EARLY START CAN INHIBIT LEARNING: A GEOMETRIC EXPLANATION

Olga Kosheleva
Department of Teacher Education
University of Texas, El Paso, TX 79968, USA
olgak@utep.edu

Early start: a seemingly natural idea. The age at which we teach different topics change. If it turns out that students do not learn, say, reading by the time they should, a natural idea is to start teaching them earlier.

Several decades ago, reading and writing started in the first grade, now they start at kindergarten and even earlier. At first glance, the earlier we start, the better the students will learn.

Early start: known side effects. With the early start, children may play less, their childhood may be not as careless as it used to be – but a usual expectation is that with an early start, children will learn better.

Early start: serious problems. In practice, however, early start does not always help: often, early start inhibits learning.

For example, according to (Papousek and Papousek 1977), human infants who started learning to turn their heads to specific sounds at the age of 31 days mastered this task, on average, at the age of 71 days, while infants who started learning this task at birth mastered this task, on average, at the age of 128 days.

This phenomenon is not limited to human infants: according to (Harlow 1959), an early start in training rhesus monkeys to discriminate objects decreased their peak performance level.

Numerous examples when an early start inhibits learning are presented and discussed in (Bjorklund and Green 1992), (Bjorklund and Pellegrini 2002), (Ellis and Bjorklund 2005), (Bjorklund 2007), (Remmel 2008).

Natural questions. The empirical fact that an early start often inhibits learning leads to the following natural question: how do take this phenomenon into account when enhancing student learning?

To be able to take this phenomenon into account in the learning process, we must be able to understand this phenomenon – and ideally, understand on the quantitative level.

These questions are still largely open. In (Bjorklund and Green 1992), (Bjorklund and Pellegrini 2002), (Ellis and Bjorklund 2005), (Bjorklund 2007), an attempt is made to understand why early start can inhibit learning. However, the existing understanding is still mostly on the qualitative level, and even on this level, the proposed explanations are still not fully satisfactory; see, e.g., (Remmel 2008).

More general questions. The above questions about the efficiency of the early start can be viewed as a particular case of more general questions: what is the best order of presenting the material, the order that leads to the best possible learning?

These more general questions are often very important. Many empirical studies have shown that a change in the order in which different parts of the material are presented often drastically changes the learning efficiency; see, e.g., (Van Patten et al. 1986), (Davydov 1990), (Tchoshanov 1997), (Paper and Tchoshanov 2001), (Lesser and Tchoshanov 2005), (Lesser and Tchoshanov 2006), (Kaminski et al. 2006), (Kaminski et al. 2008).

This is not only about using common sense: sometimes, the empirical results are counter-intuitive. For example: it is usually assumed that most students learn mathematical concepts better if they are first presented with concrete examples of these concepts, and they only learn abstract ideas later on. However, it turns out that empirically, the abstract-first approach for presenting the material often enhances learning; see, e.g., (Tchoshanov 1997), (Lesser and Tchoshanov 2005), (Lesser and Tchoshanov 2006), (Kaminski et al. 2006), (Kaminski et al. 2008), (Kaminski et al. 2006), (Kaminski at el. 2008).

What we do in this paper. In this paper, we attempt: to explain the negative effect of early start and, more generally, to explain the reasons why a change in presentation order can drastically change the efficiency of learning. We then show how this explanation can be used to avoid inhibition of learning – and to enhance the student learning.

Learning: a natural geometric representation. To facilitate reasoning about learning, let us start with a simple geometric representation of learning.

The process of learning means that we change the state of a student: from a state in which the student did not know the material (or does not have the required skill) to a state in which the student has (some) knowledge of the required material (or has the required skill).

Let s_0 denote the original state of a student, and let S denote the set of all the states corresponding to the required knowledge or skill. We start with a state which is not in the set S ($s_0 \notin S$), and we end up in a state s which is in the set S.

On the set of all possible states, it is natural to define a metric $d(s, s')$ as the difficulty (time, effort, etc.) needed to go from state s to state s'. Our objective is to help the students learn in the easiest (fastest, etc.) way. In terms of the metric d, this means that we want to go from the original state $s_0 \notin S$ to the state $s \in S$ for which the effort $d(s_0, s)$ is the smallest possible.

In geometric terms, the smallest possible effort means the shortest possible distance. Thus, our objective is to find the state $s \in S$ which is the closest to s_0. Such closest state is called the *projection* of the original state s_0 on the set S.

Learning complex material: geometric interpretation. The above geometric description of learning as a transition from the original state s_0 to its projection on the desired set S describes learning *as a whole*. Our objective is to find out which order of presenting information is the best. Thus, our objective is to analyze the *process* of learning, i.e., learning as a multi-stage phenomenon. For this analysis, we must explicitly take into account that the material to be learned consists of several pieces.

Let S_i, $1 \le i \le n$, denote the set of states in which a student has learned the i-th part of the material. Our ultimate objective is to make sure that the student learns all the parts of the material. In terms of states, learning the i-th part of the material means belonging to the set S_i. Thus, in terms of states, our objective means that the student should end up in a state which belongs to all the sets $S_1, \ldots, S_n$ – i.e., in other words, in a state which belongs to the intersection $S \stackrel{\text{def}}{=} S_1 \cap \ldots \cap S_n$ of the corresponding sets S_i.

In these terms, if we present the material in the order S_1, S_2, ..., S_n, this means that:

- we first project the original state s_0 onto the set S_1, resulting is a state $s_1 \in S_1$ which is the closest to s_0;
- then, we project the state s_1 onto the set S_2, resulting is a state $s_2 \in S_2$ which is the closest to s_1;
- ...
- at the last stage of the cycle, we project the state s_{n-1} onto the set S_n, resulting is a state $s_n \in S_n$ which is the closest to s_{n-1}.

In some cases, we end up learning all the material – i.e., in a state $s_n \in S_1 \cap \ldots \cap S_n$. However, often, by the time the students have learned S_n, they have somewhat forgotten the material that they learned in the beginning. So, it is necessary to repeat this material again (and again). Thus, starting from the state s_n, we again sequentially project onto the sets S_1, S_2, etc.

The above geometric interpretation makes computational sense. The above "sequential projections" algorithm is actually actively used in many applications; see, e.g., (Gubin et al. 1967), (Stark and Yang 1998), (Kontoghiorghes 2006). In the case when all the sets S_i are convex, the resulting Projections on Convex Sets (POCS) method actually guarantees (under certain reasonable conditions) that the corresponding projections converge to a point from the intersection $S_1 \cap \ldots \cap S_n$ – i.e., in our terms, that the students will eventually learn all parts of the necessary material.

In the more general non-convex case, the convergence is not always guaranteed – but the method is still efficiently used, and often converges.

The simplest case: two-part knowledge. Let us start with the simplest case when knowledge consists of two parts. In this simplest case, there are only two options.

The first option is that we begin by studying S_1; then, we study S_2, then, if needed, we study S_1 again, etc. The second option is that: we begin by studying S_2; then, we study S_1, then, if needed, we study S_2 again, etc.

We want to get from the original state s_0 to the state $\widetilde{s} \in S_1 \cap S_2$ which is the closest to s_0. The effectiveness of learning is determined by how close we get to the desired set $S = S_2 \cap S_2$ in a given number of iterations.

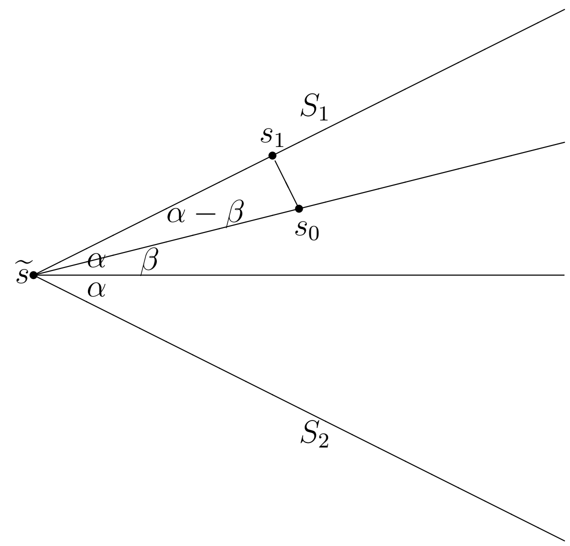
In the case of two-part knowledge, it is natural to conclude that the amount of this knowledge is reasonably small – otherwise, we would have divided into a larger number of easier-to-learn pieces.

In geometric terms, this means that the original state s_0 is close to the desired intersection set $S_1 \cap S_2$, i.e., that the distance $d_0 \stackrel{\text{def}}{=} d(s_0, \widetilde{s})$ is reasonably small.

Since all the states are close to each other, in the vicinity of the state $\widetilde{s}$, we can therefore expand the formulas describing the borders of the sets S_i into Taylor series and keep only terms which are linear in the (coordinates of the) difference $s - \widetilde{s}$. Thus, it is reasonable to assume that the border of each of the two sets S_i is described by a linear equation – and is hence a (hyper-)plane: a line in 2-D space, a plane in 3-D space, etc.

As a result, we arrive at the following configuration. Let 2α denote the angle between the borders of the sets S_1 and S_2, so that the angles between each of these borders and the midline is exactly α. Let β denote the angle between the direction from $\widetilde{s}$ to s_0 and the midline. In this case, the angle between the border of S_1 and the midline is equal to $\alpha - \beta$.

In the first option, we first project s_0 onto the set S_1. As a result, we get the following configuration:



Here, the projection line $s_0 s_1$ is orthogonal to the border of S_1. From the right triangle $\triangle\, \widetilde{s} s_0 s_1$, we therefore conclude that the distance $d_1 \stackrel{\text{def}}{=} d(\widetilde{s}, s_1)$ from the projection point s_1 to the desired point $\widetilde{s}$ is equal to $d_1 = d_0 \cdot \cos(\alpha - \beta)$.

On the next step, we project the point s_1 from S_1 onto the line S_2 which is located at the angle 2α from S_1. Thus, for the projection result s_2, we will have

$$d_2 = d(s_2, \widetilde{s}) = d_1 \cdot \cos(2\alpha) = d_0 \cdot \cos(\alpha - \beta) \cdot \cos(2\alpha). \quad (1)$$

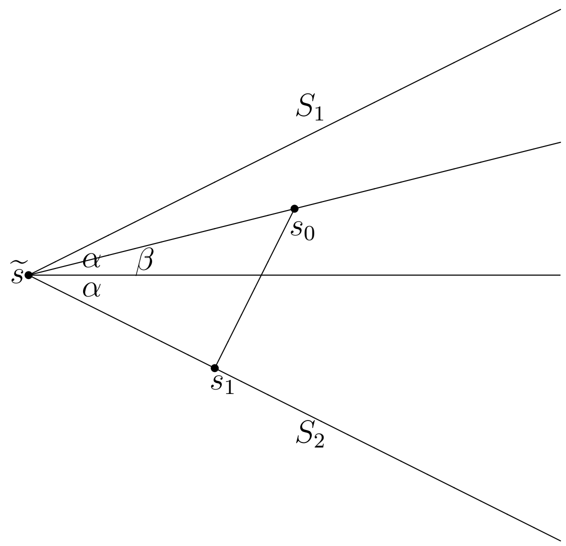
After this, we may again project onto S_2, then again project onto S_1, etc. For each of these projections, the angle is equal to 2α, so after each of them, the distance from the desired point $\widetilde{s}$ is multiplied the same factor $\cos(2\alpha)$.

As a result, after k projection steps, we get a point s_k at a distance

$$d_k = d(s_k, \widetilde{s}) = d_0 \cdot \cos(\alpha - \beta) \cdot \cos^{k-1}(2\alpha) \quad (2)$$

from the desired state $\widetilde{s}$.

In the second option, we start with teaching S_2, i.e., if we first project the state s_0 into the set S_2. In this option, we get the following configuration:



Here, we have $d_1 = d_0 \cdot \cos(\alpha + \beta)$,

$$d_2 = d(s_2, \widetilde{s}) = d_1 \cdot \cos(2\alpha) = d_0 \cdot \cos(\alpha + \beta) \cdot \cos(2\alpha), \quad (3)$$

$$\ldots$$

$$d_k = d(s_k, \widetilde{s}) = d_0 \cdot \cos(\alpha + \beta) \cdot \cos^{k-1}(2\alpha). \quad (4)$$

Since, in general, $\cos(\alpha - \beta) \neq \cos(\alpha + \beta)$, we can see that a change in the presentation order can indeed drastically change the success of the learning procedure.

Conclusion: dependence explained. Thus, our simple geometric model explains why the effectiveness of learning depends on the order in which the material is presented.

Towards specific recommendations. Let us extract more specific recommendations from our model. According to the above formulas, starting with S_1 leads to a more effective learning than starting with S_2 if and only if

$$d_0 \cdot \cos(\alpha - \beta) \cdot \cos^{k-1}(2\alpha) < d_0 \cdot \cos(\alpha + \beta) \cdot \cos^{k-1}(2\alpha), \quad (5)$$

i.e., equivalently, if and only if $\cos(\alpha - \beta) < \cos(\alpha + \beta)$. Since for the angles $x \in [0, \pi]$, the cosine $\cos(x)$ is a decreasing function, we conclude that projection of S_1 is better if and only if $\alpha - \beta > \alpha + \beta$. Thus, we arrive at the following recommendation:

Recommendation. To make learning more efficient, we should start with studying the material which is further away from the current state of knowledge. In other words, we should start with a material that we know the least.

This ties in nicely with a natural commonsense recommendation that to perfect oneself, one should concentrate on one's deficiencies.

This recommendation is also in a very good accordance: with the seemingly counter-intuitive conclusion from (Tchoshanov 1997), (Lesser and Tchoshanov 2006), (Kaminski et al. 2006), (Kaminski et al. 2008), that studying more difficult (abstract) ideas first enhances learning, and with the human infant studies (Papousek and Papousek 1977) according to which a concentration on teaching, to human infants, skills that they can easily learn is detrimental in the long run.

General case: analysis of the problem. What happens in the general case, when instead of only two knowledge components, we have a large number of different components? In the beginning, it still makes sense to project to the set S_{i_1} which is the farthest from the original state s_0.

After this original projection, in the general case, we still have a choice. We can project to any set S_{i_2}, $i_2 \neq i_1$, in which case the current distance d_1 to the desired state is multiplied by the cosine $\cos(\alpha_{i_1 i_2})$ of the angle between the corresponding sets S_{i_1} and S_{i_2}. After k steps, we get the original distance multiplied by the product of the corresponding cosines.

Our objective is to find the best order, i.e., the sequence S_{i_1}, S_{i_2}, $\ldots$, S_{i_n} that covers all n sets $S_1, \ldots, S_n$ and for which the corresponding product

$$\cos(\alpha_{i_1 i_2}) \cdot \cos(\alpha_{i_1 i_2}) \cdot \ldots \cdot \cos(\alpha_{i_n i_1}) \tag{6}$$

attains the smallest possible value.

Usually, it is easier to deal with the sums than with the products. To transform the product into a sum, we can use the fact that minimizing the product is equivalent to minimizing its logarithm, and the logarithm of the product is equal to the sum of the logarithms. Thus, minimizing the product (6) is equivalent to minimizing the sum $D(i_1, i_2) + \ldots + D(i_n, i_1)$, where

$$D(i,j) \stackrel{\text{def}}{=} \log(\cos(\alpha_{ij})). \tag{7}$$

In other words, we arrive at the following problem. We have n objects with known distances $D(i,j),\ 1 \le i,j \le n$. We must find a way to traverse all the objects and come back in such a way that the overall traveled distance is the smallest possible.

This is a well-known problem called a *traveling salesman* problem. It is known that in general, this problem is NP-hard (see, e.g., (Papadimitriou 1994), and in many cases, there exist reasonable algorithms for solving this problem; see, e.g., (Applegate et al. 2006).

Recommendations: general case. To find the optimal order of presenting the material, we must solve the corresponding instance of the traveling salesman problem, with the distances determined by the formula (7).

Acknowledgment. The author is thankful to Mourat Tchoshanov for valuable discussions.

References

D. L. Applegate, R. M. Bixby, V. Chvátal, and W. J. Cook, *The Traveling Salesman Problem*, Princeton University Press, Princeton, New Jersey, 2006.

D. F. Bjorklund, *Why Youth is Not Wasted on the Young: Immaturity in Human Development*, Blackwell Publishing, 2007.

D. F. Bjorklund and B. L. Green, The adaptive nature of cognitive immaturity, *American Psychologist*, 47:46–54, 1992.

D. F. Bjorklund and A. Pellegrini, *The Origins of Human Nature: Evolutionary Developmental Psychology*, American Psychological Association, Washington, DC, 2002.

V. Davydov, *Types of Generalizations in Instruction: Logical and Psychological Problems in the Structuring of Schhol Curricular*, NCTM, Reston, Virginia, 1990.

B. J. Ellis and D. F. Bjorklund (eds.), *Origins of the Social Mind: Evolutionary Psychology and Child Development*, Guilford Press, New York, 2005.

L. G. Gubin, B. T. Polyak, and E. V. Raik, The method of projections for finding the comon point of convex sets, *USSR Computational Mathematics and Mathematical Physics*, 7:1–24, 1967.

H. F. Harlow, The development of learning in the rhesus monkey, *American Scientist*, 47:459–479, 1959.

J. A. Kaminski, V. M. Sloutsky, and A. F. Heckler, Do children need concrete instantiations to learn an abstract concept?, In: R. Sun and N. Miyake (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society, Vancouver, BC, July 26–29, 2006*, Lawrence Erlbaum, Mahwah, NJ, 2006, pp. 411–416.

J. A. Kaminski, V. M. Sloutsky, and A. F. Heckler, The advantage of abstract examples in learning math, *Science*, 320:454–455, April 25, 2008.

E. J. Kontoghiorghes, *Handbook of Parallel Computing and Statistics*, CRC Press, Boca Raton, Florida, 2006.

L. M. Lesser and M. A. Tchoshanov, The effect of representation and representational sequence on students' understanding, In: G. M. Lloyd, M. Wilson, J. M. L. Wilkins, and S. L. Behm (eds.), *Proceedings of the 27th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education PME-NA'2005*, Roanoke, Virginia, October 20–23, 2005.

L. M. Lesser and M. A. Tchoshanov, Selecting representations, *Texas Mathematics Teacher*, 53(2):20–26, Fall 2006.

S. Paper and M. A. Tchoshanov, The role of representation(s) in developing mathematical understanding, *Theory into Practice*, 40(2):118–127, 2001.

C. H. Papadimitriou, *Computational Complexity*, Addison Wesley, 1994.

H. Papousek and M. Papousek, Mothering and the cognitive head-start: Psychobiological considerations. In H. R. Schaffer (Ed.), *Studies in Mother-Infant Interaction*, Academic Press, London, 1977.

E. Remmel, The benefits of a long childhood, *American Scien-*

tist, 96(3):250–251, May-June 2008.

H. Stark and Y. Yang, *Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets and Optics*, Wiley, New York, 1988.

M. A. Tchoshanov, *Visual Mathematics*, ABAK Publ., Kazan, 1997 (in Russian).

J. Van Patten, C. Chao, and C. Reigeluth, A review of strategies for sequencing and synthesizing instruction, *Review of Educational Research*, 56(4):437–471, 1986.