# Learning guidance based on elimination singularity phenomenon

*Tomohiro Washino**, *Tadashi Takahashi*
*Corresponding Author: d1523001@s.konan-u.ac.jp,
takahasi@konan-u.ac.jp
Graduate School of Natural Science, Konan University
8-9-1 Okamoto, Higashinada, Kobe, Japan

**Abstract**

*In the study of mathematics, when two items have concepts in common, understanding the relationship between them can be subject to "overgeneralization." In this study, a technique introduced in previous research [1] to analyze overlap singularity through a simulation using a neural network on the loss surface is extended to elimination singularity and the effect of overgeneralization.*

*In the elimination singularity examined here, students who answered a series of questions correctly immediately after studying the topic of combinations showed a superficial understanding of the topic, while students who answered the questions partially correctly following the passage of time after learning combinations were affected by overgeneralization. Findings from this analysis are used to formulate learning guidance for teachers of mathematics.*

## 1 Introduction

In learning theory, plateauing is a phenomenon in which learning is affected by singular regions and stagnates. When two concepts, A and B, are learned in the order of A and then B, it is possible that the later-learned B may affect the concept understanding of the earlier-learned A. This is called the "overgeneralization of A as B". In a previous study of overlap singularity, students who gave partially correct answers immediately after learning combinations exhibited a natural way of understanding as part of the developmental learning process, while students who answered partially correctly with the passage of time after learning combinations displayed a deeper understanding [1]. To address overgeneralization stagnation, we consider a singular region due to elimination singularity, defined later as the state where only the group of the two student groups.

## 2 Preparation for analysis

In this study, we used the concept of overgeneralization to investigate changes in student group understanding of permutations and combinations [1]. To consider the loss surface and singularity in learning, we use Kullback information as the error function. Fundamental problems in learning theory have been described in the literature [1], [2], [3]. To test semantic comprehension in the field of permutations and combinations in the unit "number of cases" in the high school mathematics curriculum, we created a set of test questions as follows: (Questions (a) and (c) are permutation problems, question (b) is a combinatorial problem, and question (d) is an interrelationship problem.) (a) "How many ways can three people be taken from a group of nine people and arranged in a row?" (b) "How many ways are there to choose two cards out of seven?" (c) "How many ways can a group of four students arrange themselves in a row?" (d) "How many ways are there to choose one chairman, one vice-chairman, and one secretary from the ten members of a committee? Concurrent posts are not permitted." An examination of scoring and problems of semantic comprehension and interrelationships is reported elsewhere [1].

We define two learning stages : (1) The stage in which partially correct answers are influenced by the study of combinations between the first and second examinations. (2) The stage in which correct answers are obtained by considering the interrelationship with combinations through the re-learning of permutations between the second and third examinations. Three student groups are defined based on questions (a) through (c): (i) Students who achieved full points in the permutation problems and partial points in the combination problem. (ii) Students who achieved partial points in the permutation problems and full points in the combination problem. (iii) Students who achieved full points in both the permutation and combination problems. We compared student groups (i) and (iii) in learning stage (1), and student groups (i) and (ii) in learning stage (2).

**Definition 1** *For inputs $x$, $\theta_0$, the output $Y$ of a two-layer neural network is defined as follows:* $Y := f(x, \theta_0) = w_3 \tanh(w_1 x) + w_4 \tanh(w_2 x)$.

**Definition 2 (Elimination singularity)** *Elimination singularity is defined as passing through the parameter region* $R_1 := \{\theta \in \mathbb{R}^4 | w_3 = 0\} \cup \{\theta \in \mathbb{R}^4 | w_4 = 0\}$.

In this case, the learning model can be expressed as follows: $f(x, \theta) = w_4 \tanh(x, w_2)$, $w_3 \tanh(x, w_1)$. The coordinate transformation from the parameters $\theta = (w_1,\ w_2,\ w_3,\ w_4)$ to the new parameters $\xi = (a,\ b,\ v,\ w)$ can be defined as follows: $a = w_2 - w_1$, $b = \frac{w_3 - w_4}{w_3 + w_4}$, $v = \frac{w_3 w_1 + w_4 w_2}{w_3 + w_4}$, $w = w_3 + w_4$. According to Amari [2], we set $(v, w)$ to the optimal solution $(v^*, w^*)$, and then consider the trajectory of the parameters $(a, b)$ to the optimal solution.

For the two student groups, let c be the average score for the combinatorial problem, d be the average score for the permutation problems, and e and f be the number of students assigned these problems, respectively. The weights of the neural network can then be determined as follows: $w_1' = c$, $w_2' = d$, $w_3 = \frac{e}{e+f}$, $w_4 = \frac{f}{e+f}$. Next, we consider students in each of the three examinations as a subset. In order to keep the parameters $v$, $w$ for the overall set and for the three subsets constant, we applied corrections to the subset weights $w_1'$, $w_2'$. For the two groups of students, the score obtained by subtracting 0.5 from the average of the sum of the permutation and combination scores from the first to the third examinations is the $g$ point. In each examination, the scores obtained by subtracting 0.5 from the average of the sum of the permutations and combinations are $h_1$, $h_2$, $h_3$, respectively. At this time, the correction is determined by $w_i = w_i' \times \frac{g}{h_i}$.

Table 1 shows the average score, number of students, and the correction coefficients for students who gave correct answers (left two columns) and partially correct answers (right two columns) to the interrelationship question (d) in the three examinations.

Table 1: Examination results

| 1st to 3rd examinations | (i) (full, partial) | (iii) (full, full) | (i) (full, partial) | (ii) (partial, full) | 2nd examination | (i) (full, partial) | (iii) (full, full) | (i) (full, partial) | (ii) (partial, full) |
|---|---|---|---|---|---|---|---|---|---|
| Average $c, d$ | 0.228571429 | 0.5 | 0.22826087 | 0.279411765 | Average $c, d$ | 0.25 | 0.5 | 0.25 | 0.276666667 |
| Number of people $e, f$ | 70 | 57 | 23 | 85 | Number of people $e, f$ | 3 | 23 | 8 | 60 |
| Overall average $g$ | 0.350393701 | | 0.268518519 | | Overall average $h_2$ | 0.471153846 | | 0.273529412 | |
| Total number of people | 127 | | 108 | | Total number of people | 26 | | 68 | |
| | | | | | Correction factor $\frac{g}{h_2}$ | 0.743692753 | | 0.981680605 | |
| | | | | | Correction $w_1, w_2$ | 0.185923188 | 0.371846376 | 0.245420151 | 0.271598301 |
| 1st examination | (i) (full, partial) | (iii) (full, full) | (i) (full, partial) | (ii) (partial, full) | 3rd examination | (i) (full, partial) | (iii) (full, full) | (i) (full, partial) | (ii) (partial, full) |
| Average $c, d$ | 0.234126984 | 0.5 | 0.214285714 | 0.245 | Average $c, d$ | 0.125 | 0.5 | 0.21875 | 0.313333333 |
| Number of people $e, f$ | 63 | 5 | 7 | 10 | Number of people $e, f$ | 4 | 29 | 8 | 15 |
| Overall average $h_1$ | 0.253676471 | | 0.232352941 | | Overall average $h_3$ | 0.454545455 | | 0.280434783 | |
| Total number of people | 68 | | 17 | | Total number of people | 33 | | 23 | |
| Correction factor $\frac{g}{h_1}$ | 1.381262125 | | 1.15564932 | | Correction factor $\frac{g}{h_3}$ | 0.770866142 | | 0.957507895 | |
| Correction $w_1, w_2$ | 0.323390735 | 0.690631062 | 0.24763914 | 0.283134083 | Correction $w_1, w_2$ | 0.096358268 | 0.385433071 | 0.209454852 | 0.30001914 |

Taking parameter $a$ as the difference between the average score for the permutation problems and the average score of the combinatorial problem, we considered the balance between the overgeneralization of permutations as combinations and the overgeneralization of combinations as permutations. Parameter $b$ is defined as the ratio between the number of students in the two groups to be considered.

In terms of the two concepts (permutations and combinations), we can visualize the understanding (learning trajectory) of the student group on the learning loss surface. More details on the construction of neural networks and a learning loss surface are given in the literature [1], [4]. Here, the state of understanding is displayed as points for the 1st examination (blue), 2nd examination (red), 3rd examination (green), and 1st to 3rd examinations (yellow). Student groups (i) and (iii) are shown on the left side of Figure 1; student groups (i) and (ii) are shown on the right.

# 3 Singularity phenomena in mathematics education

We considered elimination singularity, near-elimination singularity, and first convergence in information science. Cross-overlap singularity is described elsewhere [4].

We will first examine elimination singularity using the ratio of the number of students in the two student groups to be considered. Here we consider only the group of students who scored full points for the permutation problems ($b = 1$) or the combination problems ($b = -1$).

Next, we consider near-elimination singularity using the ratio of the number of students in the two student groups. This approaches the situation described by $b = 1$ or $b = -1$ above. Finally, we consider fast convergence for permutation and combination questions, whereby the difference between the mean score for students with partially correct permutation answers and that for students with partially correct combination answers is given a partial score. If the difference is large ($a > 0$), it becomes even larger without passing through $a = 0$.

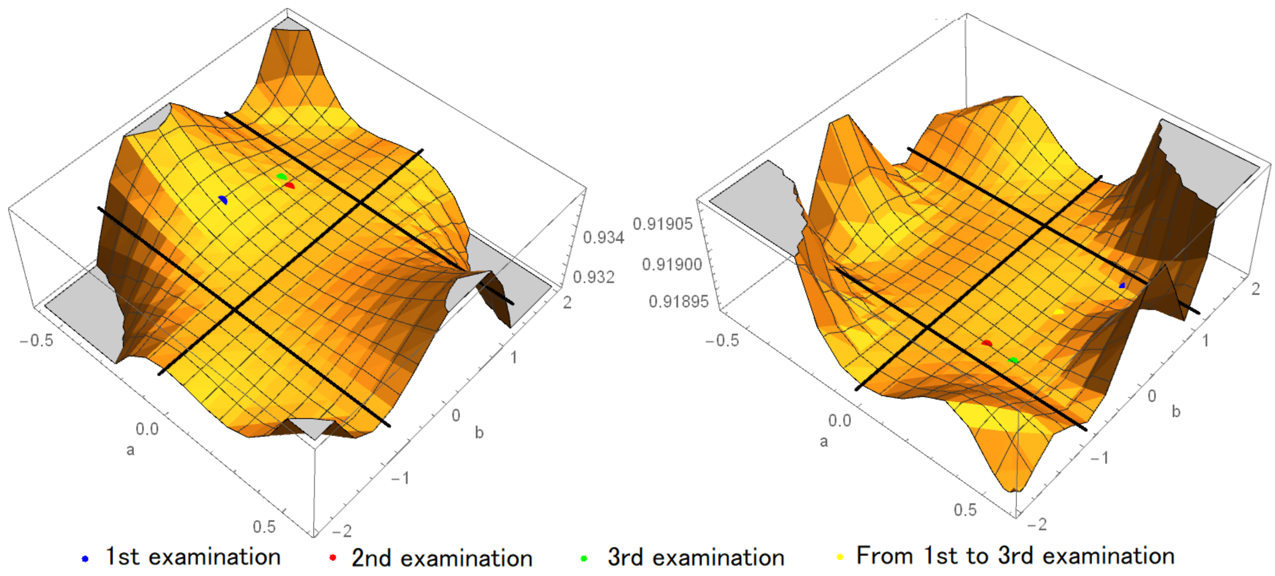


Figure 1: Learning loss surface for semantic understanding

# 4 Simulations

In the following sections, the results obtained from real data are used to perform simulations in which the initial values are changed in 0.05 increments from $-0.6$ to $0.6$ for $a$ and from $-1.1$ to $1.1$ for $b$ [1], [4]. In the figures, the initial values for the first (blue), second (red), and third (green) examinations are indicated by $\circledcirc$, and the true distribution is indicated by $\times$. For the dynamics from the first to the second examination, the initial value (simulation) is shown in blue, and the value after the simulation is shown in red. For the dynamics from the second to the third examination, the initial value (simulation) is shown in red, and the value after the move is shown in green.

## 4.1 Analysis of students who answered correctly with only superficial understanding

Immediately after learning combinations, students answered the interrelationship question (d) using permutations. We analyzed the change from the first to the second examination to target the students who gave correct answers. We now consider the process of understanding the two concepts for students who gave completely correct answers to the permutation problems, partially correct answers to the combination problem, and completely correct answers to both the permutation and the combination problems.

In the following, we consider the case where learning about combinations begins after learning about permutations. At the beginning of the study of combinations, we assume a situation that involves only the group of students who scored full points for the permutation problems, and overgeneralization using combinations as permutations occurs. Since there are few students who fully understand combinations, overgeneralization of permutations as combinations does not occur. Therefore, we consider this unbalanced state to be an elimination singularity.

#### 4.1.1 Simulation for changing $a$

First, we consider the dynamics of the simulation in which $a$ is varied. The evolution of parameters $a$, $b$ on the $a\,b$ plane and the dynamics on the loss surface are shown in Figure 2.

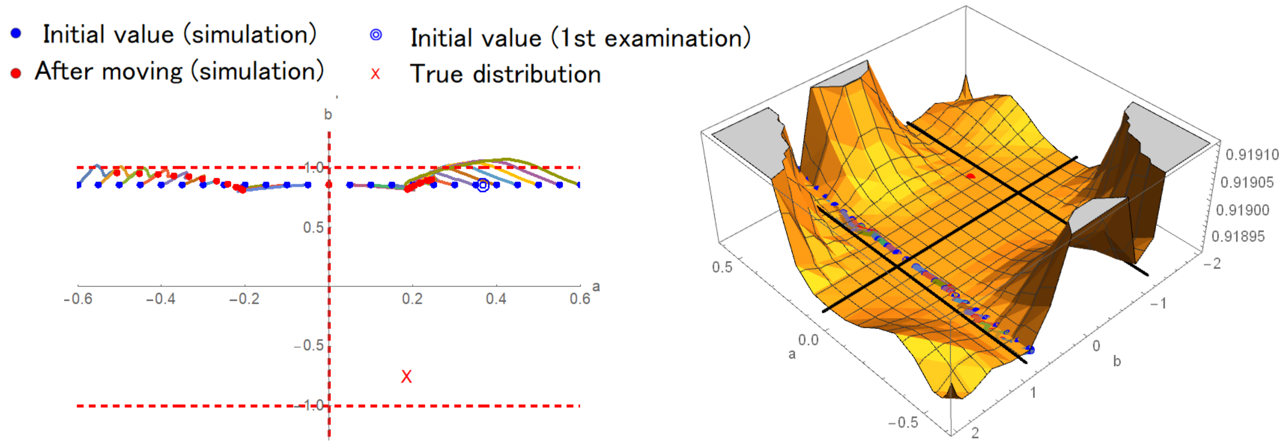


Figure 2: $a$ change: correct answer

The simulation results and transformation for the schema are shown in Table ?? .

Table 2: Simulation results and schema transformation

| Simulation | Result | Schema transformation |
|---|---|---|
| Overgeneralization of combinations as permutations increases | Only the group of students who scored full points for the permutation problems | Permutations |
| Overgeneralization of permutations as combinations increases | Proportion of students does not change | None |

Learning begins from the state $b = 0.85$, where the proportion of students who answer the permutation questions correctly and the combination questions partially correctly is larger and more unbalanced than the proportion of the group of students who answer both the permutation and combination questions correctly.

When the overgeneralization of combinations as permutations increases ($a$ decreases), the critical line $b = 1$ is affected, and elimination singularity occurs. Only students who answer the permutation questions correctly and the combination questions partially correctly are included; there are no students who answer both the permutation and combination questions correctly. If the overgeneralization with combinations as permutations is smaller than $a < -0.3$, the learning of combinations does not progress and a transformation of the permutation schema takes place.

As overgeneralization of permutations as combinations increases ($a$ increases), near-elimination singularity occurs due to an approach to the critical line $b = 1$ and a return to the original state, stagnating at $a = 0.3$. Since the ratio of students who answer the permutation questions correctly and the combination questions partially correctly to students who answer both the permutation and the combination problems correctly does not change from the initial state, the learning of combinations does not progress, and a transformation of the permutation schema does not occur.

### 4.1.2 Simulation for changing $b$

Next, we consider the dynamics of the simulation when $b$ is varied. The evolution of the parameters $a$, $b$ on the $a\,b$ plane and the dynamics on the loss surface are shown in Figure 3.

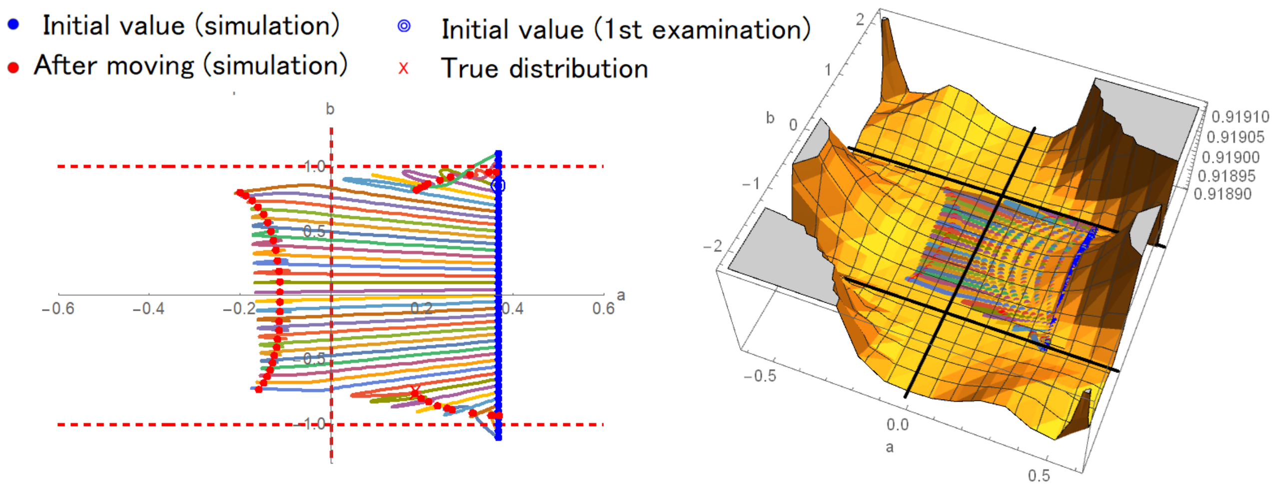


Figure 3: $b$ change: correct answer

The simulation results and transfer are shown in Table 3.

Table 3: Simulation results and transfer

| Simulation | Result | Transfer |
|---|---|---|
| Increase the percentage of students who answer the permutation question correctly and suppress overgeneralization of permutations as combinations | Positive transfer | |
| Increase the percentage of students who answer both the permutation and combination questions correctly | Overgeneralization of combinations as permutations becomes larger | Negative transfer |

Learning starts from the state $a = 0.36$, where the overgeneralization of permutations as combinations is large.

At first, as the proportion of students who answer the permutation questions correctly and the combination questions partially correctly becomes larger ($b$ increases), a near-elimination singularity occurs due to an approach to the critical line $b = 1$ and a return to the original state. A positive transition occurs because the large overgeneralization of permutations as combinations is suppressed.

As the proportion of students who answer both the permutation and combination questions correctly becomes larger ($b$ decreases), a cross-overlap singularity occurs beyond the critical line $a = 0$. The state of overgeneralization of permutations as combinations is large, $w_1$, $w_2$ becomes 0.35 from $v = 0.35$, and there is no difference from the overgeneralization of combinations as permutations. Furthermore, the overgeneralization of combinations as permutations becomes large, which leads to a negative transition.

### 4.1.3 Examination of teacher's teaching strategy

By simulating the change in $a$, $b$, the transformation of the schema and transfer can be assumed, allowing the teacher to devise an effective teaching strategy.

By increasing the overgeneralization of combinations as permutations from the change in $a$, the schema of the permutation changes, but the learning of combinations does not progress. From the $b$ change, increasing the proportion of students who answer both the permutation and combination questions correctly increases the overgeneralization of combinations as permutations, so that a negative transition occurs even as the learning of permutations progresses.

Although the problem is answered correctly (absent the influence of "choosing", students think in terms of permutations), it is not affected by the overgeneralization of permutations as combinations due to an insufficient understanding of combinations. This can be considered a superficial understanding of the developmental process.

Guidance in combinations is necessary. Furthermore, it is necessary to provide guidance for deepening understanding by considering the interrelationships between permutations and combinations. Students need to be taught to make correct judgments in problems that require understanding the interrelationships between permutations and combinations.

## 4.2 Analysis of students who give partially correct answers due to overgeneralization

In the interrelationship problem (d), after learning permutations, the student who was correct by using permutations was partially correct by mistakenly using combinations after some time had passed since learning permutations. Students who scored partial points for the permutation problems and full points for the combination problem were compared with students who scored full points for the permutation problems and partial points for the combination problem. Consider the process of understanding the concept. After learning combinations, time passes. The case in which it is time to start re-learning permutations is considered below:

Assume a situation in which, after learning combinations and time passes, there is only the group of students who scored full points for the combination problems, and overgeneralization using permutation as combinations occurs. Since there are no students who fully understand permutations, overgeneralization of combinations as permutations does not occur. Therefore, we consider this unbalanced state to be an elimination singularity.

#### 4.2.1 Simulation for changing $a$

First, we consider the dynamics of the simulation in which $a$ is varied. The evolution of parameters $a$, $b$ on the $a\,b$ plane and the dynamics on the loss surface are shown in Figure 4.

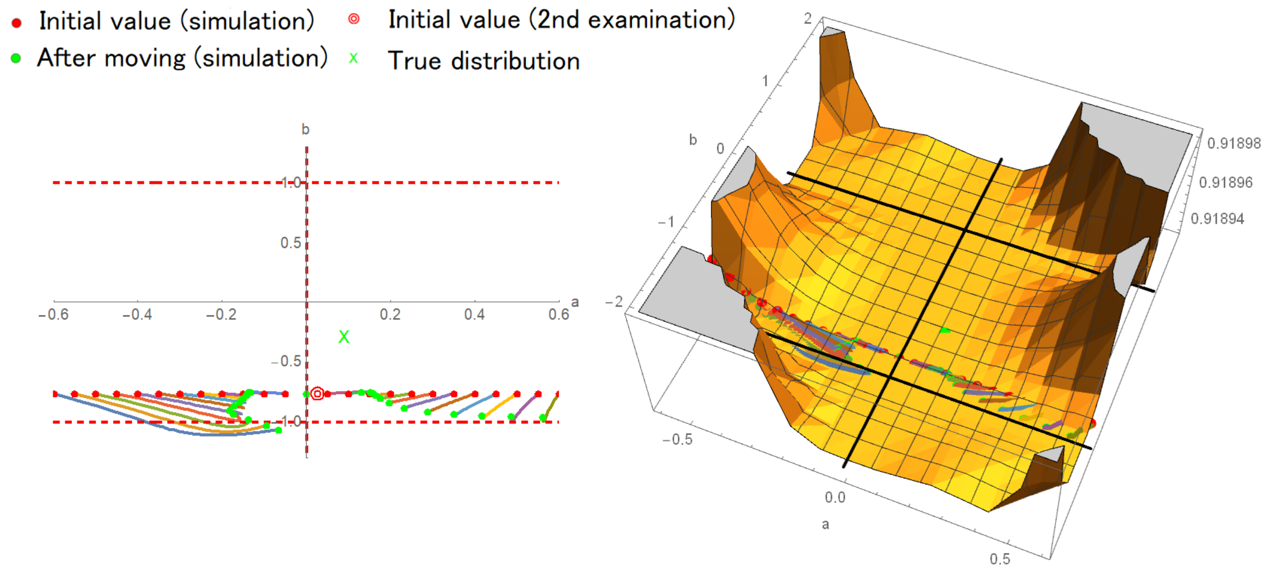


Figure 4: $a$ change: partially correct answers

The simulation results and transformation of the schema are shown in Table 4.

Table 4: Simulation results and schema transformation

| Simulation | Result | Schema transformation |
|---|---|---|
| Overgeneralization of permutations as combinations increases | Only the group of students who scored full points for the combination problems | Combination |
| Overgeneralization of combinations as permutations increases | Proportion of students does not change | None |

Learning begins from the state $b = -0.76$, where the proportion of students who answer the combination questions correctly and the permutation questions partially correctly is larger and more unbalanced than the proportion of students who answer the permutation questions correctly and the combination questions partially correctly.

When the overgeneralization of permutations as combinations increases ($a$ increases), the critical line $b = -1$ is affected, and an elimination singularity occurs. There are only students who answer the combination questions correctly and the permutation questions partially correctly. There are no students who answer the permutation questions correctly and the combination questions partially correctly. If the overgeneralization with permutations as combinations is larger than $a > 0.3$, the learning of permutations does not progress and a transformation of the combinations schema takes place.

As the overgeneralization of combinations as permutations increases ($a$ decreases), a near-elimination singularity occurs due to an approach to the critical line $b = -1$ and a return to the original state, stagnating at $a = -0.15$. Since the ratio of students who answer the combination questions correctly and the permutation questions partially correctly to the students who answer the permutation questions correctly and the combination questions partially correctly does not change from the initial state, the learning of combinations does not progress, and a transformation of the permutation schema does not occur.

### 4.2.2 Simulations for $b$ change

Next, we consider the dynamics of the simulation when $b$ is varied. The evolution of parameters $a$, $b$ on the $a\,b$ plane and the dynamics on the loss surface are shown in Figure 5.

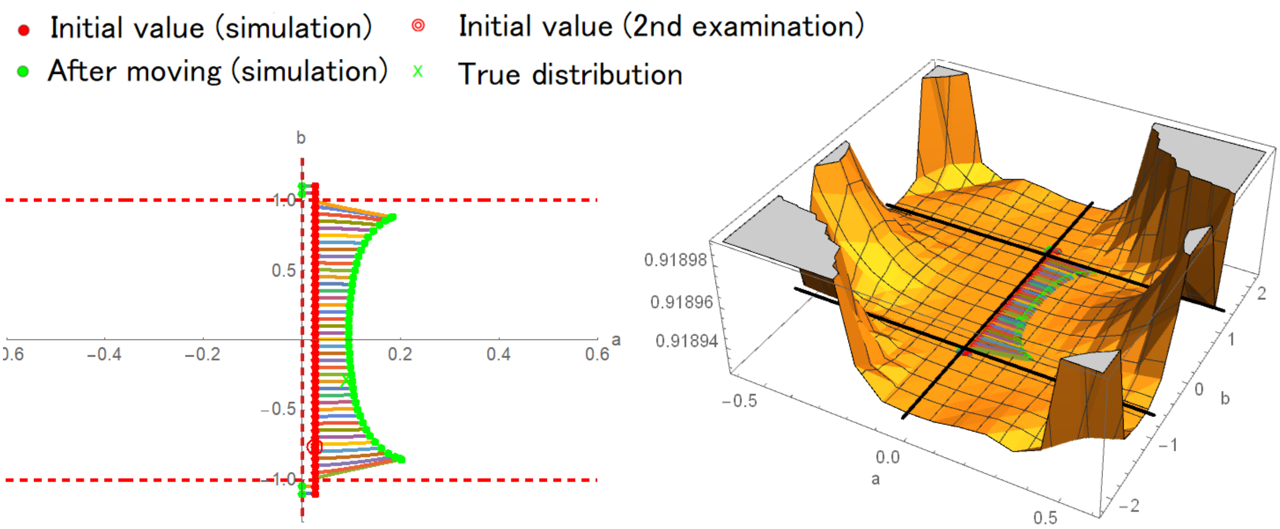


Figure 5: $b$ change: partially correct answers

The simulation results and transfer are shown in Table 5.

Table 5: Simulation results and transfer

| Simulation | Results | Transfer |
|---|---|---|
| Increase the proportion of students who score full points for the permutation problems and suppress large overgeneralization of permutations as combinations slightly | Negative transfer | |
| Increase the proportion of students who score full points for the combination problems | A large overgeneralization of combinations as permutations becomes even larger | Negative transfer |

Here, we start learning from $a = 0.026$, a state in which the overgeneralization of permutations as combinations is highly biased.

As the percentage of students who score partial points for the permutation problems and full points for the combination problem increases ($b$ decreases), fast convergence occurs without the influence of the critical line $a = 0$. The difference from the overgeneralization of permutations as combinations is even greater, so that a negative transition occurs.

When the proportion of students who score full points for the permutation problems and partial points for the combination problem increases ($b$ increases), fast convergence occurs without the influence of the critical line $a = 0$. The large overgeneralization of permutations as combinations is suppressed, so that a negative transition occurs.

By relearning permutations, the overgeneralization of permutations as combinations can be suppressed from $a = 0.18$ to $a = 0.094$.

### 4.2.3 Direction of teacher guidance

By simulating the change in $a$, $b$, the transformation of the schema and transfer can be assumed, allowing the teacher to devise a teaching strategy.

By increasing the overgeneralization of permutations as a combination from the change in $a$, the schema of the combinations changes, but the learning of permutations does not progress. From the $b$ change, increasing the proportion of students who score full points for the permutation problems increases the overgeneralization of permutations as combinations, so that a negative transition occurs even as the re-learning of permutations progresses, but the overgeneralization of permutations as combinations can be slightly suppressed.

Although the problem is answered partially correctly (under the influence of "choosing," students think in terms of combinations), it is greatly affected by the overgeneralization of permutations as combinations. This can be considered a partially correct answer affected by overgeneralization. Therefore, it is considered necessary to re-learn permutations, paying attention to understanding permutations even after learning combinations.

# References


[1] Washino T., and Ohashi S. (2023), Learning guidance based on the overlap singularity phenomenon, Scientiae Mathematicae Japonicae(to appear).

[2] S. Watanabe (2006), Algebraic geometry and statistical learning theory, Morikita Publishing Co., Ltd.

[3] S. Amari (2014), New developments in information geometry, Saiensu-sya Co., Ltd.

[4] Washino T., and Takahashi T. (2021), On the analysis of singularity structure in learning, Proc. of the 16th Asian Technology Conference in Mathematics(https://atcm.mathandtech.org/EP2021/regular/21892.pdf).