TITLE: Greater neural pattern dissimilarity in retrieval stage of spaced learning is associated with higher meaningfulness and better memory.

Liang Yaowen, He Yuying

Brain and cognitive neuroscience research certer, 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 of two encoding increases, and believe that this conflicts with the classical encoding variability theory. In this experiment, we used a new experimental paradigm and a longer lag to let the Chinese university students to learn English-Chinese word pairs, and we introduced the idea of meaningful learning into the learning stage to improve the traditional keyword method experiment task. Finally, by comparing the difference (spaced learning - one-time learning) vs. difference (massed learning - one-time learning) in the final test (retrieval) stage, we get different results: there is no significant difference in ERP, At the same time, the neural spatial 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 the two experimental evidence. On the contrary, they reflect different aspects of the process of spaced learning. Based on the evidence of different neural representation evidence of encoding and retrieval, this paper presents an idea of reintegrating three classical theories, eliminates the opposition between encoding variability and defective processing, and summarizes and classifies the experimental paradigm of spaced learning.

Key word

Spaced learning, Neural representation similarity, keyword method, retrial learning, reconsolidation

Intro

Spaced learning effect refers to the phenomenon that a longer learning interval can bring better memory effect when learning twice or more (Toppino & Gerbier, 2014). If we break this process down into the steps of study1-lag-study2-retention interval-final test. The focus of spaced learning is the optimal interval between study1 and study2. 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 retention interval(RI) (Li & Dekeyser, 2019), so that it can take less time to get better learning results.

A brief review and reconstruction of the theory.

There are three basic learning theories: deficiency handling mechanism, coding variability mechanism and retrieval mechanism in learning stage (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, but experiments have shown that in many cases, the learning time is not the influencing factor of learning results ("a spreading activation theory of memory," 1983).

So is the interval as long as possible? No, although the time interval is lengthened, it can increase the effort degree of retrieval, but if it is too long, it will lead to the decrease of the success rate of retrieval (Metcalfe et al., 1994), that is, there is no reminding process (Bui et al., 2014a), and the failed retrieval has no promotion effect on memory (Bui et al., 2014b; 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, so as to obtain a better memory effect. Therefore, this theory focuses on the definition of the maximum length of RI (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 mode of item and cue of recall, and richer context can help us improve the success rate of retrieval (Glenberg, 1979a). But this theory also has the problem of cue overload (Watkins, 1975a). 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, and the study-phase retrieval theory is generally combined with one of the other two theories (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 and Rubin (1998)). Few researcher combine encoding variability theory with encoding variability one. Furthermore, some studies claim that they support deficiency-processing theory while opposing the encoding variability theory (Xue et al., 2011), but in our opinion, there is no conflict between deficiency-processing and encoding variability, they only describe different aspects of the whole spaced learning process, and we will explain a new integration idea of these three theories at the end.

Review and supplement of experimental paradigm.

Behavioral paradigm classification.

There are three major trends in the spaced learning topic: interval tends to span the lag between two studies, experimental materials tend to be more complex and meaningful, and observation means expand from basic behavioral indicators to observation means of neural science (citespace analysis chart). This is basically consistent with the trend of memory research ().

The classification of behavioral paradigms can be divided according to the length of lag and experimental materials: 1.short lagVS long lag (Gerbier & Top Pino, 2015); 2.meaningless materials VS meaningful materials.

The paradigm of short lag+ meaningless materials is the most common and used, and most of them can stably make lag effect. Long lag+ semantically related materials can also make spacing effect stably. However, experiments on short lag+ semantic materials or long lag+ meaningless materials are few. Why? 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 ceiling effect. Similarly, for meaningless materials, long lag brings complete forgetting to most item, which leads to floor effect.

	Short lag	Long lag
Meaningless	Often used	few
Meaningful	few	less

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 VS final test stage.

We expressed the whole experimental mode as Encoding1 -- LAG --Encoding2 -- Retention Interval -- Final test (Retrial). The first idea of comparing encoding can be simply presented as: similarity of massed-learning (encoding 2-encoding 1) VS similarity of spaced learning (encoding 2-encoding 1), which we call encoding-similarity, ES. However, we believe that encoding stage differences do not reflect memory storage differences, which will be discussed in detail in the last.

Here we put forward another way to compare retrieval process, which is more in line with the theoretical definition-to investigate the long-term memory state rather than encoding difference, which may be changed by other factors besides memory. We don't directly compare spaced learning condition and massed condition, but compare it with the retrieval of once-learning (study – retention interval – final test) result, which can be simply expressed as dissimilarity (spaced–once) vs 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 by Xue Gui (Feng et al., 2019; Y. Lu et al., 2015; Xue et al., 2010); 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 appearing with EEG recording, so it can only be abandoned; 4.long lag+ meaningful material +retrieval, which has not been studied.

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

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 is related to memory retrial performance. We doubt whether this difference can be revealed in this experiment, because some nuclear magnetic experiments have shown that although repeated learning can enhance memory, it does not blindly bring about the activation and enhancement of brain regions, 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 illogical to imagine that while the memory is strengthen, retrieval process will cause more energy consumption. Therefore, we assume that the difference between once vs spaced vs massed may be difficult to detect with simple EEG amplitude analysis.(Hashimoto et al., 2011)

On the other hand, we draw lessons from the previous neural pattern similarity (NPS) (cavanagh et al., 2018; Haxby et al., 2001; Z. Lu&Ku, 2020a) and Spatio temporary 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 Spatio temporary pattern dissimilarity based on ERP(STPDS). We made some changes on the basis of 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, this can greatly reduce the amount of computation. The whole analysis process is shown in the figure below:

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

Learning materials and learning tasks.

As to the theory of encoding variability, we know that it depends on the thought of context drift (Glenberg, 1979b; Klatzky, 1975), but experiments have shown that richer contexts may not bring better memory effect, because there is a cue-overload situation (Watkins, 1975b), that is, different contexts are difficult to be integrate or conflict with each other, but the memory effect of the subjects is reduced. 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. To some extent, this is a special kind of interaction, and we can see the shadow of meaningful learning in some previous experiments on elaboration (Bradshaw & Anderson, 1982).

Keyword method with meaningful learning property.

The keyword method learning task is often used in memorizing word-pair (Pressley et al., 1982). Some experiments require the subjects to memorize Russian-English pair 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 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, 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 detailly, the subjects actually transformed the study of pair into two relatively simple associative learning. That is, zddney-"zdwan-yeh"-dawn and dawn–imagine–building, but this process has strong subjective characteristics, which means that although this elaboration thoughts conforms to the thinking habits of the experiment designer, it does not necessarily match the thoughts pattern of the subjects, and the subjects are not given the necessary post-test after the completion of the study to understand 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, regarding associative learning, we can simply use the strength of association to score subjectively (Nelson et al., 2004), but this way completely ignores the nature of association, just like here. The first associative learning studies the pronunciation similarity between foreign words and keywords, and the second associative learning studies the derivation from keywords to translation. How can we evaluate the strength of zddnie-dawn and dawn-building association? The two association 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 a English-Chinese word pair, disrupt–破坏. We presume the subject don't know the word 'disrupt' before, and '毁坏' is the Chinese translation of this word.

In the practice stage, we let the subjects learn 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 mature and unknown homologous root '-rupt'（destruction）and the familiar word 'ripe' through 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 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 disrupts can be split into dis-+-rupt, and dis- can be connected to the familiar word dispart or disperses, in addition to -rupt described before, we can establish such a logical clue: it splits dispersedly which means destruction

Therefore, the whole logical process is transformed into a process of two association: disrupt = dis (dispart)+rupt (rip)；split on all sides= destruction（破坏）. At each test stage, we ask the subjects to score their logical thinking process. We think that the characteristics of arbitrariness are inversely proportional to the logicality of the association of words, because in the testing stage, we not only ask the subjects to recall Chinese definitions, but also ask them to recall the keyword used and score the logicality of their whole thoughts.

It is worth noting that although the subjects we are looking for don't know word pair, 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 simple paronym.

Word pair	simple paronym	Association1	Association2
Unknown	Known		

We expect to get the result that spaced learning can bring about memory improvement, and this memory improvement is accompanied by refining our understanding of materials, 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 (New Oriental Examination Research Center, 2016b, 2016a) is subtracted from GRE vocabulary (ED.M, 2010). From the rest 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 are relatively reasonable. 27 students whose CET-4 scores are between 430 and 550 points are selected. These 27 students stopped taking part in the subsequent experiments and only made vocabulary assessment.

Accessing and grouping vocabulary lists.

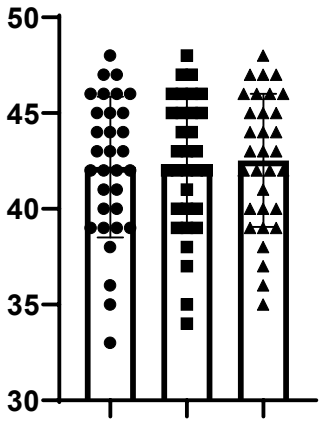
First, present the words and judge whether they know each other. In order to avoid not remembering the meaning temporarily, then present the words and translation, and ask them to judge the familiarity again. If you feel familiar with them, you will give 5 points, and if you haven't seen them at all, you will score 0 points. It is worth noting that as long as the subjects choose to know at the beginning question, the system will give this word 5 points by default.

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

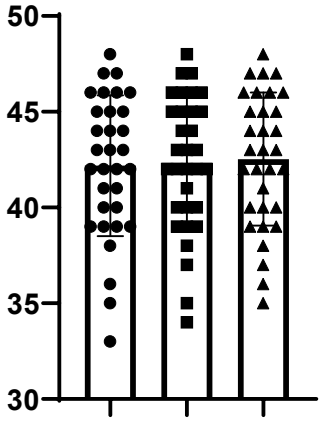
There is no significant difference in familiarity evaluation scores among the three groups, $F(2,93)=0.0747$, $P=0.9281$, and there is no significant difference in word length, $F(2, 93) = 0.002125$, $P=0.9979$.

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

pre8 单词评分结果



pre8 单词评分结果



描述统计量

	N	极小值	极大值	均值	标准差
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
有效的 N （列表状态）	40				

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. Besides cue-target, 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.

conifer
针叶树

结构:
cone, 锥形。
-fer, 携带，含有，词源同 infer，prefer偏爱（先拿），ferry因其锥形松果而得名。

逻辑:
含有锥形的果实 = 针叶树（比如松树）

1 重来　　2记住了

conifer
针叶树

construction：
cone, a solid or hollow shape 。
-fer, carry，simple paronym：infer，prefer
Logical inference：
Carries cone-shaped fruit= 针叶树

1 repeat　　2 remembered

Average split

B c is divided into six list……: a1a2, b1b2, c1c2 … and the familiarity is equal distributed .

Experiment 1

Experimental method

Experimental design

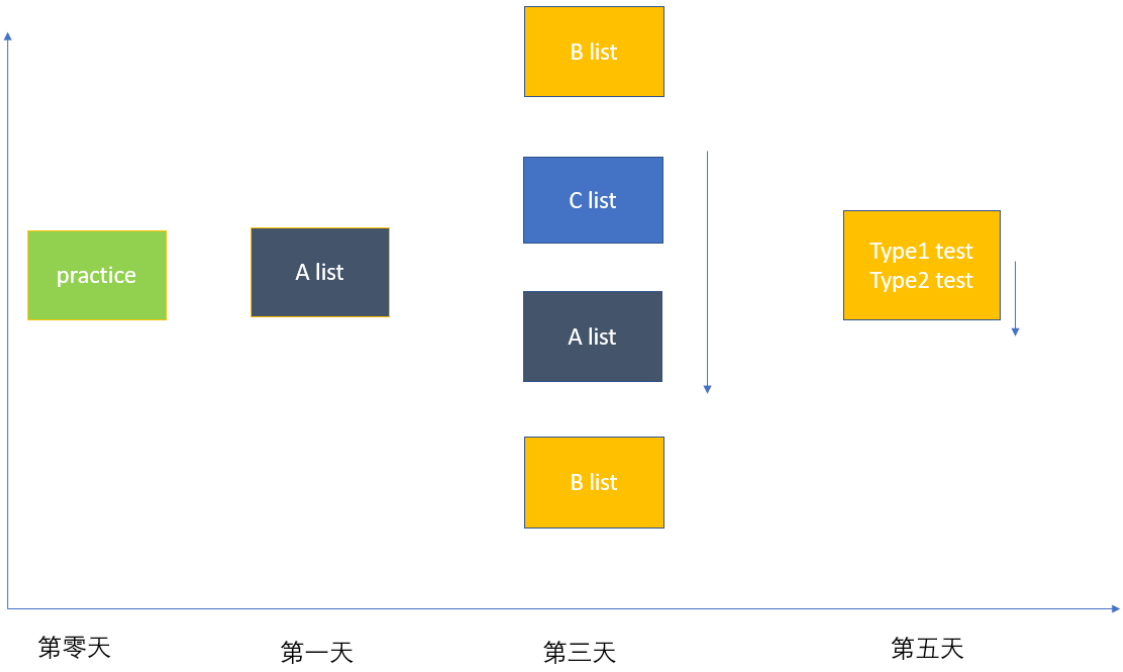
The first experiment is one-factor intra-subject design, and the three levels are one-time learning, spaced learning and massed learning. In this experiment, spaced learning is defined as cross-day learning, which require an interval longer than 48 hours, while massed learning is defined as cross-hour learning which require an interval shorter than 2 hours. In the final test stage, two indicators were collected: a three-class judgment : can recall; can't recall, but can recollected when seeing Chinese options; The other is the logical score of word-pair. If you can't remember at all, the logic score will be 0. The test detail will be explained below.

Participants' choice

Twelve college students from Chongqing universities were recruited in the experiment. Their intelligence was normal, and 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 bonus.

experimental procedure

Total 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 complete word list learning of 10 words including processes of study and initial test. If not all of words were filled out (more than one), the subjects were asked to repeat the whole procedure, and afterwards they were asked whether they had mastered this learning idea and let them describe it.

* Learning procedure.

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

According to the foregoing, we regard the KWM as a combination of two association, one is to associate English new words with English simple familiar words, and the other is to derive Chinese definitions from these English simple familiar words. The logical score is an evaluation of the whole of the two association. In the initial test, it is required to memorize 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 whole word cannot be associated or can just be learned by rote, and 10 indicates that the corresponding familiar words can be completely associated, 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().

Trial of learning

Self-pace Initial test

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

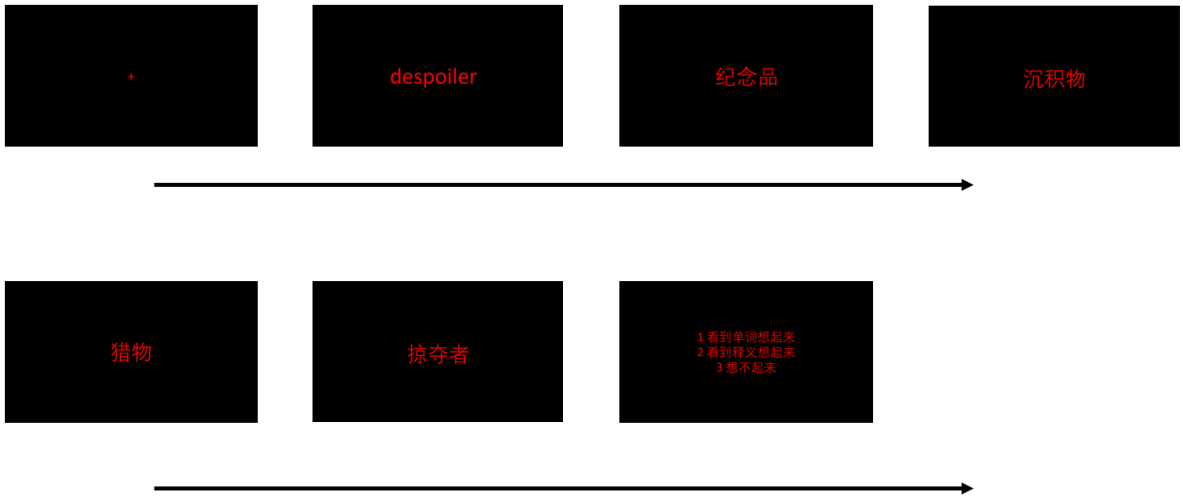
figure: As a new word, the word despoiler can be split into de- and spoil, and de- can be associated with the familiar word "break". logic: the person who destroys from top to bottom = predator. There are two layers of connections here, the first is to split association, the second is the logic of combination, and the meaning score is an evaluation of the whole.

test procedure

The first kind of test.

The first type of test requires judging three situations. In order to ensure the honesty of the subjects, the subjects will be prompted to conduct spot checks after the end. In order to avoid extra study caused by too many spot checks, only one or two spot checks were actually made.

The test process 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 word’s meaning immediately; 2 remembered it until you see the correct interpretation; 3 can't remember.



Type ii test.

Participants were asked to recall the whole train of thought from English words to Chinese translation at the moment they saw the words, and then scored the deduction process. Tell them it is the same process they trained in study phase and if they cannot recall the meaning of the word or thinking the process is not that making sense, they could just give a lower score.

experimental result

Test1 result (total score of grade 3):

If the subject choose option1(recalled word's meaning immediately), one point will be added to the sum, etcetera. Then every participant get a total score for each condition.

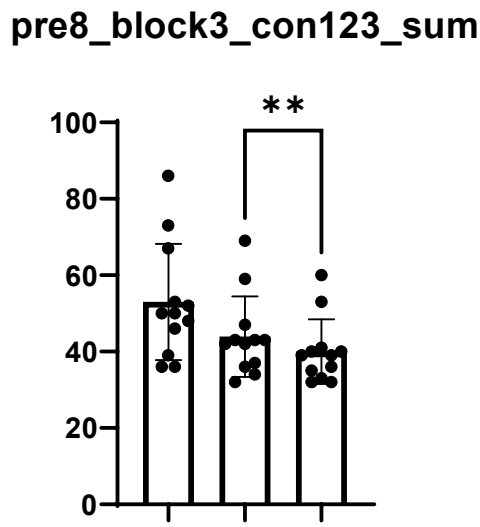


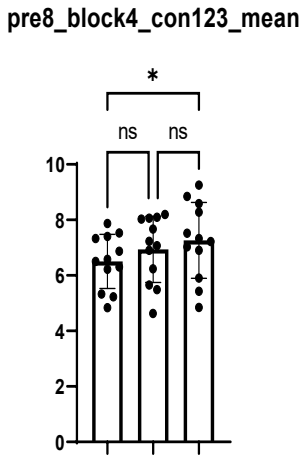
图 1: 记忆总分

Figure 1: Total memory score.

At first, it is verified whether there is obvious spaced learning effect in the learning process of the subjects. The results show that the main effect of 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 secondary learning will have better effect than spaced learning(), so we only need to do paired sample t test on the total recall scores of spaced learning and centralized learning, without correction, t=3.764, df=11, P=0.0031, R squared=0.5629. The results show that it has obvious spaced learning effect.t=3.764, df=11

Learning effect: spaced > massed.

Test2 results:



图：意义性评分
Figure: Meaning 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 meaning 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 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 understanding.

Meaning: spaced > massed > once.

Experiment 2

The recall rate can be used as an indirect measure of memory strength with feedback retrieval initial test (Bjork & Bjotk, 1992; Kornell et al., 2011), from the above behavior results, it can be seen that spaced learning will bring better memory strength, meanwhile, 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 logical level and memory strength. The reason behind this is probably related to reconsolidation and activation diffusion, and 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 pressure during extraction, 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, but it will not lead to obvious amplitude differences.

Experimental method

** experimental design.*

The learning process is exactly the same as Experiment 1. In the testing stage, Type 1 test was changed to old-new recognition task while type 2 test remain the same. The EEG signal was record at the test stage.

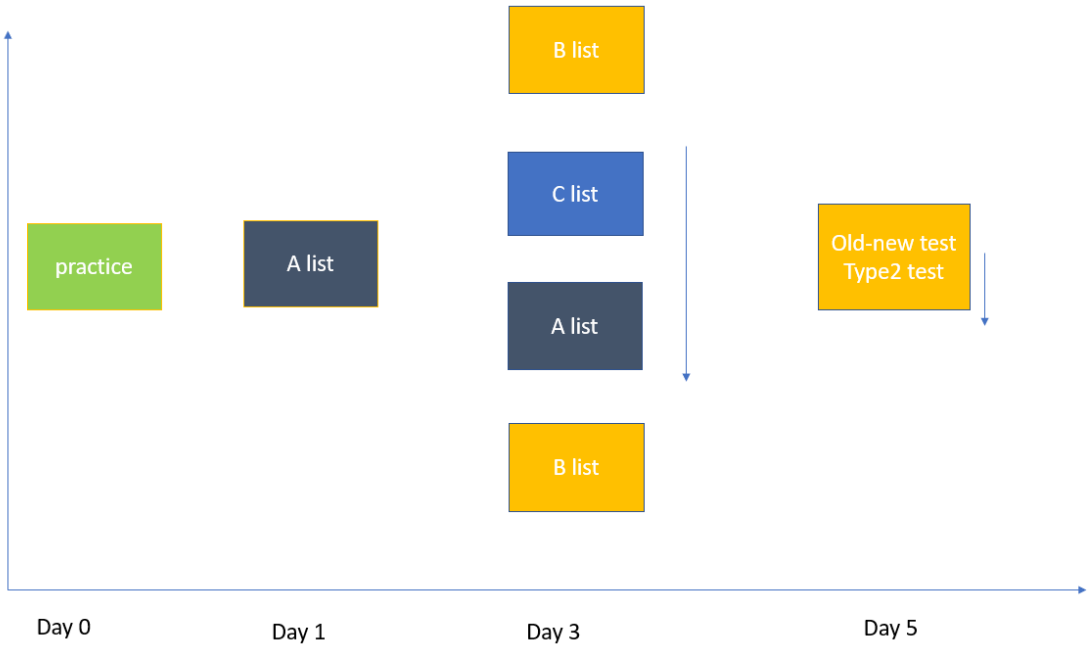


Figure1.1 General experimental process.

**实验材料*

** 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.

** participants' choice.*

Thirty-four university students from Chongqing University of Arts and Sciences (Five were excluded due to lateness, 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.

experimental result

recognition test.

29个被试正确率分析

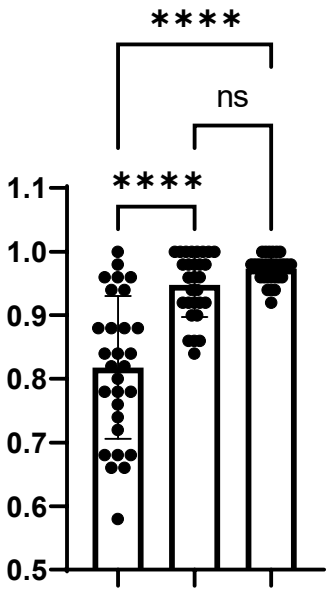


Figure Hit rate analysis of graph recognition.

The variance analysis of the old-new judgment experiments showed that the main effects of one-time 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.

再认正确率上依然是，分散>集中>一次

Recognition rate is still, spaced > massed.

Difference in resolution.

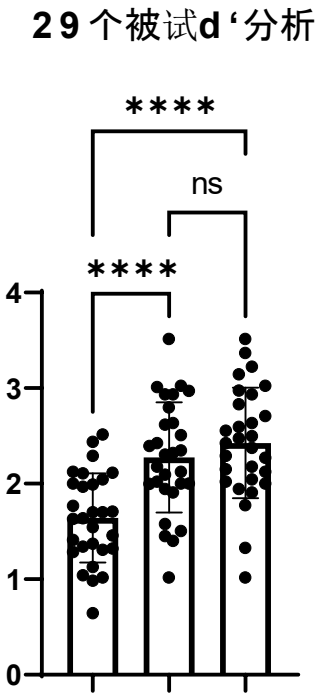


Figure D' analysis of graph recognition.

In the same way, 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 dispersion and concentration (t=2.141, df=28, P=0.0411<0.01) showed no significant difference. It shows that there is no significant difference in resolution between dispersion and concentration.

Discriminatory dispersion ≈ concentration > dispersion.

Behavior-significance score.

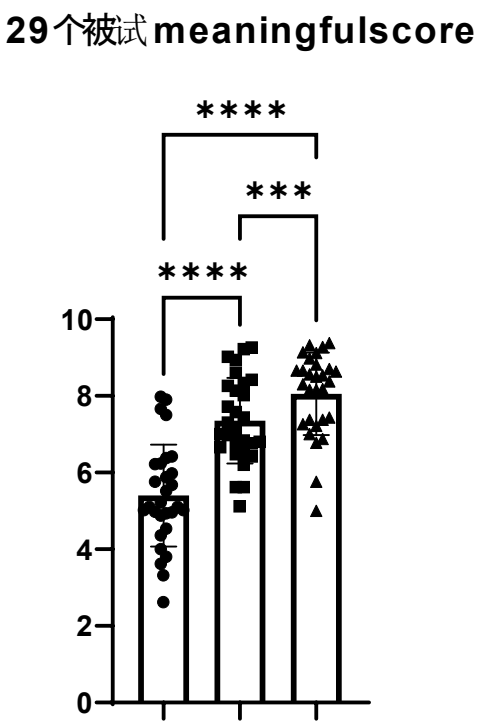


Figure: Meaning score.

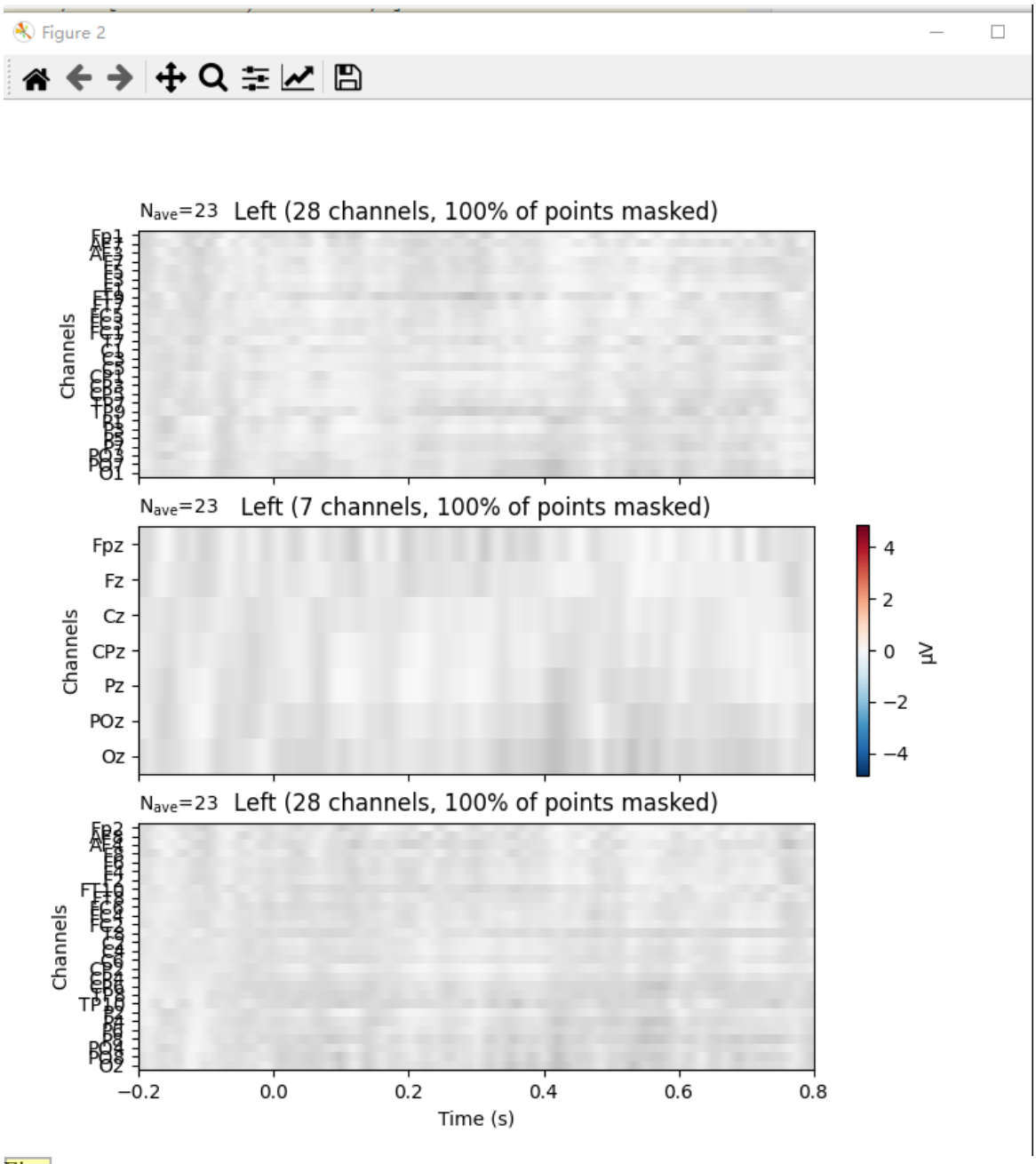
The variance analysis of the logical scores of the three learning styles showed that the main effect was significant, $F=51.25$, $p < 0.0001$;

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

* EEG -erp.

We calculated ERP difference(spaced-once) VS ERP difference (massed - once) and got no significant cluster.

If ERP is used for analysis, the significance difference may be related to the middle and late components (Blanchet et al., 2007; Schleepen et al., 2014), The code from MNE (Gramfort et al., 2013) is used to search for significant cluster of ERP. This code is based on the algorithm idea put forward by Gramfoort et al., aiming at setting a more reasonable threshold to define whether ERP components are significant or not (S. M. Smith & Nichols, 2009). After calculation, no significant cluster is found. It can be seen that although spaced learning brings better memory effect and more meaningful changes, it cannot be reflected by simple ERP amplitude.



Figure

* EEG -TPDS and STPDS

Every subject has three corresponding ERP, taking 100ms as the time window, sliding one

by one, and calculating NPS of once vs massed (red line) and once vs spaced(blue line) in each time window. Spearman correlation coefficient is used for correlation calculation between neural pattern in each time window. Here, Neurora code is borrowed to calculate Dissimalarity. Specifically, we calculated RDM with codes from Neurora (Z. Lu&Ku, 2020b), and then extracted the corresponding neural pattern dissimilarity value from it.(Z. Lu & Ku, 2020b).

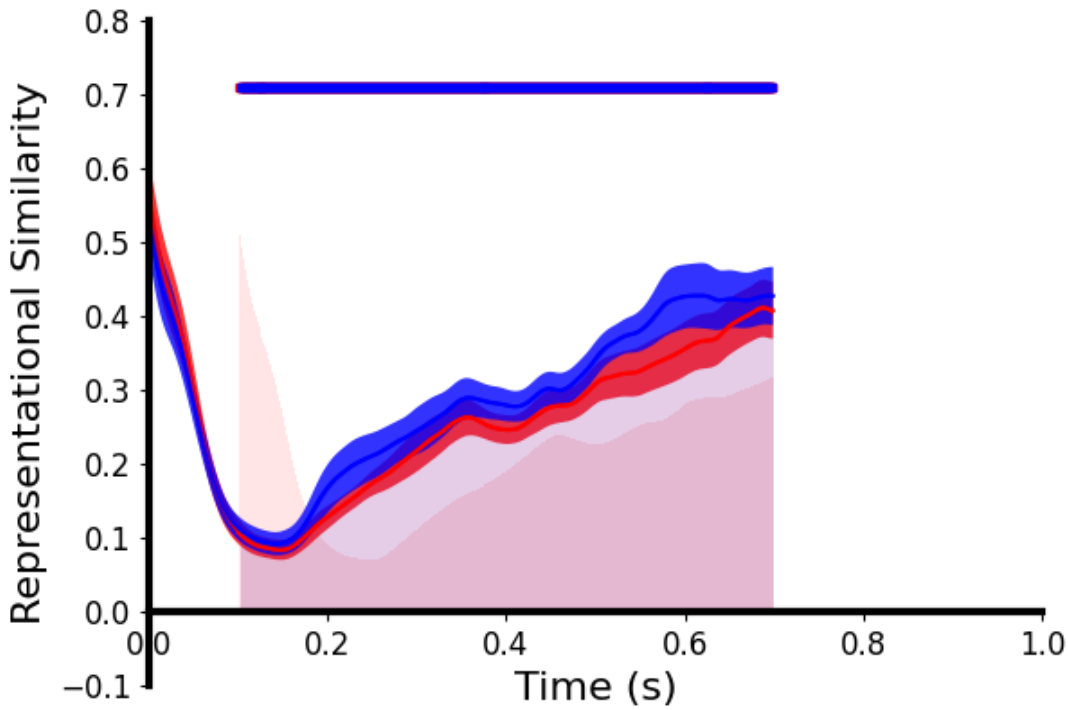
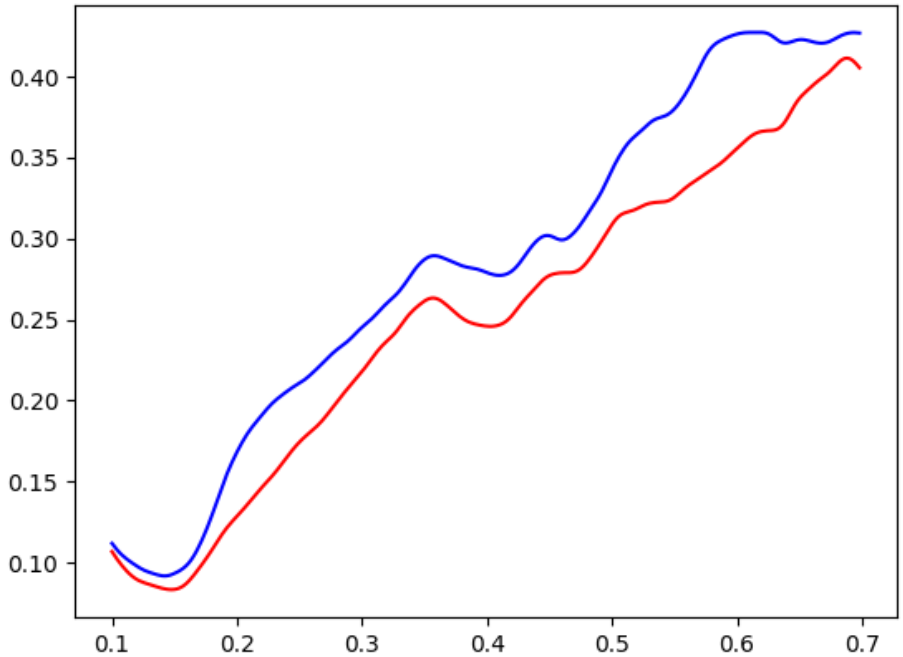


Figure neural pattern dissimilarity.

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 basically always greater than that (concentrated vs once).In order to check whether this difference is significant, STPDS is carried out. Note: The reason why STPS basing on trial comparation is not carried out here is that the word groups used are different in nature, so within-item analysis cannot be achieved.



As time goes by, the difference of representation gradually increases. However, it is still below the random level (using the comparison between the baselines in the first 200ms as the baseline).

We do statistics on neural pattern dissimilarity(NPDS), which calculate pair t-test at each time point and draws a heat map, and the area of the picture framed means significant difference(Z. Lu&Ku, 2020b).(Z. Lu & Ku, 2020b)。

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 extraction of semantic memory in time, which shows that there are differences in the extraction of this part: STPDS(spaced - once) > STPDS(massed - once).

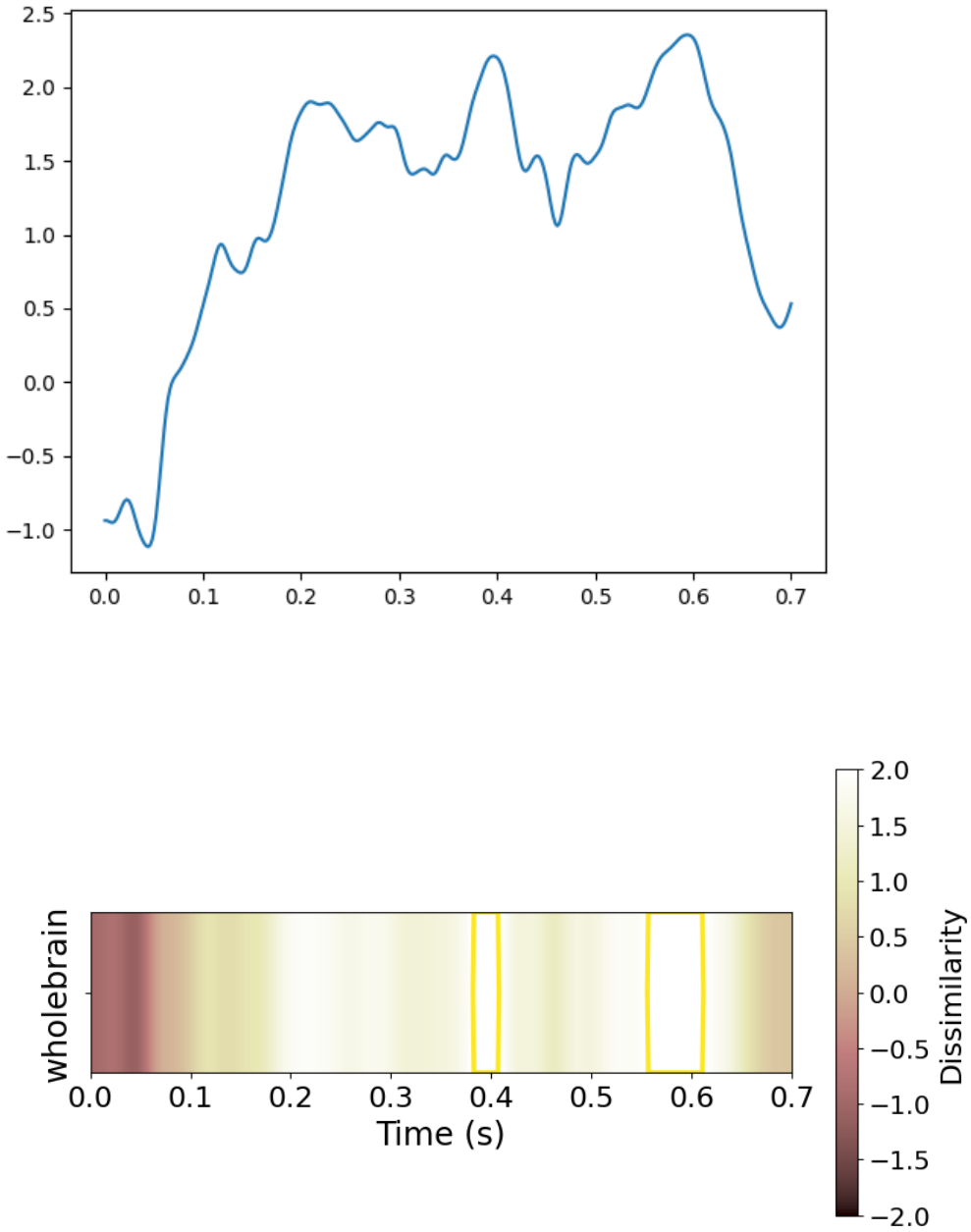


Figure: temporal pattern dissimilarity (TPDS)

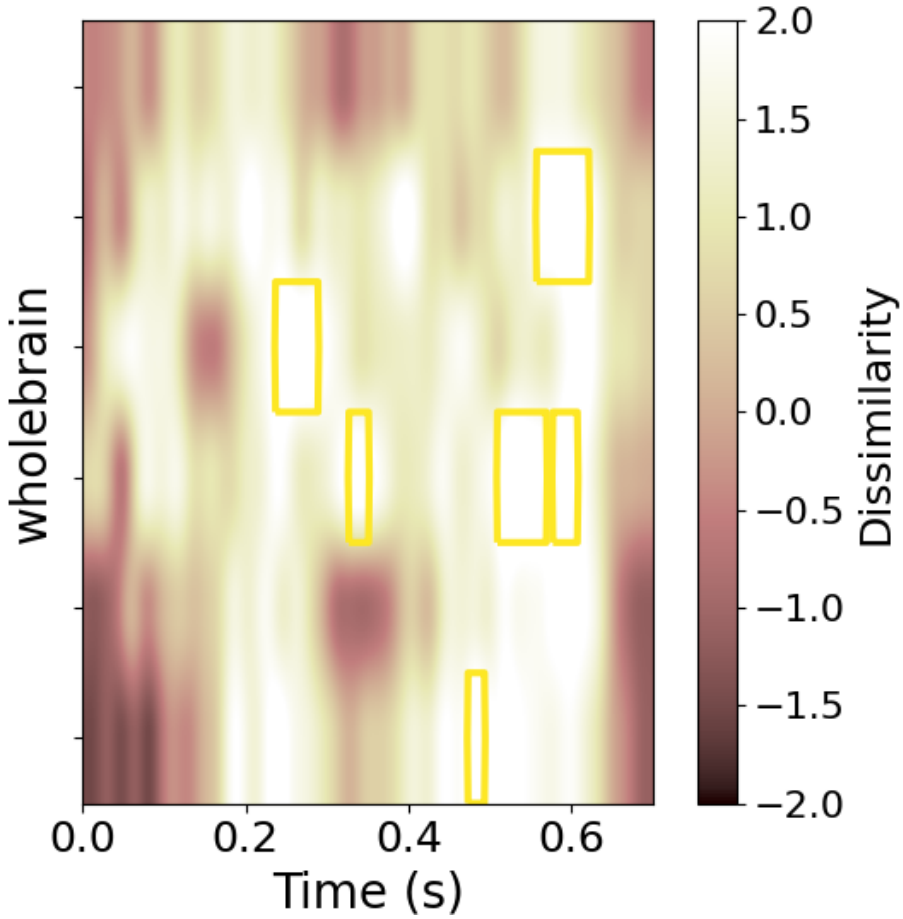
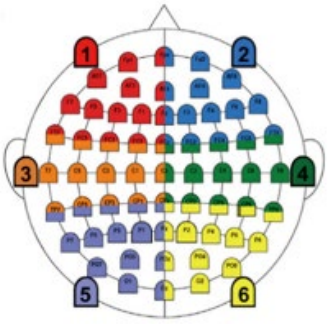


Figure: spatiotemporal pattern dissimilarity (STPDS)



Division of brain regions.

discuss

discuss

Step 1: restate the conclusions and achievements.

The first experiment shows that spaced learning can get a better recall rate, and accordingly, the logical score has also 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 meaningful score has also been improved, which shows that the enhancement of memory strength is related to the improvement of Meaningful Fulleness 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 4, that is, near the parietal lobe, while the late difference mainly comes from the area 2 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 temporal lobe. The difference here shows that the semantic transformation or semantic extraction method of spaced learning is quite different from the results of one-time learning, which is greater than the difference between centralized learning and one-time learning. However, the difference in the late frontal lobe can be guessed to be caused by the least control resources needed in spaced learning.

Step 2: Comparing this result with the previous results, the similarities or differences, it is agreed that the previous results are still improved (combined with the literature). Put theoretical integration and paradigm integration here?

Theoretical renewal.

On the whole, 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 an 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 of all, let's rethink why the similarity of representation of two encoding increases instead of decreases in the case of spaced learning. We believe that's 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? Because 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 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 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 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, 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 increasing of 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 restudy without reminding. In this way, the difference in the encoding stage probably comes from the difference between the two learning methods, even if both cases are successfully recall 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 who regards secondary learning from spaced learning and massed learning as cognitive processing of the same nature, and then think 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 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 secondary learning of spaced learning is more similar to the encoding processing of primary learning, but the memory finally formed has greater changes.

In this way, we put forward another idea of 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 triggered by spaced learning, that is, when. When lag is too short, the difference between two encoding 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, and a considerable part of retrieval can occur, but new information is incorporated, so the learning effect at this time may be the best.

Encoding-Variability Mechanisms try to describe the internal mechanism of effective decentralized learning, that is, how. 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

A spreading activation theory of memory. (1983). *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 261–295. https://doi.org/10.1016/S0022-5371(83)90201-3

Bjork, R., & Bjotk, E. (1992). A new theory of disuse and an old theory of stimulus fluctuation. *Essays in Honor of William K. Estes, Vol. 1991: From Learning*

Theory to Connectionist Theory, 1935–1967.

Blanchet, S., Gagnon, G., & Bastien, C. (2007). Event-related potential study of dynamic neural mechanisms of semantic organizational strategies in verbal learning. *Brain Research*, *1170*, 59–70. https://doi.org/10.1016/j.brainres.2007.07.024

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

Bui, D. C., Maddox, G. B., Zou, F., & Hale, S. S. (2014a). 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

Bui, D. C., Maddox, G. B., Zou, F., & Hale, S. S. (2014b). 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. (1979a). 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

Glenberg, A. M. (1979b). 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

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science (New York, N.Y.)*, *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. (2020a). 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

Lu, Z., & Ku, Y. (2020b). 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

Schleepen, T. M. J., Markus, C. R., & Jonkman, L. M. (2014). Dissociating the effects of semantic grouping and rehearsal strategies on event-related brain

potentials. *International Journal of Psychophysiology*, *94*(3), 319–328. https://doi.org/10.1016/j.ijpsycho.2014.09.007

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. M., & Nichols, T. E. (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. (1975a). 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

Watkins, M. J. (1975b). 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

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.