Are Images Crucial for Learning? Observation of a Picture Superiority Effect in Working Memory

Jeremias Lino Ferrao (j.l.ferrao@student.rug.nl)
Department of Artificial Intelligence, University of Groningen
9747 AG Groningen, The Netherlands

Abstract

In the present paper, we examined if the working memory capacity is greater for images rather than words. Previous research revealed the existence of a picture superiority effect which stated that images are easier to remember than words. However, we sought to test this effect ourselves and discuss our results considering specific literature. In our study, we conducted two free recall tests in which the items presented were images or words. In these tests, we measured and analyzed the number of correctly recalled items. Our results revealed that significantly more images were recalled than words. Therefore, we validate the picture superiority effect. We can also conclude that the working memory capacity for images is greater than words.

Introduction

The eminence of vision over all other senses in the brain is doubtless. Hutmacher (2019) supports this claim by remarking that the bulk of printed research focuses on sight in contrast to other senses. Thus, it is logical to deduce that the memory for recalling short-term events (Henceforth called Working Memory (WM)) would possess a greater capacity for images rather than words. This effect has been aptly named as the picture superiority effect (Paivio et al., 1968) and states that images are more likely to be remembered than words. However, as described by Purtill (1970), the core purpose of science is to dispute norms and build theories on previous research. Therefore, we sought to test this effect ourselves and analyze the results from our study.

The inclination of the brain towards recalling visual data has great application in learning. As mentioned by Clark and Paivio (1991), it is easier to understand concepts when they are conveyed through pictures rather than words. Clark and Paivio (1991) also reveal that the learning rate of a student is proportional to the number of items that can be recalled at once. Thus, if more items are recalled when these items are represented by images, then it is pivotal that the picture superiority effect is well studied.

Earlier research on WM by Miller (1956) underlined that WM could only recall around 7 items. Likewise, Hayes (1952) proved that the amount of data in an item did not influence this fixed number of recalled items. A drawback of both studies is that the material comprised items such as digits and words. It is unclear if items such as images would counter their claims. This is because images contain a greater amount of data than words (Fais et al., 1996).

Contrary to these papers, Hilton (2001) proposed that the brain takes less effort to recall visual data compared to auditory data. From his study, Hilton (2001) surmised that the brain uses special systems to recall visual data. If images and words are presented visually, will more items be recalled when these items are images rather than words? This is a question that arises from Hilton (2001).

In the present paper, we seek to extend the above studies and determine if the picture superiority effect is valid. Our research question investigates if more items are recalled when these items are images rather than words in a free recall test. Based on the importance of vision, as stated earlier in the paper, we believe that this is indeed the case. Thus, our hypothesis is that more items are recalled when these items are images rather than words in a free recall test.

To test our hypothesis, we will conduct two free recall tests for images and words. The material will either comprise a set of images or a set of words. First, we will present this material to the participants who we will instruct to recall as many items as possible. The type of item presented first will be randomized. After this presentation, the participant will have to list down the items recalled in any order. The participant then has to repeat the test for the second type of item. Finally, we will measure and analyze the number of correctly recalled items in both tests.

Methods

Participants

22 students from the University of Groningen (12 male, 10 female) voluntarily took part in the experiment. The mean age (in years) of the tested population was 20.1, and the range was 18 to 23. They were recruited from various social spaces on the Zernike Campus and were offered no reward for participating in the experiment. The only prerequisite was that the participant had to be fluent in English and possess no mental or learning disorders. The participant signed a consent form prior to the experiment. This form was approved by the Department of Artificial Intelligence.

Materials and Stimuli

We selected 28 words for the experiment, and each word was paired with a corresponding image. The word was obtained using the following constraints:

- The word had to be a noun from the top 10 thousand frequently used words. We used a data set by Davies and Kim (2019) to get this information. This was necessary to ensure that the participant would be familiar with the word and memorize it.
- The word needed to possess an imageability value of 500 to 700 from the MRC Psycholinguistic Database (Wilson,

1988). Imageability refers to the likelihood that reading a word would evoke a mental image. A high imageability value ensures that a concrete image exists for the word.

- The word must only comprise one or two syllables. This constraint was added to ensure that the participant could quickly rehearse the words as they appeared.

The image for each of the words was obtained from the top search results in Google Images using the keyword "{Word} Clip art". Such an example of a keyword is "Apple Clip art". The image solely needed to depict the word itself with no other item in the background.

In the experiment, we divided these 28 word-image pairs into two sets of 14. One set comprised solely images, whereas the other set comprised words only. This division was randomized using the random library in the Python programming language (Van Rossum, 2020). Because of randomization, the words and images presented were different for each participant. The words were displayed with a font size of 128 in the font Calibri and the images were depicted at a constant size of 500 x 500 pixels. These items were viewed at a distance of about 50 cm from a 15.6-inch computer screen.

Procedure

Participants were tested individually in a quiet room on the Zernike Campus. A supervisor was present to ensure the smooth flow of the experiment. The type of item first presented to the participant was randomized using Python. For example, if the participant was presented with words in the first test, then they would be presented with images in the second test. The entire experiment was conducted using a custom Python program on a laptop with a keyboard to input data.

After exchanging greetings with the supervisor, the participant signed a consent form and inserted their age and gender into the program when prompted by it. The program then displayed instructions that informed the participant about the details of the experiment. More precisely, the participants were informed that they would either be presented with a sequence of images or words and that they had to recall as many of these items as possible. The order in which they recalled the items did not matter.

The test began once the participant confirmed they understood the instructions. 14 randomly arranged items were instantaneously presented one after the other, with each item appearing on the screen for 3 seconds. This time interval of 3 seconds was necessary to guarantee that the participant could recognize and remember the items presented to them. We arrived at this number through manual testing. These items were presented in the centre of the screen as per the specifications in the "Materials and Stimuli" subsection.

Once the presentation was complete, the participant had to recall the items in any order and input them into the program. A break of 10 seconds was provided to the participant before restarting the process for the second item type. We arrived at this number after we observed participants wanted to begin the second test immediately. During the break, the participant was instructed by the program to remain seated and focus on the computer screen. Once 10 seconds passed (as displayed in a timer by the program), the participant had to press a button to start the second test. This manual restart was necessary to ensure that the participant was paying attention to the screen so that they could memorize the next set of items. Finally, when the second test was completed, the experiment concluded, and the participant left the room.

Design

- **Independent Variable (IV):** The IV manipulated in our experiment is the type of item that is shown to the participant. This IV comprises two conditions, which are whether images or words are presented.

- **Conditions:** We used a within-subject design to get meaningful data from a small sample size. Thus, each participant had to perform the test for both conditions. The first condition assigned to a participant was randomized to negate any effect that could carry over between tests. Thus, 11 participants did the experiment with the images first. Another detail to minimize carry-over effects is that the items presented in the second test did not include any of the items in the first test. Besides this, the items presented, along with the order in which the items appeared, were randomized in a test. This randomization was done by the program using Python.

- **Dependent Variable (DV):** The DV measured is the number of correctly recalled items. The program calculated this value by comparing the participant's input to the list of items presented in the test. After the experiment, this data was then inspected to account for the usage of synonyms and differing spellings. For instance, if an aeroplane was depicted, and the participant entered "airplane" then we would count it as correctly recalled.

Results

We used the data of 21 participants and calculated the mean number of images and words recalled. The data of an additional participant was completely excluded because of a disturbance in the testing environment.

Since our hypothesis states that more items are recalled when these items are images rather than words in a free recall test, we compared the mean number of images and words recalled. We used the mean as a measure of central tendency instead of the median since our data did not possess a skew towards any direction. We compiled our results in a bar plot, as illustrated in Figure 1.

Figure 1 depicts the mean number of items recalled in the two conditions along with their standard error. We chose the standard error as a measure of deviation, since we wanted to illustrate the average difference in the number of items recalled in a condition.

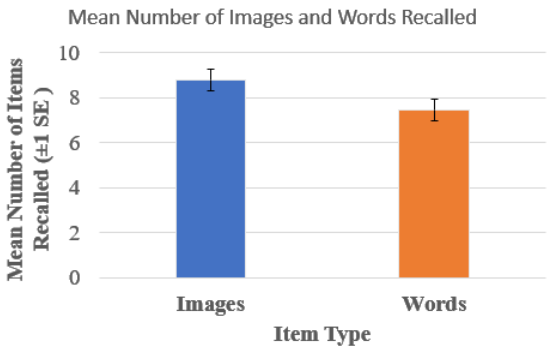

Figure 1: This graph represents the mean number of items recalled by participants when these items were images or words. The blue and orange bars show the mean number of images and words recalled, respectively. The Standard Error (SE) is also provided to show the average difference in the number of items recalled for the two item types

.

Our results reveal that, on average, around 1.4 more items were remembered when these items were represented as images rather than words. Hence, the mean number of images recalled was greater than the mean number of words recalled.

To further determine if the difference between the two groups was significant, we used a one-tailed paired sample t-test to determine the likelihood of our results occurring by chance. We chose a one-tailed t-test to determine if it was likely that more images would be recalled than words. Besides this, we needed a paired sample t-test since we had two kinds of data points. These data points were the number of images and the number of words recalled by a participant.

To use a t-test, we need to calculate the degrees of freedom (df) which is obtained using the expression N - 1. Here, N stands for the sample size of the population. Since we possessed 21 participants, the df in our t-test is 20.

Finally, the t-test revealed that significantly more items were likely to be recalled when they were represented as images (M = 8.8, SD = 1.9) rather than words (M = 7.4, SD = 2.6), $t(20) = 1.9$, $p = .035$.

Because of our significant p-value, we reject the null hypothesis, which states that an equal number of items are recalled when these items are images or words. Instead, we propose that more items are recalled when these items are images rather than words in a free recall test.

Discussion

Our Research

In the present paper, our research question examined if more items would be recalled when these items are images rather than words in a free recall test. Our hypothesis, which was proven to be true, stated that more items are recalled when these items are images rather than words in a free recall test.

To test our hypothesis, we conducted free recall tests in which the items presented were images or words and measured the number of correctly recalled items. The recall tests were automated to ensure a standard procedure for all participants. The order in which the items were presented to the participants were also randomized to mitigate any carry-over effect.

Finally, the results revealed that the mean number of items remembered was greater when the items were images compared to words. We also performed a t-test, which revealed that there was a significant difference between the two groups. Therefore, this t-test validated our hypothesis.

Background Literature

Our study validates the picture superiority effect as observed by Paivio et al. (1968) which stated that images are more likely to be remembered than words. This is because more images were recalled than words in our study.

Additionally, our findings support Hilton (2001)'s claim that the brain possesses a specialized system for recalling visual data. In our study, the mean number of images recalled was greater than the mean number of words. This result draws parallels to Hilton (2001)'s findings that more items were recalled when they were represented through visuals rather than audio.

Finally, our results appear to contradict Hayes (1952) conclusion that the number of recalled items does not vary significantly depending on the amount of information in the item. Since the mean number of images recalled was significantly higher than the mean number of words recalled, our study seems to show that the amount of information in an item affects the number of recalled items. We can come to this conclusion since images contain a greater amount of information than words (Fais et al., 1996). A potential cause of our contradiction is that Hayes (1952) used words and digits in a free recall test, whereas we used images and words. Therefore, Hayes (1952) would not have observed the picture superiority effect in his study.

Bigger Picture

In the bigger picture, the picture superiority effect has great significance in learning as Clark and Paivio (1991) reveals that the learning rate of a student increases with the number of items that can be recalled at a time. Since our study reveals that more items are recalled when represented by images, we can deduce that the use of images would contribute towards a greater learning rate. Therefore, we suggest that educational material such as books and courses should use more images to convey concepts to students. Such a change would enable students to study concepts in a shorter period of time. This increase in the learning rate would also allow a student to study more concepts in a period. Clark and Paivio (1991) also reveal that images enable a student to easily understand concepts. Therefore, the use of more images would enable students to gain a deeper understanding of concepts.

Additionally, we can theorize that the brain possesses a specialized system for recalling visual data. The existence

of such a system would explain our findings that more images were recalled than words. Hilton (2001) also arrived at a similar theory when he used the visual and auditory representation of an item in a free recall test. This is because his results revealed it was easier to remember visual data compared to auditory data. The existence of a specialized system for recalling visual data would explain the existence of the picture superiority effect.

Potential Problems

An aspect of our study that could lead to a potential problem is that we chose words based on a high imageability value. This choice granted us words that had concrete images. This is a problem since a considerable amount of commonly used words, in reality, do not possess concrete images. An example of such a word is "country". While we can provide images of countries on the world map, there is no simple way to represent the concept of a country as an image. Had we used words with lower imageability values, we may have got more realistic results. However, we would run into the problem of participants being unable to recognize these images. We used the prior condition as being able to recognize an image would be more important than using a greater variety of images.

Related to the problem of imageability is the fact that people from different cultures possess varying mental images of what they expect a particular word would look like. An extreme example of this condition is the difference between the depiction of a dragon in western and eastern cultures. In western cultures, dragons are depicted with a pair of wings and four legs, whereas in eastern cultures, they are depicted without wings ("Dragon", 2021). Therefore, if we presented a western dragon to someone who expected to see an eastern dragon, it is likely that they might fail to recognize it and thus under-perform in the recall test. While we did not possess such an extreme case in our study, it is possible that this difference in expectations affected our results. We did not use variations of an image for people from different cultures, since this would affect the standardized nature of our test.

Future Research

While we provided proof of a picture superiority effect in a free recall test, the exact mechanism in the brain which leads to this effect is unclear and is an interesting topic to investigate.

Similarly, we only examined if the superiority effect exists when items are presented visually by images and words. It would be exciting to determine if we observe a related superiority effect when the data is represented through other sensory mediums, such as sound and taste.

Such research would be important, as it could provide insight into how the brain stores information. Therefore, it is crucial that more studies examine superiority effects in working memory.

References

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149–210.

Davies, M., & Kim, J.-B. (2019). The advantages and challenges of" big data": Insights from the 14 billion word iweb corpus. *Linguistic Research*, *36*(1), 1–34.

Dragon. (2021). In *Encyclopedia britannica*. https://www.britannica.com/topic/dragon-mythological-creature

Fais, L., Loken-Kim, K.-H., & Morimoto, T. (1996). How many words is a picture really worth?, 2179–2182 vol.4. https://doi.org/10.1109/ICSLP.1996.607236

Hayes, J. R. (1952). Memory span for several vocabularies as a function of vocabulary size. *Quarterly Progress Report*, 338–352.

Hilton, E. S. (2001). Differences in visual and auditory short-term memory. *IU South Bend Undergraduate Research Journal*, *4*, 47–50.

Hutmacher, F. (2019). Why is there so much more research on vision than on any other sensory modality? *Frontiers in Psychology*, *10*, 2246. https://doi.org/10.3389/fpsyg.2019.02246

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science*, *11*(4), 137–138. https://doi.org/https://doi.org/10.3758/BF03331011

Purtill, R. (1970). The purpose of science. *Philosophy of Science*, *37*(2), 301–306. http://www.jstor.org/stable/186678

Van Rossum, G. (2020). *The python library reference, release 3.8.2*. Python Software Foundation.

Wilson, M. (1988). Mrc psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, Computers*, *6*, 1532–5970. https://doi.org/https://doi.org/10.3758/BF03202594