Greater Neural Pattern Dissimilarity in the Retrieval Stage of Spaced Learning is Associated with Higher Meaningfulness and Better Memory

Yaowen Liang, Yuying He, Cheng ZiJian, Luo WenBo

Brain and Cognitive Neuroscience Research Center, Liaoning Normal University

Abstract

Spaced learning research aims to find the best learning interval to improve people's learning efficiency. Although it has been studied for many years, people do not know the physiological mechanism behind it clearly. Some neural representation studies have found that under the condition of spaced learning, the representation similarity between two encoding stages increases, and believe that this conflicts with the classical encoding variability theory (Xue et al., 2010) In this experiment, we used a new experimental paradigm and a longer lag between two studies to let the Chinese university students learn English-Chinese word pairs, and we introduced the idea of meaningful learning into the learning stage to improve the traditional keyword method elaboration task. Finally, by comparing the difference between (spaced learning - one-time learning) and (massed learning - one-time learning) in the final test (retrieval) stage, we get results more fitting to the classical theories: there is no significant difference in event-related potential (ERP). At the same time, the spatiotemporal pattern dissimilarity (STPDS) had significant results in the parietal lobe of 400ms and the right frontal lobe of 600ms. However, we believe that there is no direct contradiction between our experimental evidence and the experiment mentioned before. On the contrary, they reflect different aspects of the process of spaced learning. Based on the different neural representation evidence of encoding and retrieval, this paper presents an idea of reintegrating three classical theories, eliminating the opposition between encoding variability theory and deficient processing theory. We also summarized and classified the experimental paradigm of spaced learning systematically.

Keywords

spaced learning, neural representation similarity, keyword method, retrial learning, reconsolidation, ERP, elaboration, NPS, STPDS, TPDS

Intro

Spaced learning effect refers to the phenomenon that a longer learning interval can bring better memory effects when learning two or more times (Toppino & Gerbier, 2014). For brevity's sake, we could break this process down into the steps of study1-lag (learning interval)-study 2-retention interval (RI)-final test. The focus of spaced learning is the optimal interval between study1 and study 2, which is the learning interval or lag. Under the optimal learning interval, it can avoid the inhibitory effect brought by massed learning (Feng et al., 2019) and avoid the forgetting brought by too long of a retention interval (RI) (Li & Dekeyser, 2019), so that help people take full advantage of consolidation or reconsolidation and get higher learning efficiency.

Brief Review and Reconstruction of Classic Theories

There are three basic learning theories: deficient processing mechanism, encoding variability

mechanism, and study-phase-retrieval mechanism (Toppino & Gerbier, 2014).

Deficient-Processing Mechanisms try to explain that the interval should not be too short, and if the interval is too short, the second learning will result in insufficient processing, either because of the inhibition effect or because the retrieval effort is less, which will lead to poor memory gain effect (HINTZMAN et al., 1973). For example, some researchers believe that after the first presentation of an information unit, there is a residue in short-term memory, during which the learners are temporarily unable or unwilling to deal with the second presentation (). We can imagine the most extreme situation: the stimulus continuously takes twice as long, and many experiments have shown that the learning time is not the influencing factor of learning results ("a spreading activation theory of memory," 1983).

Is it good that the interval is extended as long as possible? No, as the time interval is extended, it can increase the effort of retrieval, but if it is too long, it will lead to a decreased successful rate of retrieval (Metcalfe et al., 1994). That is, there is no reminding process (Bui et al., 2014), and the failed retrieval has no promotion effect on memory (Bui et al., 2014; Spellman & Bjork, 1992). Therefore, the emphasis of study-phase-retrieval mechanisms is to make the interval as long as possible on the premise of being able to recall successfully. This theory focuses on the definition of the maximum length of RI, which is similar to a related theory about retrial (Mozer et al., n.d.; Raaijmakers, 2003), this theory is more inclined to seek a balance between success and difficulty (McDermott, 2021).

Encoding-Variability Mechanisms emphasize that the essential reason for better learning results from spaced learning is to increase the coding traces of items and cues of recall bringing richer context to help people improve the success of retrieval (Glenberg, 1979). This theory also has the problem of cue overload (Watkins, 1975). In addition, this theory is supported by the evidence of neuroscience reconsolidation (C. D. Smith & Scarf, 2017).

In addition to these three basic theories, there are many compound theories, the study-phase retrieval theory is generally combined with one of the other two (Toppino & Gerbier, 2014). For example, some theories combine study-phase retrieval with encoding variability theory (Mozer et al., n.d.; Raaijmakers, 2003), and some theories combine study-phase retrieval with deficient-processing (Braun & Rubin, 1998). Few researchers combine encoding variability theory with encoding variability. Furthermore, some studies even claim that they support deficiency-processing theory while opposing the encoding variability theory (Feng et al., 2019; Y. Lu et al., 2015; Xue et al., 2010, 2011). However, in our opinion, there is no conflict between deficiency-processing theory and encoding variability theory, since they only describe different aspects of the whole spaced learning process, and we will explain a new integration idea of these three theories in the discussion part.

Review and Classification of the Experimental Paradigm

Behavioral Paradigm Classification

The classification of behavioral paradigms can be divided according to the length of lag and type of learning task: 1. short lag versus long lag (Gerbier & Toppino, 2015) and 2. meaningless learning versus meaningful learning.

The paradigm of short lag+ meaningless learning is the most common combination, and most of them can stably produce the spacing effect (). Long lag + semantically related materials can also lead to stable spacing effect (). However, experiments on short lag + semantic materials or long lag + meaningless materials are few, since the two combinations are difficult to produce results because, for semantic materials, there is a big gap between short lag and its best lag, which leads to a ceiling effect. Similarly, for meaningless materials, long lag brings complete forgetting to most items, which leads to a floor effect. The different combination of paradigms is shown below in Table 1.

	Short lag	Long lag
Meaningless	frequent	few
Meaningful	few	less

Table 1: Classification of Behavioral Paradigm

Selection and Introduction of EEG Indicators

In addition to the previous classification criteria, according to the different EEG collected stages, we can divide the research paradigm into two categories: encoding stage versus final test stage.

As before, we transform the whole experiment into small steps: Encoding 1 -- LAG –Encoding 2 -- Retention Interval -- Final test (Retrieval). The idea of comparing encoding state can be simply presented as: similarity of massed-learning (encoding 2-encoding 1) versus similarity of spaced learning (encoding 2-encoding 1), which we call encoding-similarity (ES). However, we believe that encoding stage differences cannot reflect memory storage differences, which will be discussed in detail in the discussion part.

Here we put forward another way to compare the retrieval process, which is more in line with the theory's definition-to investigate the long-term memory state rather than encoding difference，which may be changed by other factors besides memory. We do not directly compare the retrieval stage of spaced learning condition and massed learning, instead, we compare them with the retrieval stage of once-learning (study – retention interval – final test (retrieval)). This idea can be simply presented as dissimilarity (spaced–once) versus dissimilarity (massed-once), which we call retrieval-dissimilarity, RD for short.

These two analytical ideas, combined with the previous two commonly used behavioral paradigms, can lead to four research methods: 1. short lag + meaningless materials +ES, which has been done many times (Feng et al., 2019; Y. Lu et al., 2015); 2. Short lag + meaningless material +RD, which has not been studied yet; 3. The situation of long lag + meaningful material +encoding is difficult to conduct because the study process of meaningful materials is not a process that can be simply restricted and documented by time since the stimulus appears with EEG recording, so it can only be abandoned; 4. long lag + meaningful material +retrieval, which has not been studied too.

The fourth combination is adopted in this study.

	Short lag + is meaningless	Long lag+ semantics
Encoding similarity	have	Difficult to achieve
Retrieval dissimilarity	without	without

Table 2: Classification of Neuroscience

Two kinds of EEG indicators are used to measure dissimilarity here. One is traditional ERP. We know that a large number of experiments have shown that some components of ERP are related to memory retrial performance (). We doubt whether this difference can be revealed in this experiment, Some fMRI experiments have shown that although repeated learning can enhance memory, it does not blindly bring about the enhancement of brain regions' activation, but some regions are enhanced while some areas are weakened (Hashimoto et al., 2011), which is more in line with the thought of evolutionary psychology. It is unreasonable that while the memory is strengthened, the retrieval process will cause more energy consumption. Therefore, we assume that the difference between once, spaced, and massed may be difficult to detect with simple EEG amplitude analysis.

On the other hand, we draw lessons from the previous neural pattern similarity (NPS) (Cavanagh

et al., 2018; Haxby, 2001; Z. Lu & Ku, 2020) and Spatiotemporal pattern similarity (STPS) (Y. Lu et al., 2015) in order to meet the experimental requirements. A new analysis algorithm has been formed, which we call spatiotemporal pattern dissimilarity based on ERP (STPDS). We made some changes based on the original STPS. We replaced the comparison of trial with the comparison of ERP. At the same time, the similarity was changed to dissimilarity for thinking habit. Because the three groups of words A, B, and C in the experiment were not the same words, we could not conduct the within-item STPS study. In between-item STPS, the differences between individual words are likely to overpower the differences in memory strength. In addition, using ERP rather than EEG single-trial can greatly reduce the amount of computation. The whole analysis process is shown in the figure below:

According to the encoding variability theory, we put forward two hypotheses, that is, while spaced learning can achieve better behavior performance, first, there is no significant difference between ERP difference (spaced-once) and ERP difference (massed-once); Second, STPDS (spaced-once) > STPDS (massed-once) will be significant.

Learning Materials and Learning Tasks

As to the theory of encoding variability, we know that it depends on the thought of context drift (Glenberg, 1979; Klatzky, 1975). However, experiments have shown that richer contexts may not bring better memory effects because there is a cue-overload situation (Watkins, 1975), that is, different contexts are difficult to be integrated or conflict with each other, and bring negative effects to subjects' memory. In order to avoid this situation, we quoted the concept of meaningful study (Mayer, 2002). Mayer subdivided learning into rote learning and meaningful learning in different degrees according to the characteristics of arbitrary. This concept emphasizes understanding, and through understanding, through logical connection with the knowledge in our long-term memory, reduces the arbitrariness of the association when we produce elaboration. To some extent, this is a special kind of elaboration, and we can see the shadow of meaningful learning in some previous experiments on elaboration experiments (Bradshaw & Anderson, 1982).

Keyword Method with Meaningful Learning Properties

The keyword method learning task is often used in memorizing word-pairs (Pressley et al., 1982). Some experiments require the subjects to memorize Russian-English pairs in this way. First, they provide the pronunciation of Russian words, then they are asked to associate pictures with homophonic sounds, and then they establish contact with elaboration according to the association. For example, the Russian word for "building" (zddnie) is pronounced somewhat like "zdawn-yeh" with emphasis on the first syllable. Using "dawn" as the keyword, then one could imagine the pink light of dawn reflected in the windows of a tall building (Pressley et al., 1982).

Let's analyze this process closer. We can find out that the subjects actually had transformed the study of the pair into two relatively simple associative learning parts. That is, zddney-"zdwan-yeh"-dawn and dawn–imagine–building. However, this process has strong subjective characteristics, which means that although this elaboration's thoughts conform to the thinking habits of the experiment designer, it does not necessarily match the thought pattern of the subjects, and the subjects are not given the necessary post-test after the completion of the study to check their views on the materials themselves. The worst-case may be that they even think that rote memorization is more effective than this complex association.

Furthermore, if we want subjects to score their associative learning, what method can we adopt? Some researchers simply use the strength of association to score subjectively (Nelson et al., 2004), but this way completely ignores the nature of different association processes, just like here. The first associative learning bridges the pronunciation similarity between foreign words and keywords, and the second associative learning derives from the inferences from keywords to translation. It is unreasonable to evaluate the strength of zddnie-dawn and dawn-building association in a degree together. The two associations are not even of the same type, how would we guide the subjects to score their learning result?

At first, we introduce an index to measure the association strength, and KWM is the comprehensive embodiment of the two association strengths. The following describes an example of our modification to the traditional KWM, for example, we let the subjects remember an English-Chinese word pair, disrupt–破坏. We presume the subject did not know the word 'disrupt' before, and '毁坏' is the Chinese translation of this word.

In the practice stage, we let the subjects review a simplified version of Green's formula (Campbell, 1999), which describes the transformation forms between cognate words and can help us to establish the connection between homologous words logically and naturally. For example, this rule holds that the vowels A E I O U can be transformed into each other. For example, we can let Chinese students establish the connection between the familiar and simple word 'ripe' and unknown homologous root '-rupt'(destruction） according to this rule (if 'I 'was replaced by 'U'). This can help the subjects complete the first stage of associative learning, and it is not an unfounded and arbitrary association, it is based on real knowledge, which means high logicality. Then, using etymology, the simple familiar words and definitions linked in the previous stage are logically deduced to translation. The word disrupt can be split into dis-+-rupt, and 'dis-' can be connected to the familiar word dispart or disperses, combined with '-rupt' described before, we can establish such a logical clue: the unknown word 'disrupt' means splits dispersedly which also equal destruction

Therefore, the whole logical process is transformed into a process of two associations: 1. disrupt = dis (dispart)+rupt (rip)；2. split on all sides= destruction（破坏）. We consider that the characteristics of arbitrariness are inversely proportional to the logicality of the association of words. At each test stage including an initial test conducted immediately after the learning process and final test after around 48 hours, we not only ask subjects to recall Chinese translation according to English words but also ask the subjects to score their logical thinking process by asking questions like: 'Is the material facilitating your memorizing the word pair? How reasonable is it? Could you score for it from 0, which means no logical inference aroused by it, to 10, which means smooth inference?'

It is worth noting that although the subjects we are looking for do not know word pairs, they are familiar with the simple words after splitting. Later, we will describe in detail how to match the learning materials with the subjects. The first association depends largely on the familiarity of a simple paronym.

Word pair	simple paronym
Unknown	Known

We expected to get the result that spaced learning can bring about memory improvement, and this memory improvement is accompanied by refining our understanding of the material, which means a higher logical score.

Making and Evaluation of Experimental Materials

Screening Words and Subjects

At first, the vocabulary of CET-4 and CET-6(新东方考试研究中心, 2016b, 2016a) is subtracted from GRE vocabulary (ED.M, 2010). From the rest of the vocabulary, three college students who have passed CET-6 are randomly selected to select 200 specific words that they think they have never seen before, and the corresponding etymological explanations of the words are relatively reasonable. After that, 27 students whose CET-4 scores are between 430 and 550 points are selected to access the words chosen. These 27 students stopped taking part in the subsequent experiments and only made vocabulary assessments.

Assessing and Grouping Vocabulary Lists

First, present an English word and ask them to judge whether they know the meaning of the word. In order to avoid not remembering the meaning temporarily, then present the words and translation together, and ask them to score the familiarity of this word pair again from 1 to 5. If you feel familiar with them, you should choose the first option, if you have not seen them at all, you should choose the last one. It is worth noting that as long as the subjects choose to know at the first question, the system will give this word 5 points of familiarity by default.

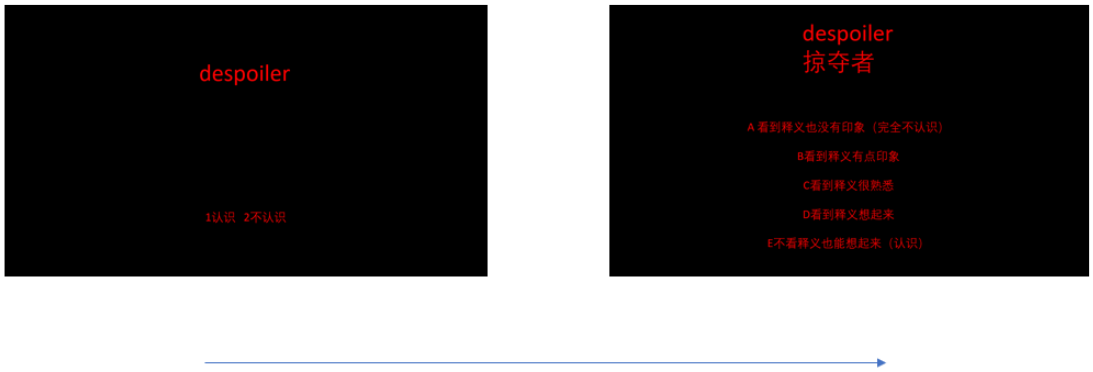


Figure 1: Vocabulary Assessing

Add all the scores of each word across all 27 students to get a sum of familiarity scores, and sort them, as shown in the following figure, select 96 words with the lowest familiarity total points. The 96 words are divided into three groups: A, B, and C. Group A and B correspond to spaced learning and massed learning conditions, while Group C corresponds to one-time learning.

There is no significant difference in familiarity evaluation scores among the three word groups, $F(2,93)=0.0747$, $P=0.9281$, and there is no significant difference in word length, $F(2, 93) = 0.002125$, $P=0.9979$.

words	sub1	sub2	sub3		subn	sum of scores
pedestal'	3	1	1	...	1	33
trajectory	1	1	1	...	1	34
juncture'	1	1	1	...	1	35
'mortuary	1	1	1	...	1	35
porcupine	1	1	2	...	1	35
'lagoon'	1	1	1	...	1	36
rayon'	4	1	1	...	1	36
skewer'	1	1	1	...	1	37
'weld'	1	1	1	...	1	37

Figure 2: Sorting Words Chosen

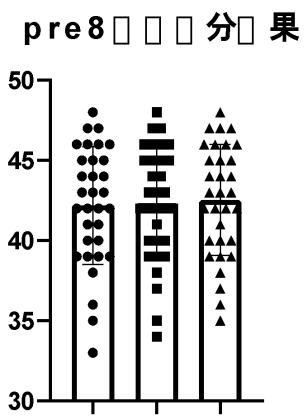


Figure 3：Familiarity Score

Descriptive statistic

	N	minimal value	max	average/mean value	standard deviation
VAR00001	40	33.00	54.00	43.7750	4.67666
VAR00003	40	35.00	55.00	44.1750	4.64586
VAR00002	40	34.00	55.00	43.9750	4.67666
Valid n (list status)	40				

Material Processing

The way of editing materials is shown in the following figure 4. Besides English-Chinese word pairs, there is also the structural analysis of words, which splits unknown words into familiar

and simple words (the first association), and finally adds the logical derivation process (the second association). All the simple cognate words here have been confirmed to belong to CET4 vocabulary.

Figure 4: Learning Material Example

Average split

A, B and C were divided into six lists……: a1a2, b1b2, c1c2 … and the familiarity is equally distributed.

Experiment 1

Experimental Method

Experimental design

The first experiment is one-factor intra-subject design, and the three levels are one-time learning, spaced learning and massed learning. In this experiment, spaced learning is defined as cross-day learning, which requires an interval longer than 48 hours, while massed learning is defined as cross-hour learning, which requires an interval shorter than 2 hours. In the final test stage, two behavior indicators were collected: a three-class judgment including 1. Can recall; 2. Can't recall, but can be recollected at the sight of Chinese options; 3. Can’t recall at all. The other is the logical score of word-pair. In addition, if subjects can't remember the words at all, the logic score should be 0. The test details will be explained below.

Participants' choice

Twelve college students from Chongqing universities were recruited in the experiment. Their CET-4 scores ranged from 425 to 500. They had not passed CET-6 or failed CET-6, with an age of 19.6±1.1, including 4 male subjects and 8 female subjects. All subjects are right-handed. Adhering to the principle of voluntary participation, this experiment was conducted after obtaining the consent of all subjects. At the same time, everyone who participated in the experiment signed the informed consent form of the experiment. After the formal experiment, the subjects will receive a certain amount of cash as a bonus.

Experimental procedure

Total Experimental Process

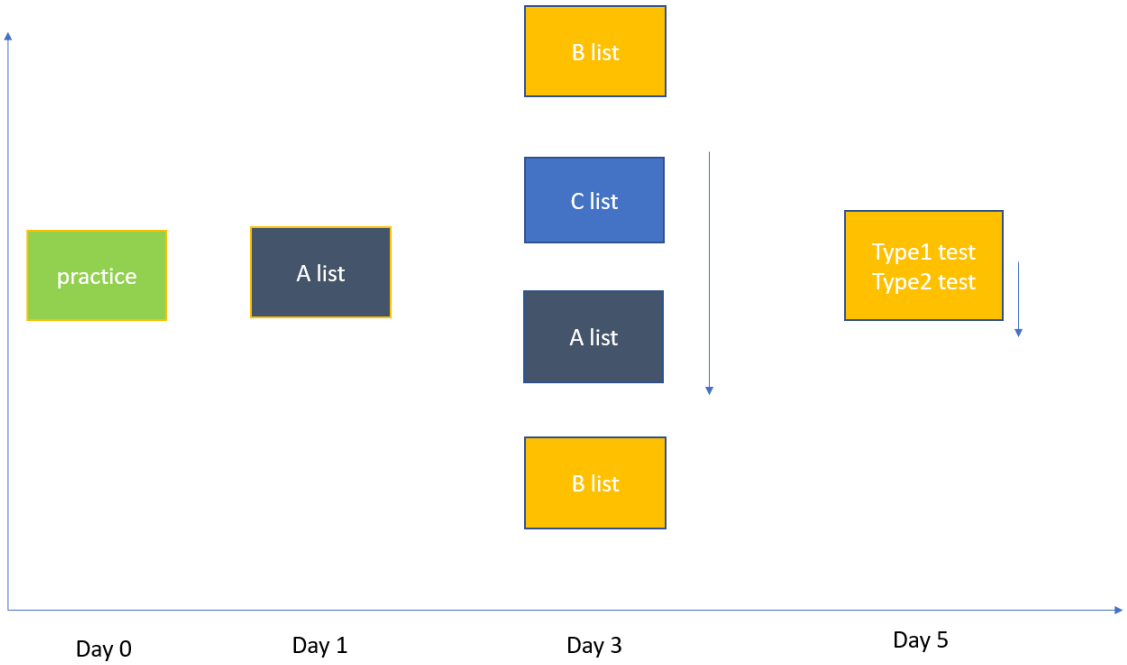


Figure 5: Overall Experimental Process

Practice procedure

Be familiar with the basics principle of Green's formula, the procedures of memorizing words, and complete the Oxford English test (). In order to avoid the effect of practice and the change of strategies, the subjects were asked to complete a 10 words list learning including processes of study and an initial test. If not all of the words were filled out (more than one or say, ten percent), the subjects were asked to repeat the whole procedure, and afterward, they were asked whether they had mastered this learning method and then asked to describe their experience.

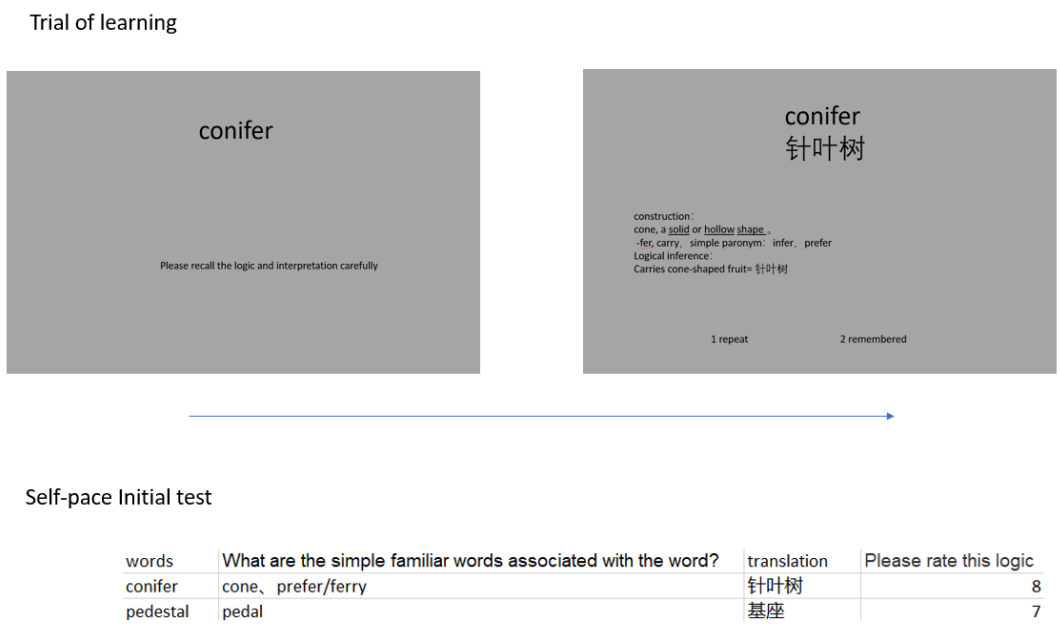
Learning Procedure

The learning process is generally divided into two steps, one is to memorize words with the help of materials provided, and the other is an initial test that requires subjects to fill in the test form, with the goal that the subjects must reach the upper limit of 100% recall rate ().

According to the foregoing, we regard the KWM as a combination of two associations, one is to associate English unknown words with simple and familiar words, and the other is to derive Chinese translation from these English words. The logical score is an evaluation of the whole of these two associations. In the initial test, it is required to type out the mediation words, that is, familiar simple cognate words (if you remember them), and give a score for the logical degree of the whole learning process. The score is based on a ten-point system, including the score of the whole KWM logical chain. 0 indicates that the word cannot be associated with long-term memory or can just be learned by rote rehearsal, and 10 indicates that the corresponding familiar words can be completely associated with existing knowledge, and the logic for deriving the Chinese meaning of new words from the familiar words is very smooth.

After the initial test is completed, check whether the subjects have completed all recall of a list. If more than two are not filled out or filled in incorrectly, the subjects are required to repeat

the learning process. This is to ensure that every word has completed the retrieval-feedback study, which means the memory strength is over the threshold of recalling ().



words	What are the simple familiar words associated with the word?	translation	Please rate this logic
conifer	cone、prefer/ferry	针叶树	8
pedestal	pedal	基座	7

Figure 6: An trial of learning

Test Procedure

Type 1 Test

The test procedure is as follows: present the word for 2 seconds, and then present four Chinese interpretations in turn, one of which is correct, and each interpretation is presented for 1.5~1.8 seconds. Ask the subjects to press the space bar when the correct answer appears, and finally ask the subjects to judge: 1. recalled the word's meaning immediately; 2 remembered it once the correct interpretation was seen; 3 can't remember.

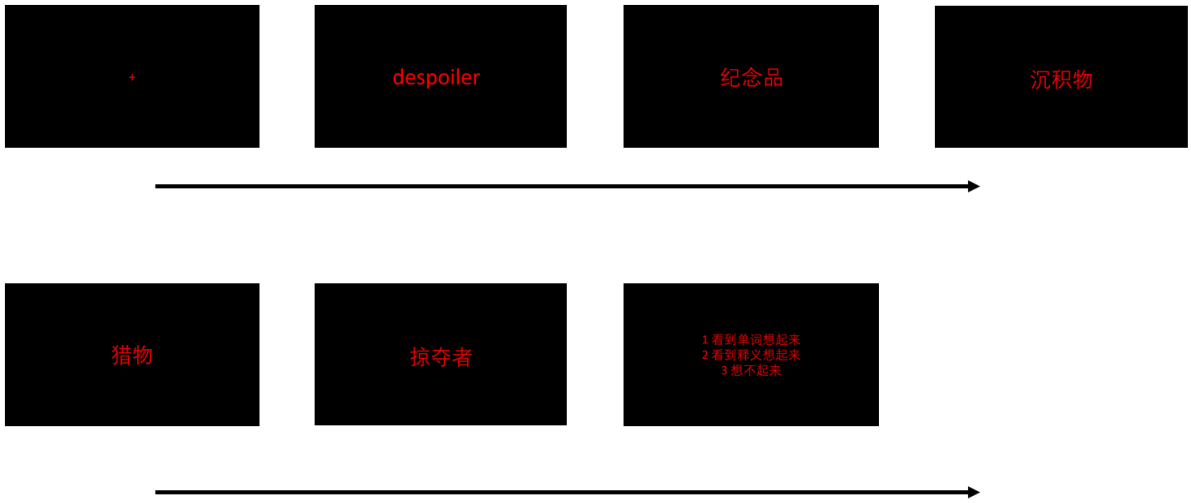


Figure 7: Test 1 procedure

Type 2 Test

Figure 8: Test 2 Procedure

Participants were asked to recall the whole train of thought from English words to Chinese translation the moment they saw the words and then scored the deduction process. They were told it is the same score process they did in the study phase and if they cannot recall the meaning of the word, they could just give a lower score.

Behavioral Data Analysis

We analyze the hit rate and logical score by condition. One-way analysis of variance and paired-sample t-test was conducted to examine the lag effect in the retrial stage.

Experimental Result

Test 1 Result (total score of memory):

If the subject chooses option 1 (recalled the word's meaning immediately), one point will be added to the sum, if the subject chooses 2 or 3, correspondingly two or three points will be added. Then every participant gets a total score for each condition.

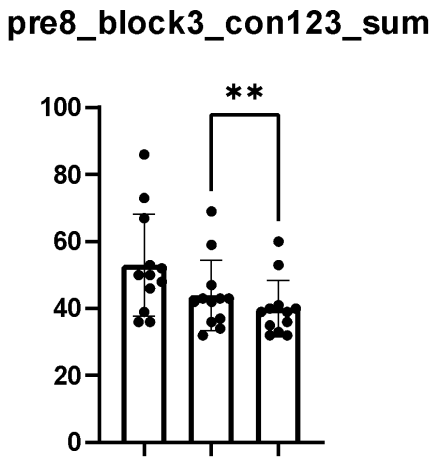


Figure 9: Total memory score.

At first, it is verified whether there is obviously spaced learning effect in the learning process of the subjects. The results show that the main effect of the learning style is significant, with F (2, 22) = 11.48, P=0.0004, R squared=0.5107. According to previous experiments, we already know that repeated learning will have a better memory effect than once learning(), so we only need to do paired sample t-test on the total recall scores of spaced learning and massed learning, t=3.764, df=11, P=0.0031, R squared=0.5629. The results show that it has an obvious spaced learning effect.

Learning effect: spaced > massed.

Test 2 Results:

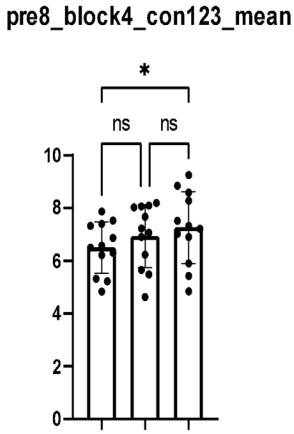


Figure 10: Average Logical Score

On the basis that spaced learning has better memory results than massed learning, we can't

help but ask whether the improvement of memory results is consistent with the logical score. We compared the average logical score among these three levels and found out the variance analysis of once learning, massed learning, and spaced learning is significant, F =7.260 P=0.0034. After repeated comparison, it can be seen that the main differences come from once learning condition and spaced learning, which indirectly shows that there are differences between massed learning and spaced learning. Generally speaking, spaced learning can bring more meaningful learning.

Logical score: spaced > massed > once

Experiment 2

The recall rate can be used as an indirect measurement index of memory strength with feedback retrieval initial test (Kornell et al., 2011; Spellman & Bjork, 1992). From the above behavior results, it can be seen that spaced learning will bring better memory strength, meanwhile, the meaningful score may rise, which shows that the subjects have a deeper understanding of the learning materials. This result makes us think that there is some correlation between the logical level and memory strength. The reason behind this is probably related to reconsolidation and activation diffusion. Spaced learning can bring more brain area connections (C. D. Smith & Scarf, 2017), more reasonable context integration, better long-term memory, and more reasonable storage construction, which should not be accompanied by stronger retrieval activation, and there is less control of resources needed during memory retrieval, but there are more different activation modes. Therefore, it is necessary to pay attention to the indicators of EEG. Here, we compare ERP (spaced-once) vs ERP (massed-once) on the one hand, TPDS (spaced-once) vs TPDS (massed-once) on the other hand (all channels together), and further compare STPDS (spaced-once) vs STPDS (massed-once).

Experimental Hypothesis

According to the encoding variability theory, spaced learning will lead to an increase in recognition rate and higher logical scores, and at the same time, it will lead to more differences in EEG representation in the retrial stage, but it will not lead to obvious amplitude differences.

Materials and Methods

Experimental Design

The learning procedure is the same as Experiment 1. In the testing stage, the type 1 test was changed to an old-new recognition task while the type 2 test remained the same. The EEG signal was recorded at the test stage.

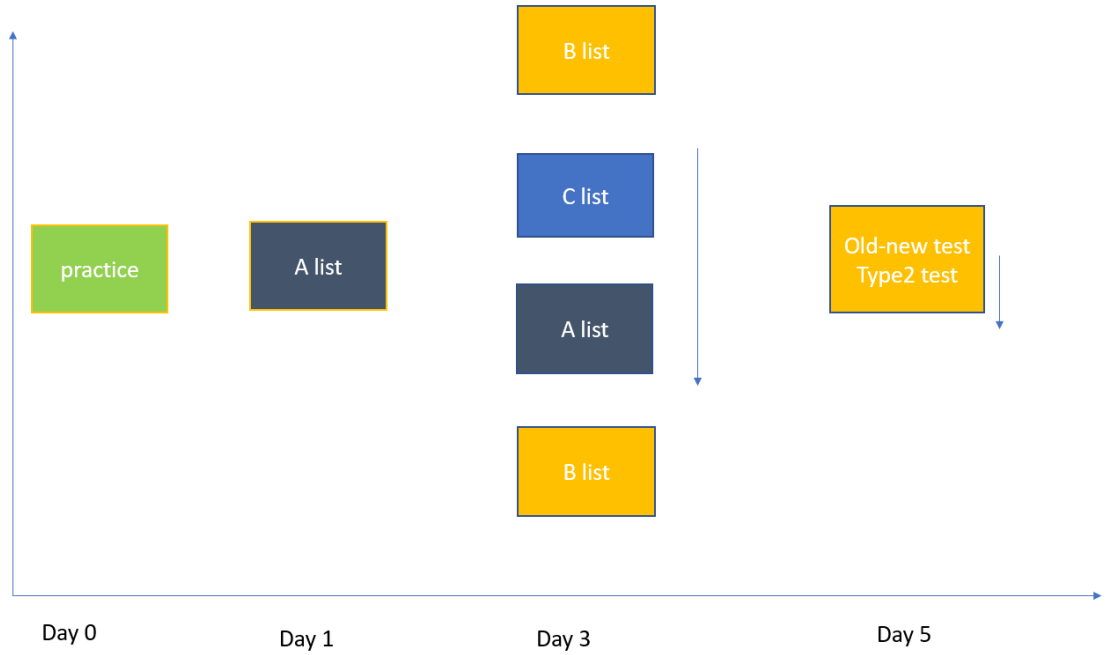


Figure 12: Overall Experimental Process of Experiment 2

Experimental Materials

In order to meet the needs of EEG experiments, the words were screened by the same method, and the table of each situation was expanded to 50 words. After the balance of familiarity and length, the difference is not significant. We also choose 100 English words with low familiarity scores as disturbance terms for the old-new tests.

Participants.

Thirty-four university students from Chongqing University of Arts and Sciences (Five were excluded due to absence, lack of sleep, poor EEG data quality, and absence) have taken the Oxford English test, and their scores are in the range. The average score in CET4 is 460±29.8 and the mean age was 19.9±1.1.

Behavioral data Analysis

EEG Recording and Preprocessing

Participants were seated about 60cm away from the computer screen in a soundproof, light-adjustable room. Continuous EEG data were recorded with a sampling rate of 512 Hz using a 64-channel BP EEG device. Ag-AgCI electrodes were mounted according to the xxx system.

We preprocessed EEG data with a Python-based toolbox called MNE () and in-house Python scripts. We re-referenced EEG data to the average of all electrodes and filtered it with a band-pass filter of 0.1-40 Hz. EEG movements were identified and corrected using an independent components analysis algorithm. The continuous data was then segmented into epochs from -200 to 800ms according to stimulus onset. The pre-stimulus interval (-200 to 0ms) was used as the baseline for baseline removal procedure. Then we apply AUTOREJECT () to reject or correct the trials automatically.

Univariate event-related potential (ERP)

We averaged the EEG responses corresponding to memorized items with confidence scores larger than 2 for each condition, then every participant got three ERPs. Then we calculated ERP difference (spaced-once) versus ERP difference (massed - once) and got no significant cluster.

The code from MNE (Gramfort et al., 2013) is used to search for a significant cluster of ERP components. This code is based on the algorithm idea put forward by Gramfoort et al., aiming at setting a more objective threshold to define whether ERP components are significant or not (S. Smith & Nichols, 2009).

TPDS and STPDS

Every subject has three corresponding ERPs, taking 100ms as the time window, sliding by one-time point, and calculating neural pattern dissimilarity (NPDS) of once versus massed (red line) and once versus spaced (blue line) in each time window.

Spearman correlation coefficient is used for correlation calculation between neural patterns in each time window. Specifically, we calculated RDM with codes from Neurora (Z. Lu & Ku, 2020) and then extracted the corresponding neural pattern dissimilarity value from it. Then every subject got two lists of values of dissimilarity, which can be drawn like figure17.

In order to check whether this difference is significant, TPDS and STPDS are carried out.

Firstly, we do statistics on neural pattern dissimilarity (NPDS), which calculate a pair t-test at each time point and draw a heat map, and the area of the picture framed means significant difference (Z. Lu & Ku, 2020). Here we label consecutive time points (≥5), which include p less than 0.01, as significant cluster figure19.

To improve spatial resolution, we further calculate the STPDS, which use the identical calculation procedural as NPDS, except we divide the whole brain into six regions

Experimental Result

Recognition Rate

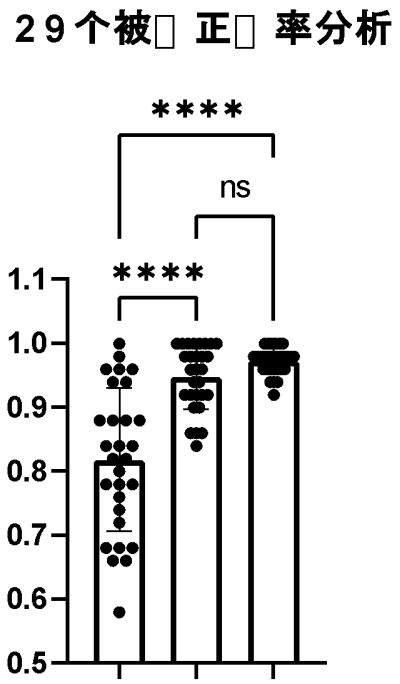


Figure 13: Hit rate analysis of recognition test.

The variance analysis of the old-new judgment experiments showed that the main effects of once learning, spaced learning and massed learning were significant F (2, 84) = 38.60, P<0.0001, and then the massed and spaced learning were tested by t, and the results were significant t=2.776, df=28, P=0.0097.

The recognition rate is still, spaced > massed.

D’

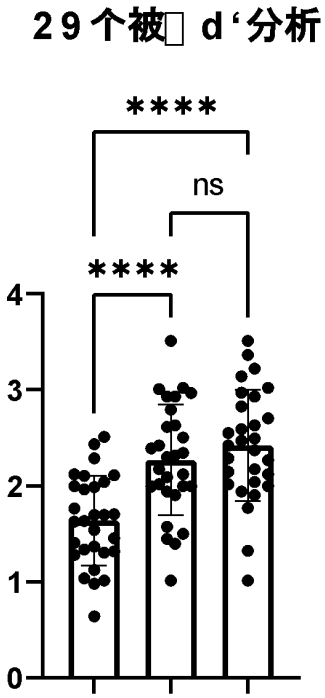


Figure 14: Analysis of Graph Recognition

The analysis of variance for three levels of d' showed that the main effect was significant ($F=17.02$, $P<0.0001$), and the t-test for spaced and massed ($t=2.141$, $df=28$, $P=0.0411<0.01$) showed no significant difference. It shows that there is no significant difference in d’ between spaced and massed.

D’: spaced ≈ massed > once

Logical Score

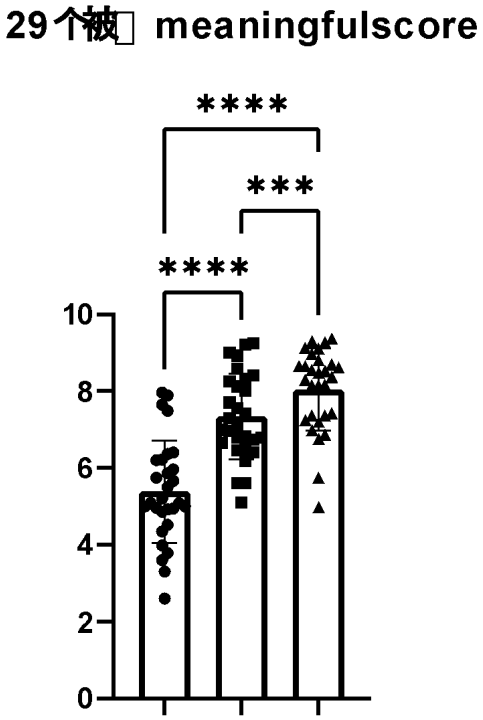


Figure 15: Logical Score

The variance analysis of the average logical scores of the three learning styles showed that the main effect was significant, $F=51.25$, $p < 0.0001$;

Therefore, the logical score is still spaced > massed > once

EEG -ERP

After calculation, no significant cluster is found. It can be seen that although spaced learning brings a better memory effect and more meaningful changes, it cannot be reflected by simple ERP amplitude.

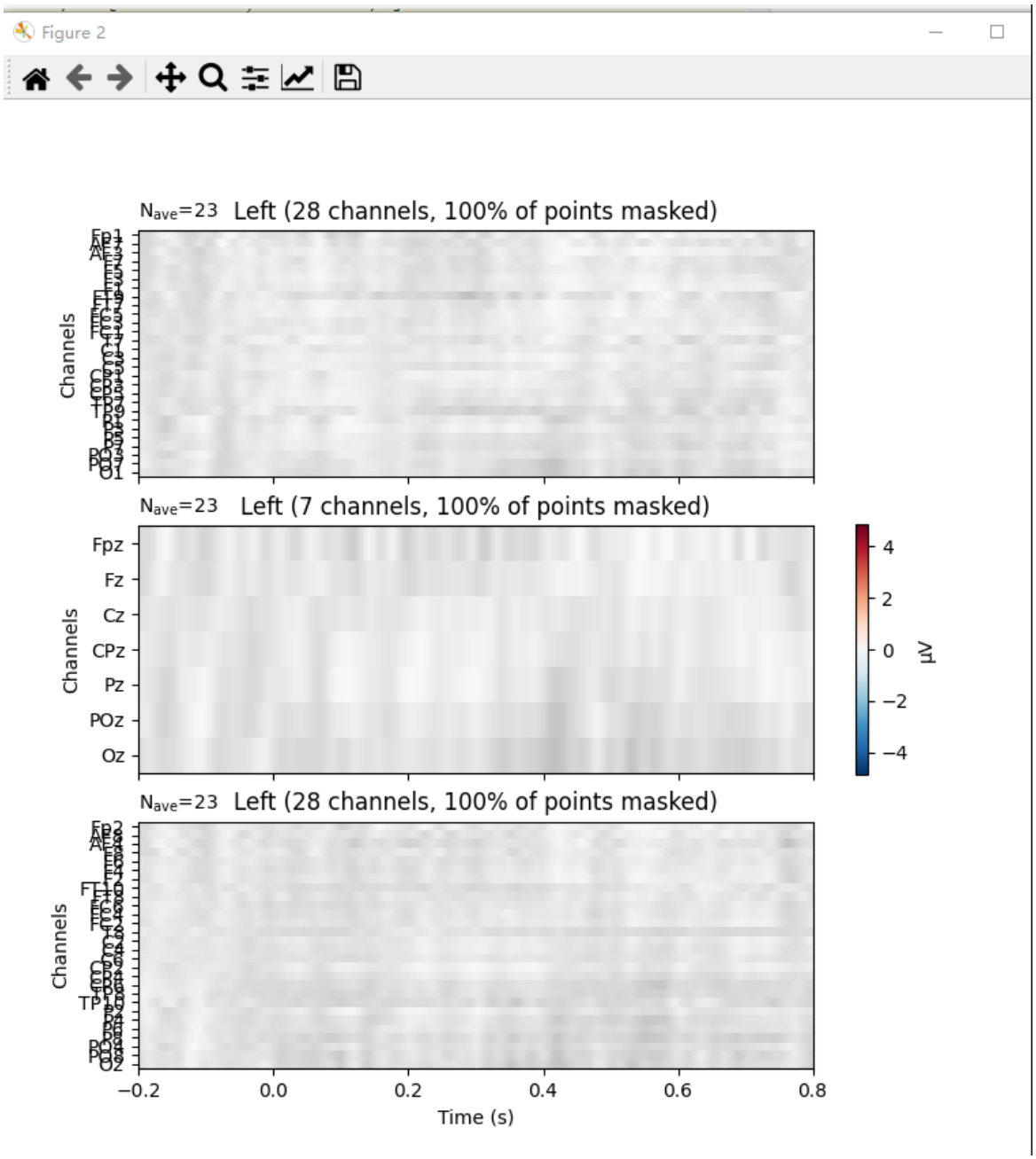


Figure 16: Significant Cluster Analysis

EEG -NPDS

It can be seen that in the early stage because the baseline part (0-0.1) is included, the dissimilarity is very large (because EEG signals fluctuate randomly). With entering the stimulation state, the representation difference begins to decrease, and then with the passage of time, the dissimilarity begins to rise. However, it can be seen that the dissimilarity (dispersed vs once) is always greater than dissimilarity (concentrated vs once).

We also find that the difference in representation gradually increases as time goes by. However, it is still below the baseline part (using the comparison between the baselines in the first 200ms as the baseline).

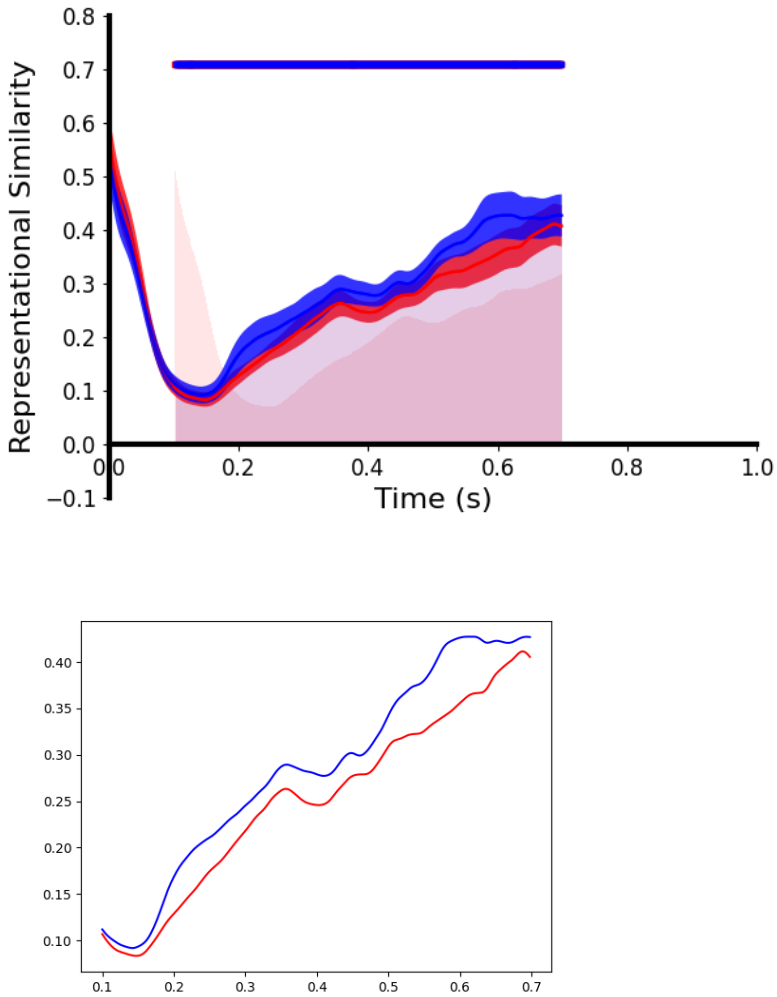


Figure 17: Neural Pattern Dissimilarity (NPDS).

TPDS and STPDS

It can be found that in the case of the whole brain, in the middle and late stage, there are significant differences around 0.4s and 0.6s. This is consistent with the retrieval of semantic memory in time, which shows that there are differences in the extraction of this part: NPDS (spaced - once) > NPDS (massed - once).

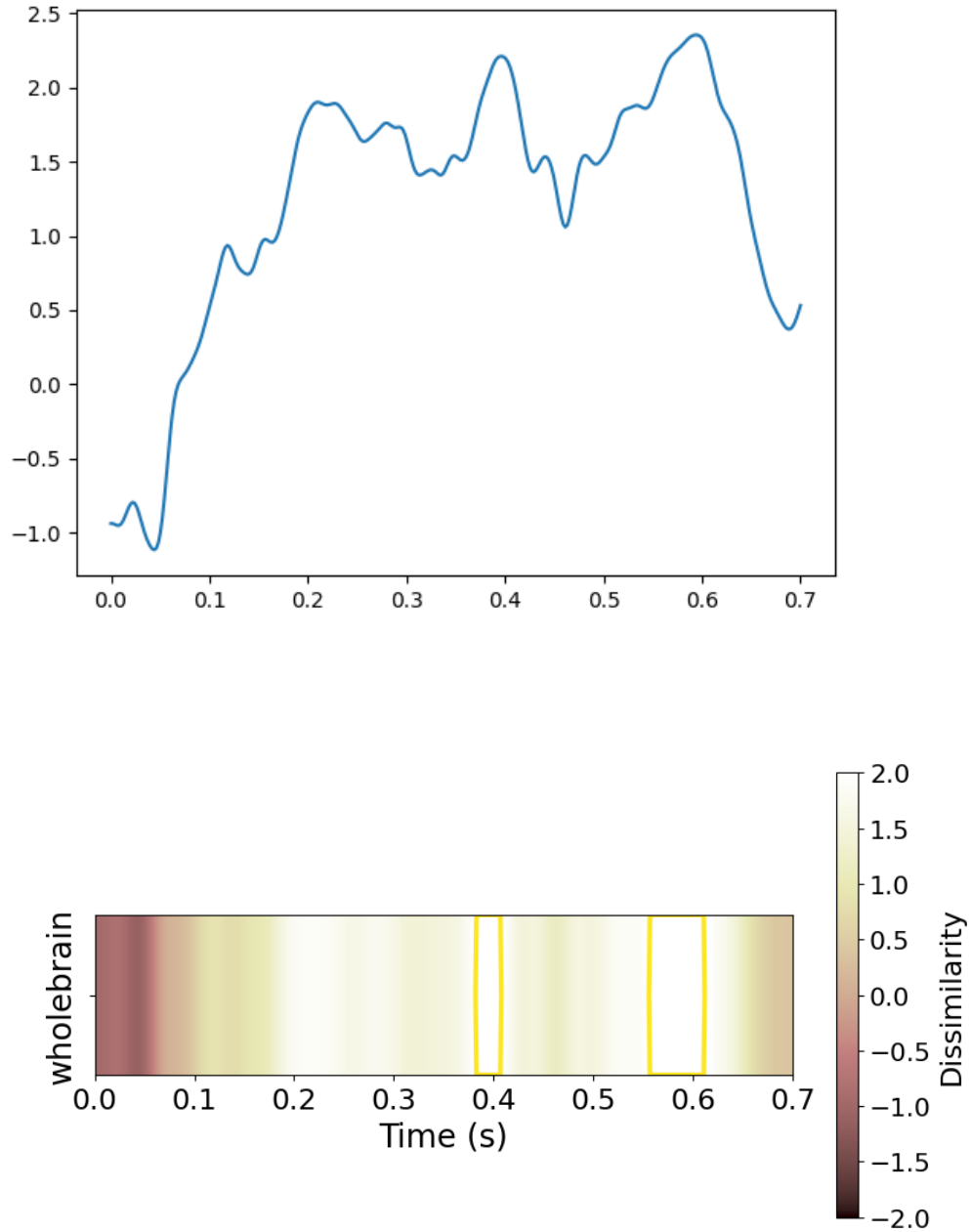


Figure 18: Temporal Pattern Dissimilarity (TPDS)

EEG – STPDS

Through STPDS calculation, we can find out that the significant region around 0.3-0.4s is region 3 and 4. The significant region around 0.5-0.6s is regions 2 and 4.

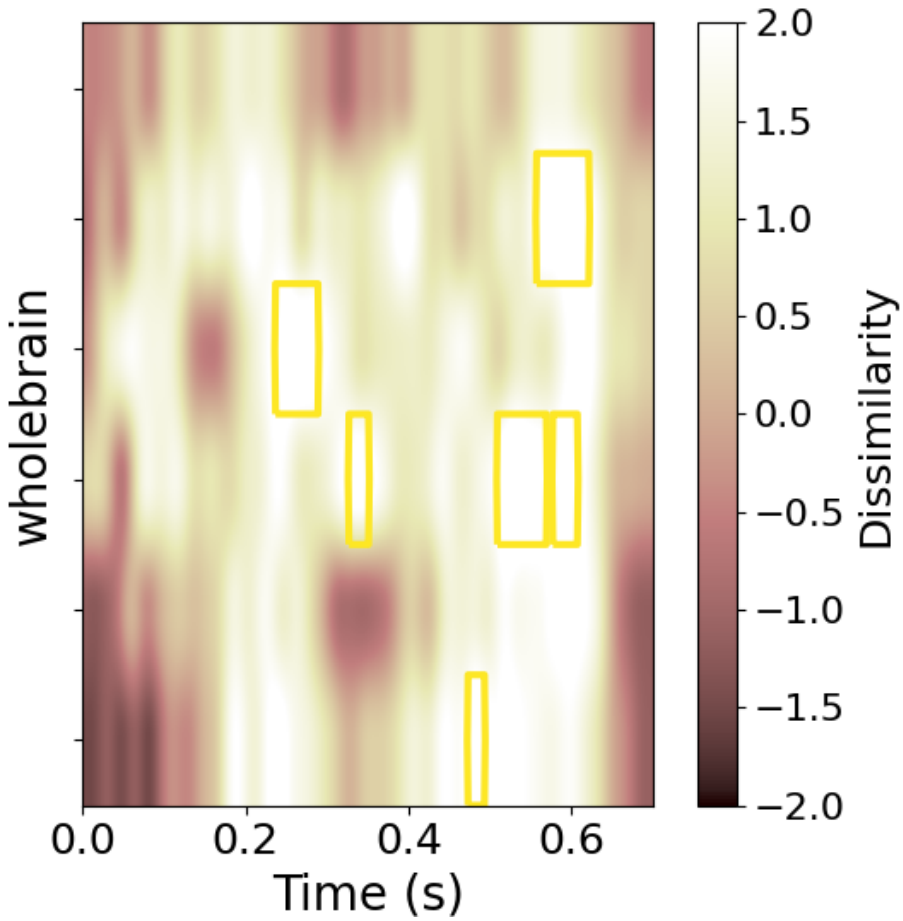


Figure 19: Spatiotemporal Pattern Dissimilarity (STPDS)

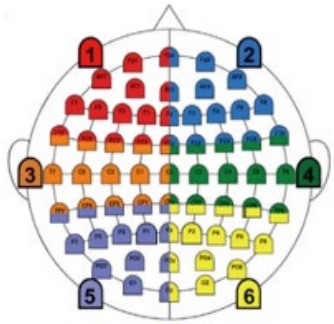


Figure 20: Division of Brain Regions.

Discussion

Part 1: Conclusions and Achievements

The first experiment shows that spaced learning can get a better recall rate, and accordingly, the logical score has increased (as shown above). Although the second experiment adopted a relatively simple old-new recognition test in the final test stage, there is still a significant increase in the correct rate. The key point is that with the better hit rate, the logical score has also been improved, which shows that the enhancement of memory strength is related to the improvement of meaningfulness in the process of learning semantic materials.

Furthermore, according to the results of experiment 2, with the increase of memory strength and meaningful score, there is no significant difference between ERP (spaced-once) and ERP (massed-once), which indicates that the improvement of long-term memory is obviously different from that of short-term memory in mechanism, it is not simply the change of activation magnitude. According to the nuclear magnetic resonance experiment in 2011 (Kornell et al., 2011), some brain regions were activated while others were weakened by repeated learning. Through the study of TPDS characterized by whole brain EEG, we can observe that at about 400ms and 600ms, TPDS (spaced-once) is significantly greater than TPDS (massed-once). Further doing STPDS, we found that the difference near 400ms mainly comes from the area 3 and 4, that is, near the parietal lobe, while the late difference mainly comes from the area 2 and 4, the right upper part of the brain-frontal lobe. This is more in line with expectations, because about 400ms happens to be the time for semantic transformation of the temporal lobe. The difference here shows that the semantic transformation or semantic retrieval process of spaced learning is quite different from the results of one-time learning, which is greater than the difference between massed learning and one-time learning. Meanwhile, the difference in the late frontal lobe can be guessed to be caused by the fewer control resources needed in spaced learning.

Part 2: Comparison of Results

Theoretical Renewal

Overall, these two experiments show that spaced learning can bring greater changes in brain storage than massed learning, and the memory of subjects becomes more logical. Moreover, this process does not blindly strengthen the activation of brain regions when extracting knowledge but produces a more different dynamic process of representation. This provides excellent proof for the theory of encoding variability. However, this does not negate the deficient processing theory, nor does it conflict with experimental results before (Xue et al., 2010). On the contrary, if we think about these two experimental phenomena together, we can get a complete picture of neural representation changes in the whole spaced learning process.

First, let's rethink why the similarity of representation of two encoding increases instead of decreases in the case of spaced learning. We believe that is because the difference in encoding does not represent the difference in memory storage. Why can't encoding processing be used to measure memory in comparing learning results between spaced learning and massed learning? There is a completely different success rate of reminding in the secondary learning of spaced learning and massed learning, and the difference in information processing nature has far exceeded the influence of memory difference.

Let's make a detailed analysis. Firstly, it is clear that the essence of spaced learning is to study the interval between primary learning and secondary learning, which is a complex psychological phenomenon (study1 – lag – study2 – RI - finaltest). When it comes to repetitive learning, we need to subdivide the two learning methods, restudy and retrieval learning (McDermott, 2021). According to whether the retrieval is successful or not, it can be divided into two forms. If retrieval is successful, it will produce retrieval effects. If retrieval fails, it will be a simple restudy. Under the condition of the same learning interval and the same retention interval, the effect of retrieval on memory strength is greater than that of restudy (McDermott, 2021). Another more appropriate theory is the Reminding Effect (Wahlheim et al., 2014). Simply put, whether the subjects know that they have learned or seen the learning materials will also affect the learning effect. If retrieval and restudy have different learning effects, we have reason to think that they have completely different neural representations (McDermott, 2021).

Furthermore, the success rate and spontaneity of retrieval are closely related to retention interval (RI) (McDermott, 2021; Wahlheim et al., 2014), in the classic experiment, it is easy to find that in the case of massed learning, that is, extremely short RI, the subjects will naturally judge, "Isn't this what I just saw?" This means that the participants remind successfully, but with the increased span of lag and the interference of information, the success rate of this judgment will decrease, that is to say, in the case of massed learning, more retrieval processes are involved, while in the case of spaced learning, it is more likely to be restudied without reminding. In this way, the difference in the encoding stage probably comes from the difference between the two kinds of cognitive processes, even if both cases are successfully recalled in the final test stage. Therefore, in Xue Gui's experiment, the representation difference in encoding does not reflect the difference in related memory, but only the difference in the proportion of encoding from top to bottom and from bottom to top, which is the difference in processing mechanism. Therefore, by default, the experiment that regards secondary learning from spaced learning and massed learning as cognitive processing of the same nature, and then thinks the difference comes from memory storage is completely wrong. Some people will question, since the spaced learning is more retrieval, shouldn't it have a better memory effect? To be clear here, the premise that retrieval is better than restudy is that they have the same RI. Accordingly, this possibility can be ruled out by comparing the similarity of representation in the retrieval stage rather than encoding.

Now, we can see the whole picture of neural representation in spaced learning. In the case of effective spaced learning, with the increase of lag between the two studies,

because there are fewer memories extracted, the process of cognitive encoding is more similar, so the similarity of representation between the two encodings rises, while the retrieval process in final test is directly affected by the extractable memories and the difference of representation rises. That is to say, the study2 process of spaced learning is more similar to the encoding processing of the study1 process, but the memory finally formed has greater changes.

In this way, we put forward another idea of theory integration: these three classical theories actually explain different aspects of spaced learning. Deficient-Processing Mechanisms and study-phase-retrieval mechanisms focus on describing the conditions triggering spaced learning. That is, when lag is too short, the difference between the two encoding processes increases. Compared with the first encoding, the second encoding has more retrieval processes from top to bottom, which leads to deficient processing. When the lag is too long, retrieval fails completely and becomes a simple restudy without reminding. Therefore, the best interval should be between these two boundaries when a considerable part of retrieval can occur, and new information is incorporated, so the learning effect at this point may be the best.

Encoding-Variability Mechanisms try to describe the internal mechanism of effective decentralized learning, that is, why. When we study repeatedly and integrate the learned knowledge with the existing knowledge, the whole memory structure will change, which is mutually confirmed with the trace transformation theory (TTT) of consolidation.

Reference

Bradshaw, G. L., & Anderson, J. R. (1982). Elaborative encoding as an explanation of levels of processing. *Journal of Verbal Learning and Verbal Behavior*, *21*(2), 165–174. https://doi.org/10.1016/S0022-5371(82)90531-X

Braun, K., & Rubin, D. C. (1998). The spacing effect depends on an encoding deficit, retrieval, and time in working memory: Evidence from once-presented words. *Memory*, *6*(1), 37–65.

Bui, D. C., Maddox, G. B., Zou, F., & Hale, S. S. (2014). Examining the lag effect under incidental encoding: Contributions of semantic priming and reminding. *Quarterly Journal of Experimental Psychology*, *67*(11), 2134–2148. https://doi.org/10.1080/17470218.2014.909506

Campbell, L. (1999). *Historical Linguistics: An Introduction*. MIT Press.

Cavanagh, S. E., Towers, J. P., Wallis, J. D., Hunt, L. T., & Kennerley, S. W. (2018). Reconciling persistent and dynamic hypotheses of working memory coding in prefrontal cortex. *Nature Communications*, *9*(1), 3498. https://doi.org/10.1038/s41467-018-05873-3

ED.M, 美]Philip Geer. (2010). *GRE 巴朗词表* (1er édition). 北京语言大学出版社.

Feng, K., Zhao, X., Liu, J., Cai, Y., Ye, Z., Chen, C., & Xue, G. (2019). Spaced Learning Enhances Episodic Memory by Increasing Neural Pattern Similarity Across Repetitions. *Journal of Neuroscience*, *39*(27), 5351–5360. https://doi.org/10.1523/JNEUROSCI.2741-18.2019

Gerbier, E., & Toppino, T. C. (2015). The effect of distributed practice: Neuroscience, cognition, and education. *Trends in Neuroscience and Education*, *4*(3), 49–59. https://doi.org/10.1016/j.tine.2015.01.001

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*(2), 95–112. https://doi.org/10.3758/BF03197590

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*, 267. https://doi.org/10.3389/fnins.2013.00267

Hashimoto, T., Usui, N., Taira, M., & Kojima, S. (2011). Neural enhancement and attenuation induced by repetitive recall. *Neurobiology of Learning and Memory*, *96*(2), 143–149. https://doi.org/10.1016/j.nlm.2011.03.008

Haxby, J. V. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

HINTZMAN, D., BLOCK, R., & SUMMERS, J. (1973). MODALITY TAGS AND MEMORY FOR REPETITIONS - LOCUS OF SPACING EFFECT. *Journal of Verbal Learning and Verbal Behavior*, *12*(3), 229–238. https://doi.org/10.1016/S0022-5371(73)80013-1

Klatzky, R. L. (1975). *Human memory: Structures and processes* (p. 276). W. H. Freeman.

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*(2), 85–97. https://doi.org/10.1016/j.jml.2011.04.002

Li, M., & Dekeyser, R. (2019). Distribution of Practice Effects in the Acquisition and Retention of L2 Mandarin Tonal Word Production. *The Modern Language Journal*, modl.12580. https://doi.org/10.1111/modl.12580

Lu, Y., Wang, C., Chen, C., & Xue, G. (2015). Spatiotemporal neural pattern similarity supports episodic memory. *Current Biology: CB*, *25*(6), 780–785.

https://doi.org/10.1016/j.cub.2015.01.055
Lu, Z., & Ku, Y. (2020). NeuroRA: A Python Toolbox of Representational Analysis From Multi-Modal Neural Data. *Frontiers in Neuroinformatics*, *14*, 563669. https://doi.org/10.3389/fninf.2020.563669
Mayer, R. E. (2002). Rote Versus Meaningful Learning. *Theory Into Practice*, *41*(4), 226–232. https://doi.org/10.1207/s15430421tip4104_4
McDermott, K. B. (2021). Practicing Retrieval Facilitates Learning. *Annual Review of Psychology*, *72*(1), 609–633. https://doi.org/10.1146/annurev-psych-010419-051019
Metcalfe, B. in the D. of P. J., Metcalfe, J., & Shimamura, A. P. (1994). *Metacognition: Knowing about Knowing*. MIT Press.
Mozer, M. C., Pashler, H., Cepeda, N., Lindsey, R., & Vul, E. (n.d.). *Predicting the Optimal Spacing of Study: A Multiscale Context Model of Memory*. 9.
Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407. https://doi.org/10.3758/BF03195588
Pressley, M., Levin, J. R., & Delaney, H. D. (1982). The Mnemonic Keyword Method. *Review of Educational Research*, *52*(1), 61–91. https://doi.org/10.3102/00346543052001061
Raaijmakers, J. G. W. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science*, *27*(3), 431–452. https://doi.org/10.1207/s15516709cog2703_5
Smith, C. D., & Scarf, D. (2017). Spacing Repetitions Over Long Timescales: A Review and a Reconsolidation Explanation. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00962
Smith, S., & Nichols, T. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061
Spellman, B. A., & Bjork, R. A. (1992). When Predictions Create Reality: Judgments of Learning May Alter What They Are Intended to Assess. *Psychological Science*, *3*(5), 315–317. https://doi.org/10.1111/j.1467-9280.1992.tb00680.x
Toppino, T. C., & Gerbier, E. (2014). About Practice. In *Psychology of Learning and Motivation* (Vol. 60, pp. 113–189). Elsevier. https://doi.org/10.1016/B978-0-12-800090-8.00004-4

Wahlheim, C. N., Maddox, G. B., & Jacoby, L. L. (2014). The Role of Reminding in the Effects of Spaced Repetitions on Cued Recall: Sufficient but Not Necessary. *Journal of Experimental Psychology-Learning Memory and Cognition*, *40*(1), 94–105. https://doi.org/10.1037/a0034055

Watkins, M. J. (1975). Inhibition in recall with extralist "cues." *Journal of Verbal Learning and Verbal Behavior*, *14*(3), 294–303. https://doi.org/10.1016/S0022-5371(75)80073-9

Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J. A., & Poldrack, R. A. (2010). Greater Neural Pattern Similarity Across Repetitions Is Associated with Better Memory. *Science*, *330*(6000), 97–101. https://doi.org/10.1126/science.1193125

Xue, G., Mei, L., Chen, C., Lu, Z.-L., Poldrack, R., & Dong, Q. (2011). Spaced Learning Enhances Subsequent Recognition Memory by Reducing Neural Repetition Suppression. *Journal of Cognitive Neuroscience*, *23*(7), 1624–1633. https://doi.org/10.1162/jocn.2010.21532

新东方考试研究中心. (2016a). *六级英语新大纲词汇表*. 浙江教育出版社.

新东方考试研究中心. (2016b). *四级英语新大纲词汇表*. 浙江教育出版社.