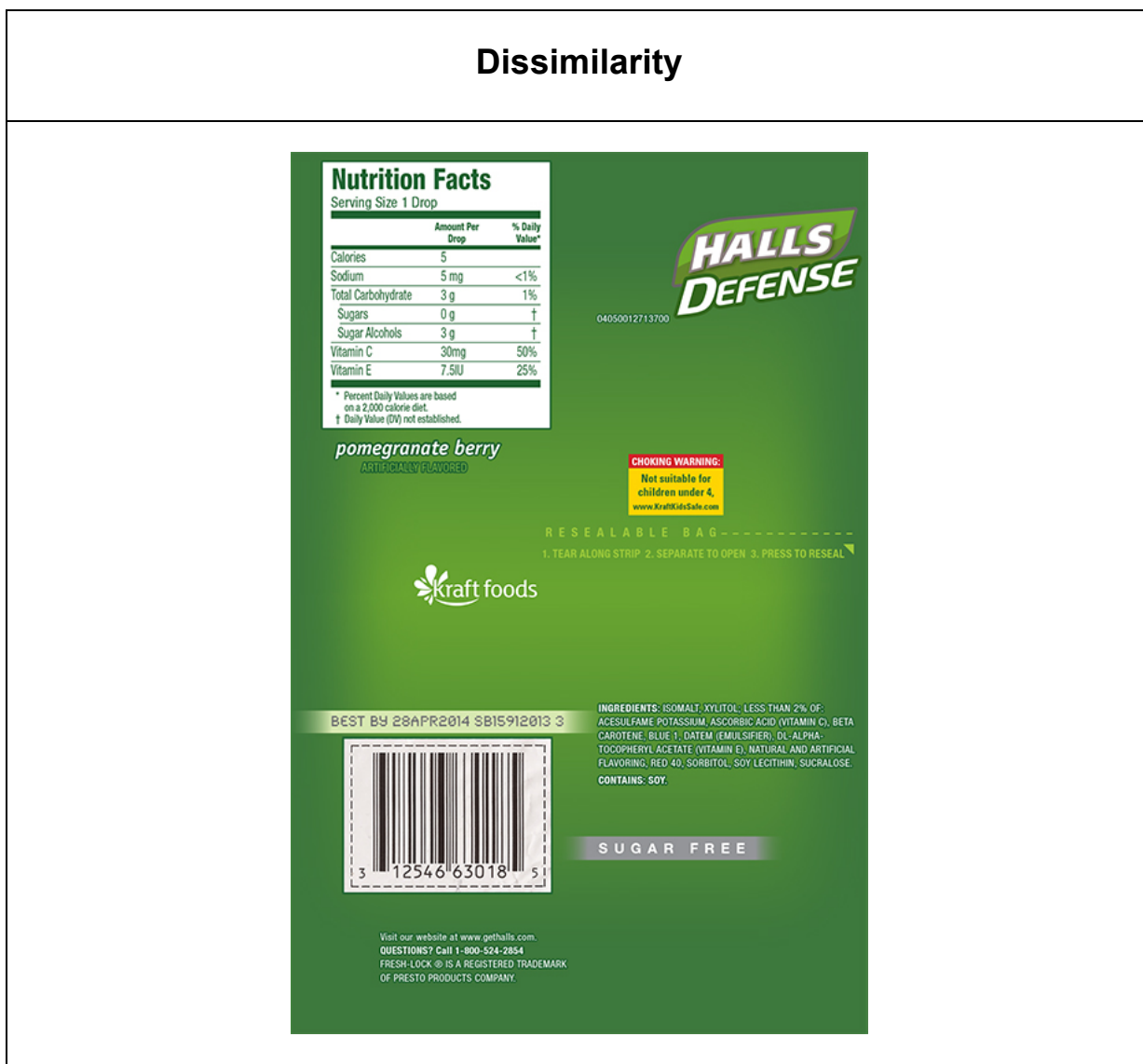


Figure 5 Elements on the Halls lozenge package arranged in accordance with the Dissimilarity MDS solution in Figure 2.

3. Layout 3 Random; Sem→Syn (0.0): We arranged elements so that there was no systematic relationship between semantic relatedness and placement on the layout.



Nutrition Facts
Serving Size 1 Drop

	Amount Per Drop	% Daily Value*
Calories	5	
Sodium	5 mg	<1%
Total Carbohydrate	3 g	1%
Sugars	0 g	†
Sugar Alcohols	3 g	†
Vitamin C	30mg	50%
Vitamin E	7.5IU	25%

* Percent Daily Values are based on a 2,000 calorie diet.
† Daily Value (DV) not established.

Phase II: 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. Before the actual search task began, the participant was asked to complete a demographic survey (Appendix 1). Then participant completed a familiarization section in which he/she was shown the 12 elements from Table 1. The 12 items were displayed, one at a time, to the participant for a minimum of three seconds. After three seconds had elapsed, a clickable next button appeared that enabled the participant to advance to the next item if they chose to do so. This forced participants to view the elements for at least three seconds while still allowing them to self-pace through the familiarization section. After participants had viewed all twelve items, they were given a practice search task to complete. The following instructions were displayed:

This part of the survey is a practice session. The 12 items you just looked at will be arranged on a single layout. You will be asked to locate one of the items and click on it. Once you have clicked the item, you will be given further instructions. When you are ready to start the practice session, click on the double arrow button at the bottom right of the screen.

Once the participant clicked on the double arrow button, a gray box appeared informing the participant which element they should search for on the trial

that was about to begin (e.g. EXPIRATION DATE) and that clicking on the 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 related to the orientation element. For example, we used batch number (“batch no”) as the orientation element in two sessions, the first with UPC code (“upc code”) as the target element, and the second with company logo as the target element. In Table 3, we see that the relatedness rating for batch no and upc code is 5, a relatively high rating, while the relatedness rating for batch no and co logo is 3, a relatively low rating.

Table 3 Orientation elements and their two target elements

<u>Orientation element</u>	<u>Target 1 (Highly related)</u>	<u>Target 2 (Less related)</u>
batch no	upc code (5)	co logo (3)
choke haz	co logo (4)	flavor (2)
co logo	prod logo (4)	excl items (3)
cont info	prod logo (6)	exp date (3)
excl items	ingred (6)	open instr (2)
exp date	batch no (4)	choke haz (2)
flavor	ingred (5)	choke haz (2)
ingred	nutr facts (6)	choke haz (2)
nutr facts	ingred (6)	open instr (3)
open instr	prod logo (3)	excl items (2)
prod logo	cont info (6)	upc code (3)
upc code	batch no (5)	choke haz (2)

Note: Numbers in parentheses indicate mean relatedness rating of the target element to the orientation element. Target 1 elements were rated high in relatedness to the orientation element. Target 2 elements were rated low in relatedness to the orientation elements. Complete relatedness ratings are found in Table 2.

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.** 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.

1) Kraft. After participants completed the search tasks, a screen appeared with the following instructions:

On this part of the survey, you will be shown a layout. Look at the layout and then answer the following question: Does the name of the product in the KRAFT FOODS logo appear again in the contact information? Click on the double arrow button at the bottom right of the screen to proceed.

When participants clicked on the double red arrow button, a gray box appeared with the following instructions visible to the right of the gray box;

When you click on the gray box on the left, you will be shown a layout. Please look at the layout and answer the question: Does the company name that appears in the KRAFT FOODS logo appear again in the CONTACT INFORMATION?

Just below the instructions were Yes/No buttons indicating that the participant was to answer the question by clicking either yes or no. When participants clicked on the gray box, the same layout—Similarity, Dissimilarity, Random—that the participants had been randomly assigned for the search task appeared, and the participant located the elements described in the instructions to complete the integration task by clicking either yes or no. The instructions remained visible to the participant as they were performing the integration

task, so they were able to refer back to the instructions if they forgot the question.

2) Doctor. After participants completed the Kraft task, a gray box appeared with the following instructions visible to the right of the gray box:

When you click on the gray box on the left, you will be shown a layout. Please look at the layout and consider the following scenario: Your doctor has warned you that it is not safe for you to use a product if BOTH these things are true: 1) the product contains beta carotene, AND 2) the amount of Vitamin E in the product is above 6 IU per serving. Based on your doctor's warning, decide whether the product shown in the layout is safe for you to use.

Just below the instructions were Yes/No buttons indicating that the participant was to answer whether the product was safe to use by clicking either yes or no. When participants clicked on the gray box, the same layout —Similarity, Dissimilarity, Random—to which the participant had been randomly assigned for the search task appeared, and the participant located the elements described in the instructions to complete the integration task by clicking either yes or no. The instructions remained visible to the participant as they were performing the integration task, so they were able to refer back to the instructions if they forgot the question.

3) Sugar. After participants completed the Doctor task, a gray box appeared with the following instructions visible to the right of the gray box:

For this part of the survey, you will again be shown a layout. On the layout is an item labeled " _____ Free", indicating that the product does not contain a certain ingredient. Find that label item, identify the substance that should not be contained in the product, and then determine from the layout whether that substance really is, or is not, contained in the product.

Just below the instructions were two buttons, one with the words “The ingredient is contained in the product,” the other with the words “The ingredient is NOT contained in the product.” When participants clicked on the gray box, the same layout —Similarity, Dissimilarity, Random—that the participants had been randomly assigned for the search task appeared, and the participant located the elements described in the instructions to complete the integration task by clicking one of the buttons to indicate that the ingredient was or was not contained in the product. The instructions remained visible to the participant as they were performing the integration task, so they were able to refer back to the instructions if they forgot the question.

Participants. We again used the Internet polling site, Mechanical Turk, to solicit 90 participants. Participants were paid \$0.55 each for their participation. We used Mechanical Turk’s built-in screening tools to limit participants to those from the United States, and to those with a hit approval rate of 95% or higher. 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 for the Search, Integration and Rating from eighty-nine (88) participants. As mentioned earlier, we used a randomizing algorithm from the Qualtrics API to assign one of the three layouts (Similarity, Dissimilarity, Random) randomly to participants. The distribution of the layouts is shown Table 4.

Table 4 Distribution of layouts

Dissimilarity	31
Similarity	29
Random	28
Total	88

Results

Measuring response times on the 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; $F(2, 2005) = 2.20, p = .11, \eta^2 = .002$. 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 ($LOG_{10}RT = a + b_1(Layout) + b_2(Semantic\ Relatedness) + b_3(Layout * Semantic\ Relatedness)$) 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 5.

Table 5 Results of multiple regression model (LOG10RT) = a + b1(Layout) + b2(Semantic Relatedness) + b3(Layout*Semantic Relatedness)

	<i>df</i>	<i>F</i>	<i>p</i>
Semantic relatedness (RU_A)	4	.864	.49
Layout (Similarity, Dissimilarity, Random)	2	2.039	.13
RU_A * Layout	8	2.485	.01

The interaction between layout and semantic relatedness shown in Table 5 occasioned additional regression analyses to determine the effect of semantic relatedness (shown as RU_A in the x axis label in Figure 7) on participants' response times (shown as LOG10 of Response Time on the y axis label in Figure 7) on the search task. As Figure 7 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.

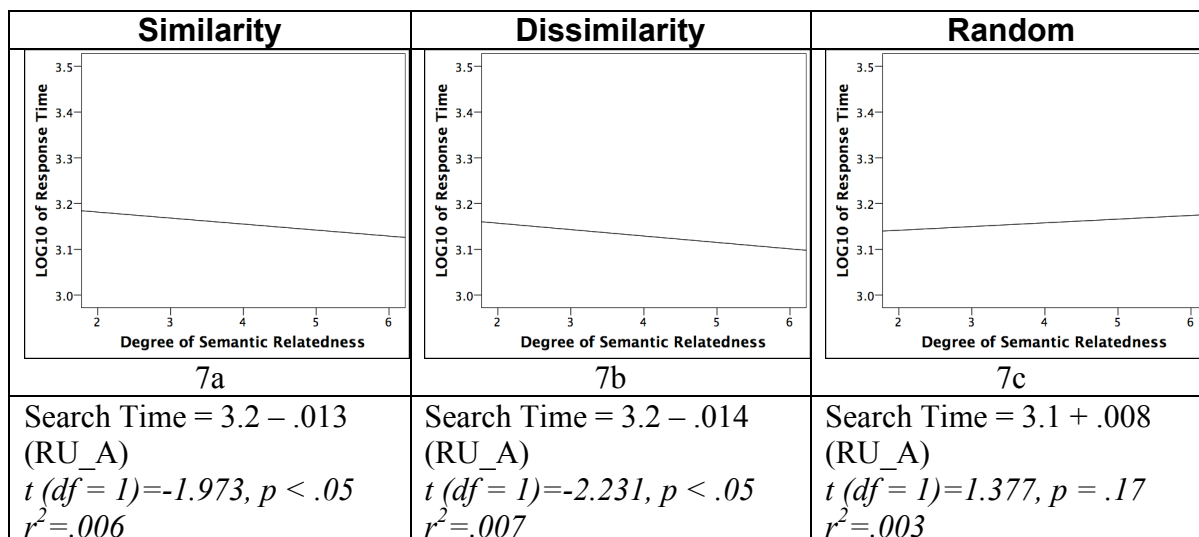


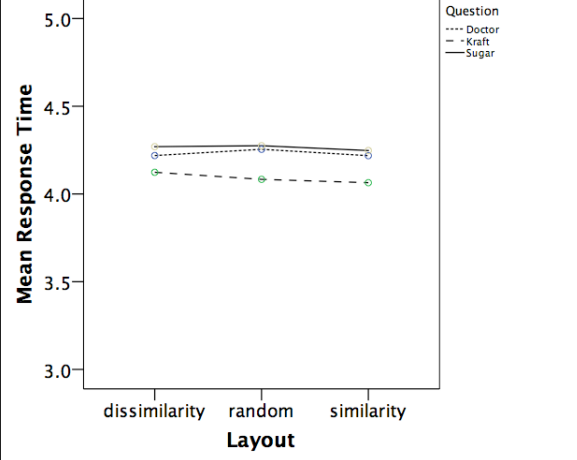
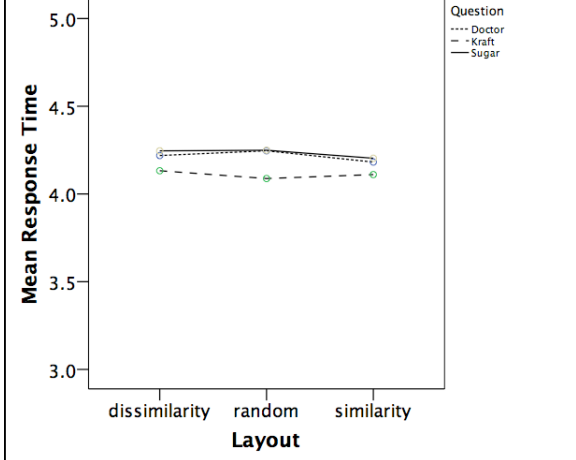
Figure 7 Search times (LOG10RT) as a function of semantic relatedness (RU_A). 7a. Similarity condition—elements with high semantic relatedness (higher RU_A) were depicted closer together than elements with low semantic relatedness. 7b. Dissimilarity condition—elements with high semantic relatedness were 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.

Measuring response times on the 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. Also, a 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, $F(4, 255) = .288, p = .89$.

All answers	Correct answers only
	
<p><i>Figure 8</i> Results of ANOVA revealing no significant differences between layouts in response times on the integration tasks, $F(2, 255) = .386, p = .68$ and no significant interaction between layout and question $F(4, 255) = .197, p = .94$</p>	<p><i>Figure 9</i> Results of ANOVA using correct answers only. There were no significant differences between layouts in response times on the integration tasks, $F(2, 255) = .478, p = .62$. and no significant interaction between layout and question, $F(4, 255) = .288, p = .89$</p>

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. This pattern of results produced the significant interaction between layout and semantic distance. In contrast, results for the integration tasks revealed

no significant differences between layouts in response times on the integration tasks and no significant interaction between layout and question.

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. As envisioned by the researchers, H1 can be summarized by the following relationship: Mapped Layout > Random layout > Reversed layout. In this relationship, a “mapped” layout is defined as a layout in which the placement of elements of elements on the layout is mapped to the degree to which those elements are semantically related. According to H1, a mapped layout on which highly related elements were placed close together as in the Similarity condition would be the optimal layout for performing search tasks rapidly. Thus, the Similarity condition should produce *decreasing* search times as a function of increasing semantic relatedness, and data indicate that this is indeed the case. Conversely, a layout with elements arranged in *reverse* of the optimal arrangement—that is, with highly related elements placed far apart, as in the Dissimilarity condition—should be the worst condition for performing search tasks rapidly.

At a minimum, H1 suggests that the Dissimilarity condition should produce *increasing* search times as a function of increasing semantic relatedness because elements with high semantic relatedness were placed farther apart on the layout. But as Figure 7 shows, the Dissimilarity condition (7b) also yielded significantly *decreasing* search times as

a function of semantic relatedness. It is not surprising that, in the Similarity condition, response time improved with increased semantic relatedness because, on the Similarity layout, semantic relatedness was mapped to physical distance. In other words, items that were semantically similar were placed closer together. If total response time consisted of eye movements to locate the targets and then physical movements to select them with the cursor, then it follows that shorter distances between targets would yield faster response times. But in the Dissimilarity condition, the layout was mapped in *reverse* of the Similarity condition, that is, items that were more semantically related were placed farther apart than items that were less semantically related. This meant that there were larger physical distances between semantically related items than in the Similarity condition. If response time was simply a function of physical distance, then the Dissimilarity condition should have produced increasing response times as semantic relatedness increased. However, the data show that this was not the case. Instead, response times improved in the Dissimilarity condition as semantic relatedness increased, just as they had in the Similarity condition.

One possible explanation for this is that the Similarity and Dissimilarity conditions provided an ordered structure, whereas the Random condition did not. Users can learn an ordered structure and consequently can develop strategies for the use of that structure. In the Similarity condition, items that were similar in meaning were consistently placed close together. Once a participant grasped this structure, he/she could shorten search times by starting the search at the orientation item and moving outward. In the case of the Dissimilarity layout however, items were arranged so that items that were similar in meaning were consistently placed far apart. Thus, a search strategy that involved starting the search

from the periphery of the layout and working inward would have produced faster search times. In both cases, the ordered structures make possible the development of such strategies.

By contrast, the Random layout does not lend itself to the development of a strategy at all. Since the placement of the items in the layout is not structured, each search item is no more likely or less likely to be near the orientation item, so no pattern can be recognized and therefore no strategy developed. Whether the development of such strategies actually takes place could be ascertained by comparing a participant's performance on search tasks that occurred earlier in the block of 30 trials with their performance on tasks that occurred later in the block of trials. An improvement in search times on trials that occurred later in the block would indicate the development and employment of a successful strategy. Unfortunately, such an analysis could not be performed in the current study because the decision was made in advance to randomize the order of the search tasks, and the order in which the search tasks were presented was not recorded. Future research should include such an analysis.

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 mated 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 correspondence 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. Even when S-R relationships are arranged in reverse of the optimal arrangement, that is, arranged in a structure that shares the least possible agreement with population stereotypes but is nonetheless orderly, performance is superior to that of conditions in which S-R relationships have been 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 current 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 in the 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: $+1.00 > 0.00 > -1.00$. As predicted, results showed that performance was uniformly better when lights and keys were in direct correspondence. However, just as in the Fitts and Deininger study and the present research, Morin and Grant found that “... a complete reversal of display-control relationships ($\tau = -1.00$) yields better performance than do the more irregular rearrangements ($-.86 \leq \tau \leq +.86$)” leading them to conclude that a subject “responds readily to orderliness, direct or reversed (high positive or negative values of τ)” (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. If the intercept for response times on the integration tasks consisted of the simple arithmetic addition of the search components for that task, then, by now, these search components might have been very fast. It is therefore not surprising that the effect of Layout was lost in the integration tasks. This could be remedied, going forward, with the use of different layouts for the integration tasks. In order to use the same elements that were used in the search task, new layouts could be created depicting a different product from the same product category. For example, the package of a different cough drop brand (e.g., Luden's) could be used with the same elements arranged identically to the Similarity, Dissimilarity, and Random layouts. This would eliminate the possibility that the effect of Layout would be lost in practice effects. Another way possible practice effects could be eliminated is by counterbalancing the order of the tasks. In the current research, all participants completed the search tasks before performing the integration tasks. In future studies, half of the participants might be asked to perform the integration tasks first. Alternatively, practice effects could also be eliminated by the introduction of additional dimensions to the integration tasks. All the tasks in the present research were performed on 2-D layouts. Future research in which search and integration

tasks are performed on 3-D layouts (boxes for example) might produce a measurable effect of Layout.

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. For example, the results of this analysis yielded a simple tool for evaluating an existing label for the extent to which its elements are arranged by semantic relatedness. By means of a Pearson correlation, physical distances between elements on the original product package that was used as the model for the layouts in this research were compared to the physical distances between elements on the Similarity, Dissimilarity, and Random layouts. Results of the correlation shown in Table 6 indicate that the elements on original label were not arranged according to an ordered mapping by semantic relatedness.

Table 6 Results of Pearson correlations comparing the physical distances of elements on the original product label with physical distances on the Similarity, Dissimilarity, and Random layouts

		Similarity	Dissimilarity	Random
Original product label	<i>r</i>	-.03	.14	-.34
	<i>p</i>	.89	.51	.10

Less obviously, the present research might provide guidance for labeling other kinds of consumer packaging, e.g., household chemicals. In some cases, this could affect consumer health and safety. Better search performance can be critical for persons who have

negative reactions to certain ingredients. Labels designed to facilitate user search tasks could potentially help people find the ingredients list on food labels and assist them in making better choices for their health and safety. Similarly, labels designed to facilitate rapid identification of hazardous chemicals could help prevent injury or death. Future research might determine additional variables that may facilitate people's visual search for components on labels, and could potentially serve as a basis for manufacturers' labeling decisions, as well as revisions to food labels regulations and guidelines. The results of the current research could be applied to existing product labels

REFERENCES

- Bettman, J.R., Payne, J.W., & Staelin, R. (1986). Cognitive considerations in designing effective labels for presenting risk information. *Journal of Marketing and Public Policy*, 5, 1-28.
- Card, S.K. (1982). User perceptual mechanisms in the search of computer command menus. *Proceedings of Human Factors in Computing Systems Conference (CHI '82)*, pp.190-196.
- Coren, S., Ward, L., & Enns, J. (2004). *Sensation and Perception*, Sixth Edition. Hoboken, NJ: John Wiley & Sons, Inc.
- Fitts, P. and Deininger, R. (1954). S-R Compatibility: correspondence among paired elements within stimulus and response codes. *Journal of Experimental Psychology*, 48(6), Dec 1954, 483-492.)
- Kosslyn, Stephen M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, 3, 185-226.
- Kruskal, Joseph B. & Wish, Myron (1978). *Multidimensional Scaling*, Fifth Printing. Beverly Hills, CA: Sage Publications, Inc.
- McDonald, J. E., Stone, J. D., Liebelt, L. S. & Karat, J. (1982). Evaluating a method of structuring the user-system interface. *Proceedings of the 26th Annual Meeting of the Human Factors Society*.

- McDonald, J. E., Stone, J. D., & Liebelt, L. S. (1983). Searching for items in menus: the effects of organization and type of target. *Proceedings of the 27th Annual Meeting Of the Human Factors Society*, pp. 834-837.
- McDonald, J. E., Dayton, Tom, & McDonald, D. R. (1988). Adapting menu layout to tasks. *International Journal of Man Machine Studies*, 28, pp. 417-435.
- Morin, R. E., & Grant, D. A. (1955). Learning and performance on a key-pressing task as a function of the degree of spatial stimulus-response correspondence. *Journal of Experimental Psychology*, 49(1), 39-47.
- Prinzmetal, W., Henderson, D., & Ivry, R. (1995). Loosening the constraints on illusory conjunctions: Assessing the roles of exposure duration and attention. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 1362-1375.
- Treisman, A.M. (1986) Features and objects in visual processing. *Scientific American*, 255, 114-125.
- Treisman, A.M. & Gelade, G. (1980) A feature integration theory of attention. *Cognitive Psychology*, 12, 97-136.
- Tullis, T. S. (1985). Designing a menu-based interface to an operating system. *Proceedings of Human Factors in Computing Systems Conference (CHI '85)*, pp. 79-84.
- Tullis, T. S., Bied, B. R. (1988). *Space Station Functional Relationships Analysis Final Technical Report*, Huntington Beach, CA: McDonnell-Douglas Astronautics Company, prepared for Ames Research Center under contract NAS2-11723.