The Multi-Items Rearrangement Task: a Faster and Reliable Method for Acquiring Similarity Matrix

Hsuan-Yu Lina

Alexei Fischerb

Klaus Oberauera

University of Zuricha, Center for Adaptive Security Research and Applications (CASRA)b

Author Note

[Include any grant/funding information and a complete correspondence address.]

Abstract

[The abstract should be one paragraph of between 150 and 250 words. It is not indented. Section titles, such as the word Abstract above, are not considered headings so they don’t use bold heading format. Instead, use the Section Title style. This style automatically starts your section on a new page, so you don’t have to add page breaks. Note that all of the styles for this template are available on the Home tab of the ribbon, in the Styles gallery.]

Keywords: [Click here to add keywords.]

The Multi-Items Rearrangement Task: a Faster and Reliable Method for Acquiring Similarity Matrix

Similarity between items is an important aspect in many area of psychology, e.g., categorization (Nosofsky & Palmeri, 1997), memory (Farrell, 2006; Jackson, Linden, Roberts, Kriegeskorte, & Haenschel, 2015; Nosofsky & Kantner, 2006), and reasoning (Heit & Rubinstein, 1994). The typical task for acquiring similarity matrix is the Paired-Comparison task (Giordano et al., 2011), which askes participants to rate the similarity of two items at once, thus the number of trials requires to compare items is

|  |  |  |
| --- | --- | --- |
|  |  |  |

Because the number of necessary comparisons to complete a full similarity matrix increase exponentially, acquiring similarity matrix is often impractical when the number of items reaches certain number. For example, the similarity matrix of 50 items requires 1225 trials, which will require two hours to complete the task (Giordano et al., 2011).

In this study, we present a new method for acquiring the similarity matrix quickly and accurately, Multi-Items Rearrangement task. The task presents multiple items at once, and participants are asked to rate those items in the single trial, which reduced the number of trials required for completing the similarity matrix drastically. In this article, we first introduce the Multi-Items Rearrangement task in detail, then we will introduce two experiments to test the reliability and validity of the Multi-Items Rearrangement task.

# Multi-Items Rearrangement Task

In the Multi-Items Rearrangement Task, multiple items are presented at once and participants are instructed to rearrange the items based on their similarity to the others, whereby the proximity between the items serves as an indicator of their similarity. Items which are closer together are regarded as similar, while items which far apart are regarded as dissimilar.

Figure 1 shows the procedure of a trial in the Multi-Items Rearrangement task. Because the Multi-Items Rearrangement task presents multiple items at once, a subset of the similarity matrix can be acquired in a single trial. Therefore, the Multi-Items Rearrangement task requires less trials to complete the full similarity matrix.

The Multi-Items Rearrangement Task divides the item pool into several subsets, and two subsets are presented in a trial, thus, the task only requires pair-wise combinations among subsets instead of pair-wise combinations among every single item. Assuming the item pool has items and is divided into subsets with items each, there are subsets. The number of trials required to completely compare all the subsets is

|  |  |  |
| --- | --- | --- |
|  |  |  |

In the case where each subset contains only item, the Multi-Items Rearrangement task requires the same number of trials as the Paired-Comparison task, since there are only two items are compared at once. If the subset contains half of the items, all the items are compared at once, and only a single trial is required to complete the similarity matrix. However, it is unrealistic to present all the in a trial. Presenting large amount of the items in a trial increases the complexity for rearrangement. In the study, we presented 8 items at once in a single trial, i.e., 4 items per subset.

Other than time efficiency, the Multi-Items Rearrangement task also provides a finer scale for reporting similarity comparing to Paired-Comparison task. In the Paired-Comparison task, participants are limited to a discrete ordinal rank scale with a narrow value range (normally between 5 and 9 points). Many pairs will fall into the same similarity rating despite there being subtle differences between pairs. In the Multi-Items Rearrangement task, the similarity between items is reported through the items’ distance to one another. The scale’s granularity is limited only by the resolution of the screen. Hence, participants are able to reflect their objective similarity more precisely.

Another advantage of the Multi-Items Rearrangement task over the Paired-Comparison task is that the Multi-Items Rearrangement task is affected less by the diagnosticity effect (Tversky, 1977) because multiple items are presented at once. Previous studies found that the item set affects the perception of the items, which then affects the similarity rating (Goldstone, 1995). For example, the similarity rating between green and blue is more dissimilar if the pair is presented alone then if the pair is presented along with red. In the Paired-Comparison task, participants are gradually exposed to the item set, hence the standard of the similarity rating changes throughout the task. In the Multi-Items Comparison task, participants are exposed to multiple items at once, thus providing an overview of the item pool and contextual information among items.

Although the Multi-Items Rearrangement task has many advantage of acquiring similarity matrix, the reliability and the validity of the task is not yet explored. Thus, we will introduce two experiments which examine the reliability and the validity on different material. Discrete features are commonly used in the psychology experiment (Allen, Baddeley, & Hitch, 2014; Luck & Vogel, 1997). Therefore, in the Experiment 1, we used the material constructed from multiple discrete features. In the Experiment 2, we used the material from a continuous dimension with objective similarity, which was commonly used in both categorization and visual working memory studies.

# Experiment 1

In the Experiment 1, we examined the reliability and the validity of the Multi-Items Rearrangement task with materials constructed from discrete features. The experiment is separated into two blocks. The first block employed the Multi-Items Rearrangement task to measure the similarity between abstract faces. The second block used the Paired-Comparison task to validate the result acquired from the Multi-Items Rearrangement task.

## Method

Participants. Ten students recruited from University of Zürich. Participants were rewarded with course credits or 30 Swiss Francs upon experiment completion.

Materials. Both Multi-Items Rearrangement task and the Paired-Comparison task shared the same set of stimuli. Color patches are used in the practice trials. The colors are randomly selected from all the possible colors in the 24 bits RGB color space. Abstract faces are used in the experiment trials. The faces are varied on four dimensions: the width between eyes, the height of eyes, the length of nose, and the position of mouth, with each dimension has two possible configurations. The faces are shown in the Figure 2.

Procedure. Experiment 1 consisted of two identical sessions, and the sessions were carried in two different days. Each session contained two blocks. The first block is the Multi-Items Rearrangement task, and the second block is the Paired-Comparison task. On average, each session takes about 45 minutes.

Multi-Items Rearrangement Task. The Multi-Items Rearrangement task consists of two practice trials and 12 experiment trials. The items were randomly assigned into 4 subsets with 4 items each. In each trials, two subsets of items were randomly scattered on the screen without overlapping. Participants were instructed to rearrange the items by using mouse to drag-and-drop the items, and the distance between items should reflect the similarity between the items. The farther the distance between items, the less similar they are perceived to be. After participants were satisfied with the arrangement of the items, they can press space bar to continue to next trial. Participants were instruction to take as long as they want to rearrange the items.

The Multi-Items Rearrangement task requires 6 trials to produce a complete similarity matrix between 16 items. We repeated the procedure twice in order to obtain a more accurate measurement for the similarity matrix. The items were rearranged into different subsets for the second repetition.

Paired-Comparison Task. The paired comparison task consists of 4 practice trials and 240 experiment trials. In each trial, two items were presented on the screen with a 9 points scale below the items. Participants were instructed to rate the similarity between the two items by clicking on the 9 points scale, with 1 to be the most similar, and 9 to be the most dissimilar. After the similarity value is selected, a blank screen appeared for 1 second and was followed by the next trial. Participants were instructed to take as long as they wanted to complete the trial. There were 10 evenly distributed breaks in the Paired-Comparison task session, participants were encouraged to take as long as they want in the break. Participants were instructed to press the space bar to continue the task after they finished the break.

The Paired-Comparison task requires 120 trials to complete a similarity matrix for 16 items. We repeated the measurement twice in order to increase the accuracy of the similarity matrix.

## Results

The similarity matrix acquired from the Multi-Items Rearrangement task is based on the Euclidian distance between items in the trial. If the distance between two items were measured multiple times, the average of the distance among measurements is used as the similarity between the items. The similarity matrix acquired from the Paired-Comparison task is based on the rated similarity between items. Similar to the Multi-Items Rearrangement task, if an items pair is rated multiple times, the similarity between the items pair is calculated as the average between ratings. The similarity matrices acquired from both tasks were normalized by rescaling the maximum dissimilarity in the matrix to 1. The normalization ensures that the similarity matrix acquired from the Multi-Items Rearrangement task and the similarity matrix acquired from the Paired-Comparison task are under the same scale.

To test the reliability of Multi-Items Rearrangement task and the Paired-Comparison task, we compared the similarity matrices acquired from the first session with the second session for each participant. The comparison between the similarity matrices is done through Random Skewers method (Cheverud & Marroig, 2007). The correlation between the similarity matrices acquired from first session and the second session for both tasks of each participant are listed in Table 1. The lowest reliability is 0.71 for the Multi-Items Rearrangement task, and 0.78 for the Paired-Comparison task. To test the validity of the Multi-Items Rearrangement task, the similarity matrices acquired from Multi-Items Rearrangement task and the similarity matrices acquired from Paired-Comparison task were again compared with the Random Skewers method. The correlation between the similarity matrices are listed in Table 1, where the lowest correlation is 0.85. To ensure both similarity matrices are aligned, we plotted the acquired similarity matrices with Multidimensional Scaling, as shown in Figure 3.

The average time required for task completion was 475.6 seconds for the Multi-Item Rearrangement task, and 905.1 seconds for the Paired-Comparison task. In both cases the break time was excluded. The time required for both tasks were compared in R (R. Core Team, 2016) with BayesFactor package (Morey & Rouder, 2015), and the data strongly supported that Paired-Comparison task takes longer than Multi-Items Rearrangement task ().

# Experiment 2

In Experiment 2, we replicated the same method applied in Experiment 1 using color patches as material in order to examine the reliability and validity of the Multi-Items Rearrangement task for continuous features.

### Method

Participants. Ten students recruited from University of Zürich. Participants were rewarded with course credits or 30 Swiss Francs upon experiment completion. All participants in Experiment 2 had not participated in the previously conducted Experiment 1.

Materials. Both Multi-Items Rearrangement task and the Paired-Comparison task shared the same set of stimuli. The faces from the Experiment 1 were used in the practice trials. Color patches were used in the experiment trials. 16 color patches were selected from a color wheel which was created in the CIE L\*a\*b\* color space with radius of 60 and centered at luminance set to 70, set to 20, and set to 38. The luminance was hold constant while and were allowed to vary. All the color patches were evenly distributed on the color wheel. The color patches are shown in the Figure 4, and the RGB values of the color patches are shown in Table 2.

Procedure. The procedure of Experiment 2 was identical to that of Experiment 1.

## Results

The similarity matrices acquired from both tasks were normalized in the same way as in Experiment 1. For the reliability test for both tasks, we again applied the Random Skewers method to compute the coefficient between the similarity matrices acquired from the first and the second session. The correlations between the similarity matrices are shown in the Table 3. The lowest attained reliability coefficient was 0.79 for the Multi-Items Rearrangement task and 0.70 for the Paired-Comparison task.

To test the validity of the Multi-Items Rearrangement task, we compared the average similarity metric acquired from both sessions of the Multi-Items Rearrangement task with the average similarity metric acquired from both sessions of the Paired-Comparison task using the Random Skewers method. The correlations are shown in Table 3, the lowest validity measure being 0.77. The similarity matrices acquired from both tasks were plotted MDS in Figure 5.

The average time required for completion was 338.3 seconds for the Multi-Item Rearrangement task, and 713.9 seconds for the Paired-Comparison task. The time required to complete both tasks were compared with BayesFactor package in R. The results have shown strong evidence for the time required to complete the Multi-Items Rearrangement task being shorter than the Paired-Comparison task ().

# Conclusion

The experiments results have shown that the Multi-Items Rearrangement task is on par with the Paired-Comparison task regarding reliability. The lowest reliability measure in both experiments is 0.70 from the Paired-Comparison task in Experiment 2 and 0.71 for the Multi-Items Rearrangement task in Experiment 1. Even in the worst case, the reliabilities from both tasks are highly reliable, hence, we concluded there is no reliability issue in the Multi-Items Rearrangement task.

The lowest validity coefficient from both experiments was 0.77 for Experiment 2, which was produced by a participant with low reliability value in the Paired-Comparison task. Low within-task reliability is an indicator of higher variance among data points of the produced similarity matrices, which in turn has a negative effect on the between-task correlation constituting the validity measure. Still, the Multi-Items Rearrangement task is highly reliable even without considering the reduction of the validity from the low reliability of the Paired-Comparison task. The result from Multidimensional Scaling has also shown that both tasks yielded almost identical spatial organization. We thus conclude that there is no validity issue in the Multi-Items Rearrangement task.

Furthermore, participants completed the Multi-Items Rearrangement task in approximately half the time it took them to complete the Paired-Comparison task.

# General Discussion

In this study, we introduced a new task for acquiring similarity matrix — the Multi-Items Rearrangement task — and two experiments to examine the reliability, the validity, and the time efficiency of the task. The experiments shown that the Multi-Items Rearrangement task is reliable and valid method of acquiring similarity matrices while taking only half of the time required by the Paired-Comparison task. The time efficiency of the Multi-Items Rearrangement task allows the experimenter to acquire the similarity matrices for individual participants when previously unfeasible, for example, when the item set is too large. Some studies applied the strategy of acquiring only a subset of the complete similarity matrix from each individual participant, and then assembling the yielded subsets into a full similarity matrix (Boles & Clifford, 1989). However, the assembled similarity is potentially polluted by inter-individual differences in similarity ratings between participants. This is especially problematic for stimuli which produce large inter-individual differences in similarity rating.

In a previous study, 16 items were presented in each trials in the Multi-Items Arrangement task with the item pool of 48 items. The Multi-Items Arrangement session required approximately 15 minutes to complete. If only 8 items were presented in a Multi-Items Arrangement trial, the measurement would require approximately 31 minutes to complete using the average duration acquired from Experiment 2. If the measurement was done with the Paired-Comparison task, using the average duration of the Paired-Comparison trial acquired from Experiment 2, the measurement would take approximately 56 minutes. The estimate time to complete the measurement with different number of items presented in a trial under different number of items in the item pool is shown in Figure 5.

The Multi-Items Rearrangement task does not come without disadvantages. The relationship between three items can be perfectly represented on two-dimensional space regardless the dimensionality of the representation. However, with four or more items, the relationship between items cannot always be represented on a two dimensional without distortion. Because the task prompts participants to reflect the similarity between items on a two dimensional plane, the measurement might miss some complex relationship between items. For example, if the item set is represented in the higher dimensional space, there must be some distortion when reflecting the items on the two-dimensional space. The Pair-Comparison task does not share the same problem because the Pair-Comparison task only compare two items at a time, which the relationship can be perfectly reflected on a one-dimensional scale.

A strategy which could be employed to reduce distortion would be to increase the number of times an item in the item pool is presented in combination with every other item throughout the task trials. Every time an item pair is present on a trial, a distance metric can be yielded for that item pair. With multiple distance metrics for the same pair, the average distance could be calculated in order to counteract the effects of distortion due to dimensional reduction. The method for constructing Multi-Items Rearrangement task trials presented in this paper, i.e. dividing the item pool into subgroups of equal size and then distributing all possible 2-subgroup combinations among trials, might ensure that every item is compared with every other item at least once, but it does not distribute the number of comparisons in a homogenous form. Some item pairs are presented multiple times while others are presented only once. By evenly distributing the number of trials presenting a particular item pair redundant distance metrics could be yielded, from which the average distance could be taken.

Although the Multi-Items Rearrangement task might cause distortion for the relationship between items because of the response scheme. The results from Experiment 1 shown that even when the items were constructed from four feature dimensions, the similarity matrix acquired through the Multi-Items Rearrangement task was strikingly similar to the distortion-free matrix acquired through the Paired-Comparison task. Hence, we believe the distortion caused by dimensional reduction in the Multi-Items Rearrangement task is likely negligible.

References

Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2014). Evidence for two attentional components in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1499–1509. https://doi.org/10.1037/xlm0000002

Boles, D. B., & Clifford, J. E. (1989). An upper- and lowercase alphabetic similarity matrix, with derived generation similarity values. *Behavior Research Methods, Instruments, & Computers*, *21*(6), 579–586. https://doi.org/10.3758/BF03210580

Cheverud, J. M., & Marroig, G. (2007). Research Article Comparing covariance matrices: random skewers method compared to the common principal components model. *Genetics and Molecular Biology*, *30*(2), 461–469. https://doi.org/10.1590/S1415-47572007000300027

Evers, E. R. K., & Lakens, D. (2014). Revisiting Tversky’s diagnosticity principle. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00875

Farrell, S. (2006). Mixed-list phonological similarity effects in delayed serial recall. *Journal of Memory and Language*, *55*(4), 587–600. https://doi.org/10.1016/j.jml.2006.06.002

Giordano, B. L., Guastavino, C., Murphy, E., Ogg, M., Smith, B. K., & McAdams, S. (2011). Comparison of Methods for Collecting and Modeling Dissimilarity Data: Applications to Complex Sound Stimuli. *Multivariate Behavioral Research*, *46*(5), 779–811. https://doi.org/10.1080/00273171.2011.606748

Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychological Science*, *6*(5), 298–304.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(2), 411.

Jackson, M. C., Linden, D. E. J., Roberts, M. V., Kriegeskorte, N., & Haenschel, C. (2015). Similarity, not complexity, determines visual working memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1884–1892. https://doi.org/10.1037/xlm0000125

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: omputation of Bayes Factors for Common Designs (Version R package version 0.9.12-2). Retrieved from https://CRAN.R-project.org/package=BayesFactor

Nosofsky, R. M., & Kantner, J. (2006). Exemplar similarity, study list homogeneity, and short-term perceptual recognition. *Memory & Cognition*, *34*(1), 112–124. https://doi.org/10.3758/BF03193391

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266–300. https://doi.org/10.1037/0033-295X.104.2.266

R. Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from http://www.R-project.org/

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352. https://doi.org/10.1037/0033-295X.84.4.327

Tables

Table 1

Reliability and Validity of Experiment 1.

|  |  |  |  |
| --- | --- | --- | --- |
| Participant | Paired-Comparison | Multi-Items Rearrangement | Validity |
| 1 | 0.98 | 0.91 | 0.95 |
| 2 | 0.87 | 0.84 | 0.91 |
| 3 | 0.78 | 0.80 | 0.91 |
| 4 | 0.79 | 0.81 | 0.85 |
| 5 | 0.87 | 0.96 | 0.87 |
| 6 | 0.81 | 0.74 | 0.91 |
| 7 | 0.94 | 0.83 | 0.87 |
| 8 | 0.95 | 0.86 | 0.86 |
| 9 | 0.91 | 0.95 | 0.91 |
| 10 | 0.85 | 0.71 | 0.89 |

Table 2

The RGB values of the color patches used in Experiment 2.

|  |  |  |  |
| --- | --- | --- | --- |
| Item | R | G | B |
| 1 | 255 | 90 | 109 |
| 2 | 255 | 97 | 65 |
| 3 | 255 | 116 | 0 |
| 4 | 255 | 137 | 0 |
| 5 | 238 | 156 | 0 |
| 6 | 204 | 171 | 0 |
| 7 | 170 | 182 | 0 |
| 8 | 141 | 188 | 49 |
| 9 | 118 | 191 | 101 |
| 10 | 107 | 190 | 145 |
| 11 | 117 | 180 | 180 |
| 12 | 149 | 176 | 204 |
| 13 | 191 | 162 | 213 |
| 14 | 234 | 144 | 206 |
| 15 | 255 | 122 | 183 |
| 16 | 255 | 101 | 150 |

Table 3

Reliability and Validity of Experiment 2.

|  |  |  |  |
| --- | --- | --- | --- |
| Participant | Paired-Comparison | Multi-Items Rearrangement | Validity |
| 1 | 0.96 | 0.92 | 0.96 |
| 2 | 0.94 | 0.85 | 0.91 |
| 3 | 0.89 | 0.89 | 0.90 |
| 4 | 0.93 | 0.86 | 0.93 |
| 5 | 0.94 | 0.89 | 0.95 |
| 6 | 0.98 | 0.83 | 0.93 |
| 7 | 0.97 | 0.95 | 0.97 |
| 8 | 0.99 | 0.84 | 0.96 |
| 9 | 0.70 | 0.79 | 0.77 |
| 10 | 0.93 | 0.83 | 0.93 |

Figures

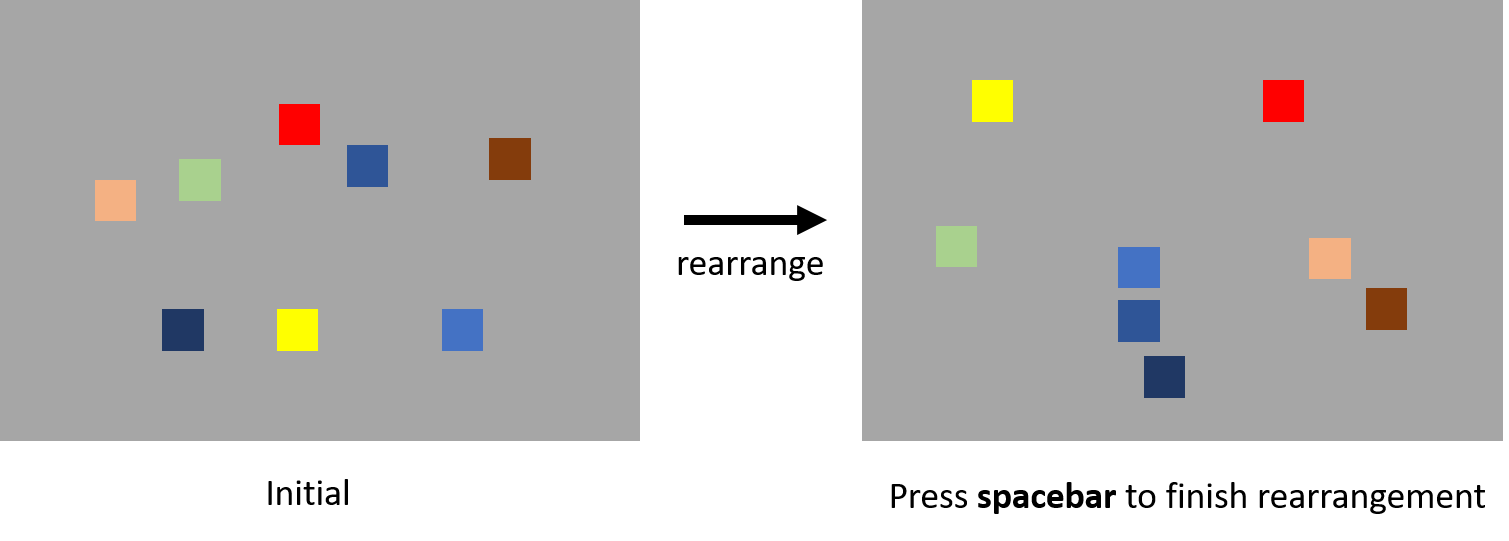


Figure 1. The procedure of the Multi-Items Rearrangement task. The left figure is the initial presentation of the items. Participants were asked to rearrange the items according to the similarity between the items by using mouse to click-and-drop. The right figure is a potential outcome after the rearrangement.



Figure 2. The material used in the Experiment 1. The faces are constructed with four dimensions: the width between eyes, the height of eyes, the length of nose, and the position of mouth.

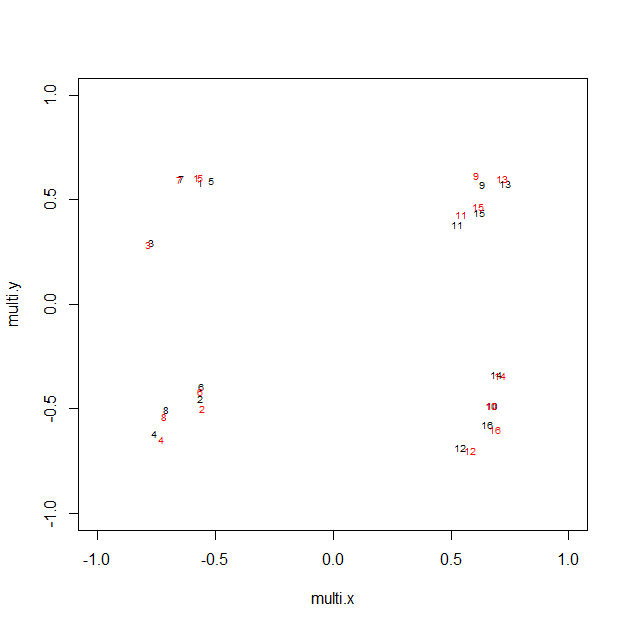


Figure 3. The MDS results of the similarity matrices acquired from the Multi-Items Rearrangement task and the Paired-Comparison task. The numbers indicate the items in Figure 2.



Figure 4. The material used in the Experiment 2. All the color patches are equality distributed on a color wheel which centers at set as 70, set as 20, and set as 38 with radius 60.

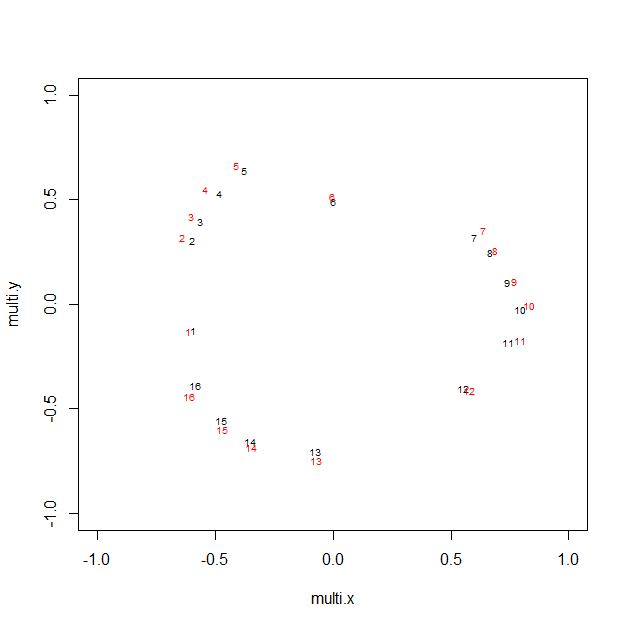


Figure 5. The MDS results of the similarity matrices acquired from the Multi-Items Rearrangement task and the Paired-Comparison task. The numbers indicate the items in Figure 4.

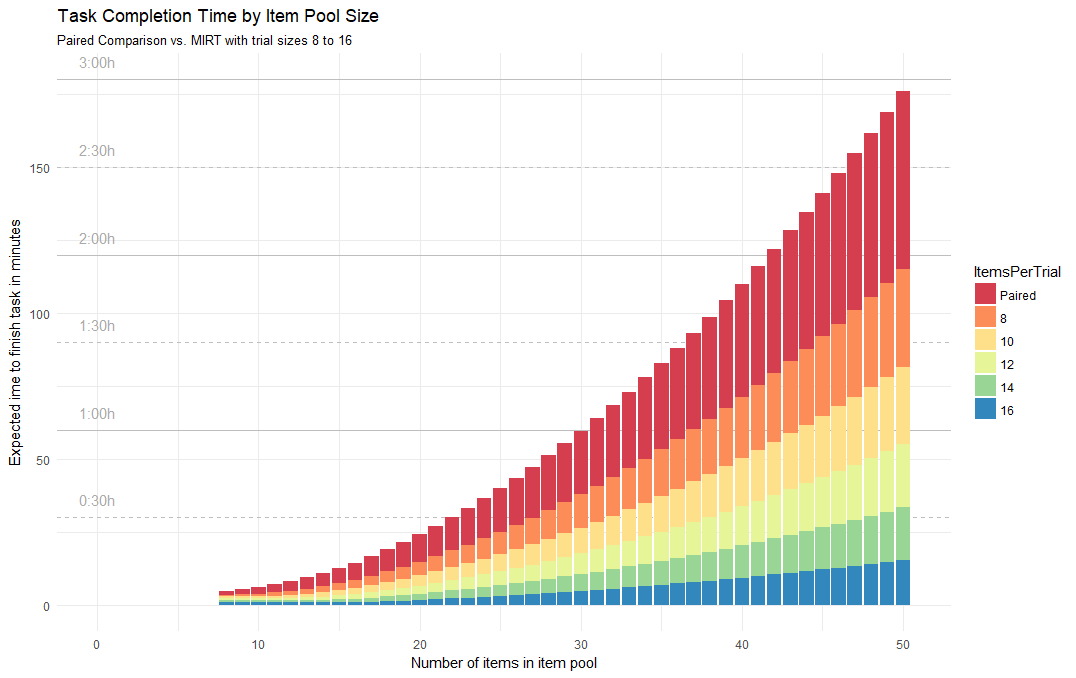


Figure 5. The estimated time for completing the similarity matric of different numbers of items in the item pool with different number of items presented in a trial. The red bar indicates the estimated time required for completing the Paired-Comparison task. The other bars represent the estimated time required for completing the Multi-Items Arrangement task with different number of items in a trial.