A Model of Knowledge Interaction

Sydney Zheng

May 2020

Abstract

Concepts in knowledge are inextricably interrelated. Not only do we see this indicated in previous studies in psychology and artificial intelligence, we also see this pattern manifest again and again when one field inspires another seemingly vastly different field, when human problems such as climate change and poverty and civil rights are shown to be intertwined, when situations in our daily life better inform our understanding of other people's situations. Here we investigate through a model how already interrelated knowledge affects how we interact with knowledge in the future. More concretely, we examine the best way an individual can interact with the environment such that they learn the most about a particular concept. We show that by creating an environment where the person learns a diversity of other concepts, with repetitions of learning the concept in question can the person most effectively learn the target concept.

I. Introduction

In society, we often like to categorize knowledge into distinct categories. "Art", "English", "Science". The study of animal behaviors and the study of couture. In school and throughout one's career, individuals often focus on a specific field of inquiry, though in their overall life they may encounter a diverse array of experiences. This work seeks to explore how seemingly different types of knowledge are interrelated within the mind, and how that affects our interaction with knowledge yet to come.

Despite our natural desire to distinguish between different types of knowledge, sometimes knowledge naturally interrelates. A designer using notions of beauty to discover efficient solutions in the travelling salesman problem (TSP), a "hard" computational problem. The unusually bulky body of a fish, fast as it darts through the water, drives the engineering of a new, high-efficiency car. Knowledge of our position in the universe, as a verdant ball of fire and rock rotating around the sun, changed our conception of our relation with God. By experience, it seems we cannot so easily distinguish our relationship with one kind of knowledge from another kind–instead, it seems they are related.

This concept of interrelated knowledge has been studied across a variety of domains. In psychology, the fan-effect describes how the more pieces of knowledge we associate with some original piece of knowledge, the more difficult it could be to remember any one connection between some piece of knowledge with the original[1][2][3].[4] Because of this effect, researchers discovered that the technique of "association splitting" could help individuals with obsessive-compulsive disorder associate less trauma with certain words.[5] Instead of only associating "fire" with

[1] John Robert Anderson. "Retrieval of propositional information from long-term memory". In: *Cognitive psychology* 6.4 (1974), pp. 451–474.

[2] Michael F Bunting, Andrew RA Conway, and Richard P Heitz. "Individual differences in the fan effect and working memory capacity". In: *Journal of memory and language* 51.4 (2004), pp. 604–622.

[3] Gabriel A Radvansky and Rose T Zacks. "Mental models and the fan effect." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17.5 (1991), p. 940.

[4] Shannon D Moeser. "The role of experimental design in investigations of the fan effect." In: *Journal of Experimental Psychology: Human Learning and Memory* 5.2 (1979), p. 125.

[5] Terence HW Ching et al. "Association Splitting for Obsessive-Compulsive Disorder: A Systematic Review". In: *Current Psychiatry Research and Reviews Formerly: Current Psychiatry Reviews* 15.4 (2019), pp. 248–260.

"danger", they could also associate "fire" with "diamonds". Priming is another psychological effect by which focusing on one idea affects our understanding of another idea. Seeing "HORN" may lead us to think "TUSK" when see "T_SK", while seeing "WORK" may lead us to think "TASK" instead[6][7].[8]

A major step in artificial intelligence was on models that could learn *between* tasks. Transfer learning[9] uses knowledge from one domain to help better understand knowledge in another domain. Specifically, the training data and testing data of the model to create some kind of prediction or result may be different, and yet the model may still yield meaningful results on the testing data. Using a shared representation to learn several tasks in parallel has shown to lead to learning any one task better[10].[11] Meta-learning models learn how to learn new models that work on specific tasks, thereby taking advantage of the generalizeable process of learning itself[12].[13] Few-shot learning uses a general structure over knowledge in order to learn a concept from just a few examples[14].[15] For example, Snell, Swersky and Zemel (2017) represent knowledge as a "space"; representative examples of a class characterize part of that space as being a concept. The characterization of a "point" within this space is based on how points of knowledge around it have already been characterized.

Within the realm of academic coaching, we see experts such as Daniel Wong recommending that students interleave subjects to understand each better,[16] to help more clearly distinguish the salient knowledge in each, as if to counteract the fan-effect. In the book, *Why hard work and specialising early is not a recipe for success*, Hoog (2020) recommends to "connect new ideas to what you already know"[17] such as by creating metaphors for one topic in the language of another. I.e., in society, government is like the brain, the military is like the immune system, the bones are like the builder who make our infrastructure.

In this work, we examine how modelling knowledge as interrelated concepts would affect how people then interact with knowledge in the future. In particular, we study how "Speakers" can use different strategies to best help "Listeners" learn a topic better given this assumption that concepts are interrelated within the mind.

II. Model

We have created two types of models: a Listener and a Speaker. The Listener model examines the implications of the idea that

[6] Evan Weingarten et al. "From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words." In: *Psychological Bulletin* 142.5 (2016), p. 472.

[7] Endel Tulving, Daniel L Schacter, and Heather A Stark. "Priming effects in word-fragment completion are independent of recognition memory." In: *Journal of experimental psychology: learning, memory, and cognition* 8.4 (1982), p. 336.

[8] *Priming*. URL: `https://www.psychologytoday.com/us/basics/priming`.

[9] Sinno Jialin Pan and Qiang Yang. "A survey on transfer learning". In: *IEEE Transactions on knowledge and data engineering* 22.10 (2009), pp. 1345–1359.

[10] Rich Caruana. "Multitask learning". In: *Machine learning* 28.1 (1997), pp. 41–75.

[11] Theodoros Evgeniou and Massimiliano Pontil. "Regularized multi–task learning". In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. 2004, pp. 109–117.

[12] Chelsea Finn, Pieter Abbeel, and Sergey Levine. "Model-agnostic meta-learning for fast adaptation of deep networks". In: *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org. 2017, pp. 1126–1135.

[13] Ricardo Vilalta and Youssef Drissi. "A perspective view and survey of meta-learning". In: *Artificial intelligence review* 18.2 (2002), pp. 77–95.

[14] Jake Snell, Kevin Swersky, and Richard Zemel. "Prototypical networks for few-shot learning". In: *Advances in neural information processing systems*. 2017, pp. 4077–4087.

[15] Victor Garcia and Joan Bruna. "Few-shot learning with graph neural networks". In: *arXiv preprint arXiv:1711.04043* (2017).

[16] Doug Rohrer. "Interleaving helps students distinguish among similar concepts". In: *Educational Psychology Review* 24.3 (2012), pp. 355–367.

[17] Roediger Brown McDaniel. *Make it stick: the science of successful learning*. Belknap press of Harvard university press, 2014.

connectivity between topics will help further learning of the topics. The Speaker model uses different strategies of teaching the topics to Listeners.

Every person, or instance of Listener or Speaker, has an internal network of knowledge composed of $N = 20$ nodes. Each node represents an abstract topic, $n \in [N]$. Connections between nodes represent how much a person associates one topic with another topic. Between two topics, n_i and n_j, the connection, c_{n_i,n_j}, is a positive integer if it exists. There is also $\bar{c}_{n_i,n_j}$, which is used in computation and represents the real value of the connection, such that $c_{n_i,n_j} = \text{floor}(\bar{c}_{n_i,n_j})$. Each node has an associated value, $v_n \in [0, 1]$, which represents the person's current amount of knowledge within that topic. The closer v_n is to the maximum of 1.0, the greater the knowledge.

From a more concrete perspective, one could imagine the nodes represent topics such as "cooking", "physics", "chemistry", and that in some person's knowledge representation, they have not only form a "physics"-"chemistry" link but a "physics"–"cooking" and "chemistry"–"cooking" link as well.

Every time the Speaker, s, speaks, they output the topic, n, they are discussing as well as its associated knowledge value, v_n^s. Each Listener will then take in this n, v_n^s pair in order to update their internal network. This update has two major steps: (1) update the knowledge at n and (2) update the connectivity of n. The Speaker will speak and the Listener will listen for $T = 1000$ time-steps. Note that the knowledge and connectivity values of the Speaker remain constant over the T time-steps, while these values change for the Listener as they update their knowledge network.

i. Listener

i.1 Updating knowledge value

In order to achieve objective (1), the Listener, ℓ, uses the following equation to compute the amount, a_t, they have learned:

$$a_t = \lambda v_n^s (c_{n,t}^\ell + 1)(v_{n,t}^\ell + \epsilon).$$

$\lambda = 0.001$ represents the learning rate, which is the same for any Listener. $v_{n,t}^\ell$ and $c_{n,t}^\ell$ represent the value and connectivity of n within ℓ at time-step t. A small amount, ϵ or 1.0, is added to these quantities to prevent the product from going to 0.0.

Essentially, the amount the Listener ℓ learns is based on (a) how knowledgeable the Speaker s is on the topic, (b) how knowledgeable they already are on the topic, and (c) the amount of connectivity n has within their knowledge network.

To justify (a), we appeal to common experience–generally, the more knowledgeable the teacher, the more one can learn from them.

To justify (b), we point to work done by Lev Vygotsky on the "zone of proximal development".[18] The theory states that children have a series of zones by which they can execute a task with guidance versus on their own. In the beginning of learning how to perform a task, the children will need help from a more experienced person. Over multiple repetitions, they will begin to be able to do it by themselves. We extend this argument to say that when a person makes learning some topic their task, they are less effective at doing it on their own in the beginning, and more effective over time. Therefore, when they already know more about the topic, they know how to learn more about the topic.

(c) is the idea we are exploring.

[18] Saul McLeod. *Lev Vygotsky*. 2018. URL: https://www.simplypsychology.org/vygotsky.html.

After computing a_t, the Listener will update their network such that

$$v^{\ell}_{n,t+1} = v^{\ell}_{n,t} + a_t.$$

i.2 Updating connectivity

The Listener will then perform the second objective, which is to update the of the topic, n. Specifically, they will update the value of $c_{n,n'}$, where n' is the previous topic, the topic chosen by the Speaker at time-step $t-1$. The Listener will use the following set of equations:

$$\bar{c}_{n,n',t} = \mu + \bar{c}_{n,n',t},$$

$$c_{n,n',t} = \text{floor}(\bar{c}_{n,n',t}),$$

where $\mu = 0.05$ represents the rate of connectivity, which is the same for any Listener.

By analogy, one can think that by discussing topic n_{t-1}, the teacher is priming the students before teaching topic n_t. This is similar to the principle of learning-based priming[19][20].[21]

ii. Speaker

The Speaker must choose a topic to teach at each of the T time-steps. The goal is to maximize the amount the students learn about a target topic, n^*. The topic chosen at time-step 1 is always n^*. We present the following four strategies that were used: fixed, inspired, inspired-repeat, feedback-repeat.

ii.1 Strategy: fixed

The Speaker presents the same topic, n^*, at every time-step. This strategy serves as the baseline for comparison.

[19] Munakata. *CCNBook/Memory - Computational Cognitive Neuroscience Wiki*. 2014. URL: `https://grey.colorado.edu/CompCogNeuro/index.php/CCNBook/Memory`.

[20] Gaurav Malhotra et al. "On the persistence of structural priming: Mechanisms of decay and influence of word-forms". In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 30. 30. 2008.

[21] Kathryn Bock and Zenzi M Griffin. "The persistence of structural priming: Transient activation or implicit learning?" In: *Journal of experimental psychology: General* 129.2 (2000), p. 177.

ii.2 Strategy: inspired

The Speaker will "ramble" around the target topic. The goal of this is to improve the connectivity of the target topic within the internal networks of the Listeners.

The Speaker chooses as per the following algorithm:

1. Perform a modified breadth-first search around the previously chosen topic, n' for $r = 3$ iterations. When searching in the current iteration from a particular node, n_j, we include in the next iteration its neighbor, n_k, if the following condition is true:

   $$\alpha > \left(\frac{1}{2}\right)^{c_{n_j,n_k}},$$

   where α is a randomly generated number. Therefore, if the connection between n_j and n_k is strong, we are more likely to include n_k in our search.

2. Assign each found topic, n_i, with a score, z_{n_i}. We get the connectivity of each topic, c_{n_i} for each n_i, and then perform min-max over all of them. Therefore,

   $$z_{n_i} = \text{minmax}(c_{n_i})$$

   and represents the relative connectivity of n_i in comparison with the other topics.

3. Compute the probability of choosing each topic n_i by normalizing z_{n_i} with regards to the sum of all the scores.

4. Choose the output topic, n, given the probabilities.

ii.3 Strategy: inspired-repeat

The Speaker will ramble, but after every fixed interval, $t_r = 10$, the Speaker is guaranteed to discuss n^*. This strategy combines the previous two strategies in the hope that by improving the connectivity of the target topic in a Listener, the Listener will more easily learn the topic n^* when it is presented again.

ii.4 Strategy: feedback-repeat

The Speaker will poll all the Listeners it is teaching in order to more likely choose a topic they already associate strongly with the target. This polling is combined with the guaranteed repeat of n^* after every t_r time-steps to ensure n^* itself is taught.

The strategy goes as follows:

1. Compute the average knowledge network among all the Listeners. Therefore, each connection, c_{n_j,n_k}, between some topics, n_j and n_k, is the average of each Listener's c_{n_j,n_k}.
2. For the set of topics that have a non-zero connection to n^*, compute a score for each. Therefore, topic n_i has the score
$$z_{n_i} = \text{minmax}\left(1 - \left(\frac{1}{2}\right)^{c_{n_i,n^*}}\right),$$
where the min-max is taken over the arguments to it for all the topics in the set.
3. Compute the probability of choosing each topic n_i by normalizing z_{n_i} with regards to the sum of all the scores.
4. Choose the output topic, n, given the probabilities.

III. Methods

Our overall goal was to evaluate the performance of each of the strategies on teaching some specific topic to all the students. A sub-goal was to identify if there was any discrepancy in how "specialist" Speakers and "generalist" Speakers taught. A specialist for some topic n has high knowledge value, v_n, for the topic and low connectivity around the topic. A generalist for some topic n has at least a medium knowledge value, v_n, and a fairly high connectivity around the topic. The primary method we used to measure "connectivity around a topic" was the size of the sub-network that contained the topic. A sub-network is a subset of the network's nodes, such that for every node in it is connected to every other node. This was chosen over using the connectivity *of* the topic, c_n, because it had a greater impact on how topics could be chosen in several of the Speaker strategies, inspired and inspired-repeat.

There were several steps in our process:

1. Create the knowledge networks.
2. Among the created knowledge networks, identify which pairs of network(person)-topic are most suitable for being considered a "specialist" or "generalist" Speaker/teacher. All of the knowledge networks would be used to instantiate Listeners/students.
3. Observe general trends of how different values for students characteristics affect their learning trajectory.
4. Observe how different strategies affect the learning trajectory of students.
5. Compare each strategy's performance given a specialist or generalist teacher.

The former three are addressed in this section, and the latter two are addressed in Results.

i. Create knowledge networks

We created 1,600 knowledge networks. Each network had $N = 20$ nodes. During the process, we assigned values for each node and the connections between nodes.

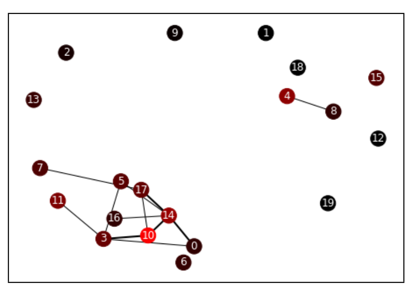


Figure 1: *An example of a generated knowledge network. The more red the node, the higher the associated value.*

The values were clipped between 0 and 1 from a Gaussian. We chose a mean of 0.25 since people often have a basic knowledge of most subjects. We chose a standard deviation of 0.25 so that a few of the values drawn could reach near the maximum of 1.

The links were created through an iterative process. At each iteration, a link between some nodes n_i and n_j was more likely to be created if (1) the values of these nodes were higher and (2) these nodes were already more connected through previous iterations. There was a parameter to tune the likelihood that the resulting network would be more "generalist"–that would encourage the creation of links between nodes. We leave the reader to find the details in the code.

ii. Identify teachers

We identified a set of specialist person-topic pairs by choosing the pairs where (1) the person's knowledge value for the corresponding topic fell within a specific range and (2) the topic's sub-graph size fell within a specific range. The heat-map in figure 2 demonstrates our choice of range for the first and second criteria. We ended up choosing pairs that had a knowledge value over 0.75 and a sub-graph size between 2 and 4. In total, there were 76 specialist person-topic pairs.

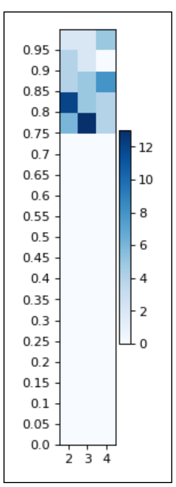


Figure 2: *Heat-map of how many person-topic pairs were found that fit the particular specialist-like combination of knowledge value (x-axis) and sub-graph size (y-axis).*

The set of potential generalist person-topic pairs was created through a similar process, with the addition that the average knowledge value of nodes within the sub-graph of the topic should be above some threshold. The heat-map in figure 3 illustrates the distribution of the average knowledge value of the sub-graph given the knowledge value of the person for the topic. We chose a threshold of 0.3.

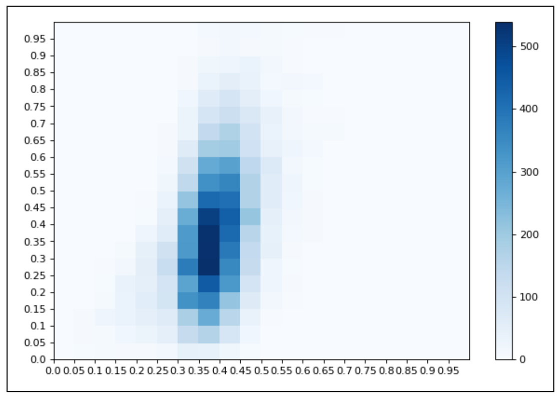


Figure 3: *Heat-map for how many person-topic pairs were found that had the particular combination of knowledge value (x-axis) and average value of nodes in the sub-graph (y-axis).*

Knowledge values of the person on the topic had to be at least 0.25. The size of the topic's sub-graph was chosen to be at least 5.

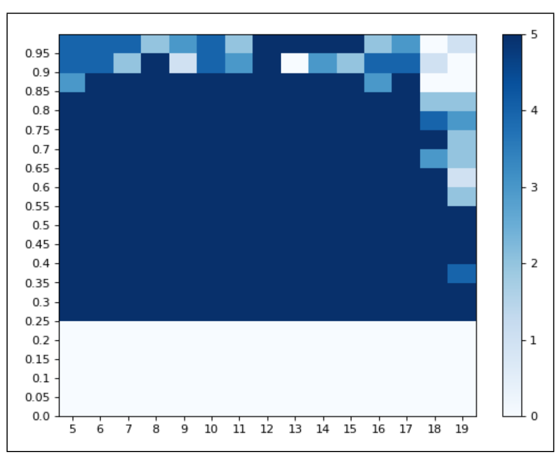


Figure 4: *Heat-map of how many person-topic pairs were found that fit the particular generalist-like combination of knowledge value (x-axis) and sub-graph size (y-axis) after filtering on the average value of nodes in the sub-graph.*

iii. Trends over time by student features

We ran the simulation of select person-topic teachers on all 1,600 students and examined how much the students in general learned over the T time-steps.

We examined how different strategies affected the trajectory of student learning. In order to take into account the teacher's own attributes, we chose teachers of varying (1) knowledge value [value] on the target topic and (2) sub-graph size [size] for both specialists and generalists. The following shows the attributes of the person-topic pairs used from both the specialist pool and the generalist pool. In the table, the values are binned in increments of 0.05.

Index	Value	Size
0	(0.75, 0.8]	2
1	(0.75, 0.8]	3
2	(0.75, 0.8]	4
3	(0.8, 0.85]	2
4	(0.8, 0.85]	3
5	(0.8, 0.85]	4
6	(0.85, 0.9]	2
7	(0.85, 0.9]	3
8	(0.85, 0.9]	4
9	(0.9, 0.95]	2
10	(0.9, 0.95]	3
11	(0.95, 1.0]	2
12	(0.95, 1.0]	3
13	(0.95, 1.0]	4

Table 1: *Table displaying the chosen attributes of specialist person-topic pairs to examine student trends over time.*

Index	Value	Size
0	(0.3, 0.35]	6
1	(0.3, 0.35]	10
2	(0.3, 0.35]	14
3	(0.3, 0.35]	18
4	(0.5, 0.55]	6
5	(0.5, 0.55]	10
6	(0.5, 0.55]	14
7	(0.5, 0.55]	18
8	(0.7, 0.75]	6
9	(0.7, 0.75]	10
10	(0.7, 0.75]	14
11	(0.7, 0.75]	18
12	(0.9, 0.95]	6
13	(0.9, 0.95]	10
14	(0.9, 0.95]	14
15	(0.9, 0.95]	18

Table 2: *Table displaying the chosen attributes of generalist person-topic pairs to examine student trends over time.*

We examined how three factors among students contributed to their learning trend: (1) their knowledge value on the target topic, (2) the size of their sub-graph around the target topic, and (3) the average value of the sub-graph around the target topic. After a qualitative analysis, we saw that the greater the person's knowledge value on the topic was, the faster the students' ability to learn the subject. The topic's sub-graph size and the topic's sub-graph average knowledge value had less of an impact though qualitatively at least appeared to follow a similar trend.

IV. Results

i. Trends over time by strategy

We show the trajectory for each strategy by using two generalist examples, and show how these trajectories change based on increased teacher knowledge value. The effect of increased sub-graph size had a similar effect as that of increased knowledge value. The specialist example follows a similar pattern, so we do not show them here. The x-axis represents time and the y-axis the amount the students learned. The graphs show the maximum learned (green), minimum learned (red), average learned (blue), and one standard deviation around the average learned (transparent blue). Though Speakers using any strategy besides fixed may choose topics that are not the target, in the graphs, we only show the moments where the Speaker did choose the target. For time-steps where the target is not chosen, we plot the y-axis as being the same as the value of the prior, most recent time-step where the topic chosen was the target.

One can see that the fixed examples are characterized by a small bump at the beginning of the trajectory, which grows more prominent with a larger teacher knowledge value. Beyond the initial bump, the trajectory levels off.

The inspired strategy was erratic, sometimes producing significant patterns of activity and sometimes showing very little activity in a way that was *not* correlated with knowledge value or sub-graph size. This effect is not shown obviously here but more so in the next sub-section. It is likely that in the cases of very little activity, the teacher ended up not choosing the target topic often enough to teach it to the students, thereby showing low learning rates.

The inspired repeat consistently showed significant learning during the course of the trajectory. The greater the knowledge value of the teacher, the sooner the students learned more. For the bottom example, one can see that the amount learning eventually decreases with time. This is due to the fact that there is an upper limit of 1.0 to the amount a person can know a topic within this model. Upon reaching this point, the student effectively stops learning.

The feedback strategy is characterized by a fairly flat trajectory of which there may be a slight rise near the end, which is more significant with a greater teacher knowledge value.

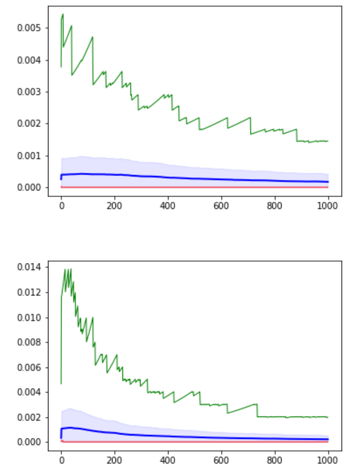


Figure 5: *Generalist index 3 (above) and index 16 (below) fixed example.*

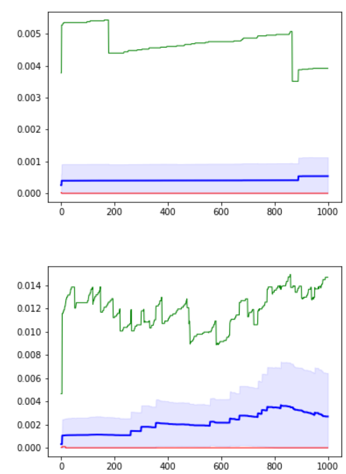
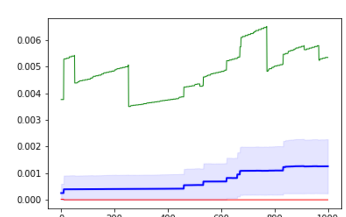


Figure 6: *Generalist index 3 (above) and index 16 (below) inspired example.*



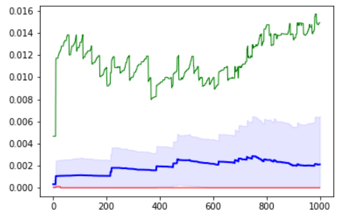


Figure 7: *Generalist index 3 (above) and index 16 (below) inspired repeat example.*

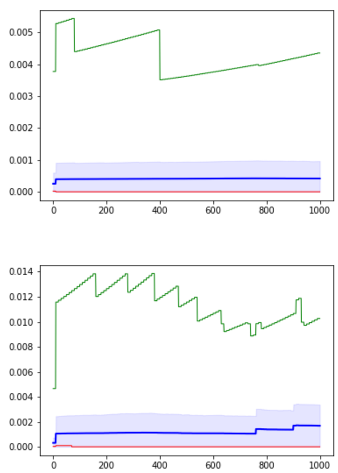


Figure 8: *Generalist index 3 (above) and index 16 (below) feedback repeat example.*

ii. Compare strategies overall

We also examined the average amount all students learned over the course of time for each strategy, for specialists and generalists, and for each teacher attribute. The following figures display a more global overview of the performance of each strategy. The entries at (0.75, 0.8], 4 for specialists and [1.0, 1.1), [19, 21) for generalists are empty because there happened not to be a generated instance that fit those slots.

The fixed strategy served as our baseline. It generally had a student learned value of around 0.2 to 0.4, increasing with increasing teacher knowledge value. Generally, specialists perform better, likely due to their overall higher knowledge value.

As mentioned before, the inspired strategy performed erratically across different teacher attributes. However, we can see in the cases where the students on average learned a significant amount, the generalists perform better.

The inspired-repeat strategy performed best overall and beat the baseline. Rather surprisingly, it appears that a *smaller* sub-graph size leads to better performance with this strategy. This may be the reason why the specialist teachers perform better than the generalists overall, as they have smaller sub-graph sizes.

Somewhat surprisingly, the feedback-repeat strategy performed worst, with students learning on average about 0.10 over the course of the time.

It would be interesting to investigate further why smaller sub-graph size leads to better performance in the inspired-repeat strategy, and why the feedback-repeat strategy had such an unexpectedly poor outcome.

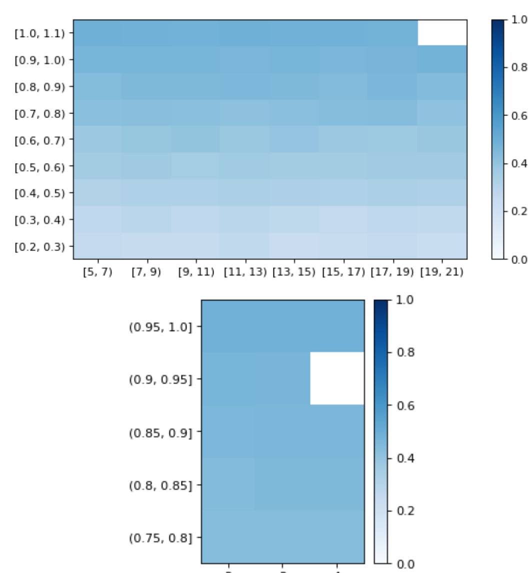


Figure 9: *Fixed strategy heat-map for generalists (above) and specialists (below).*

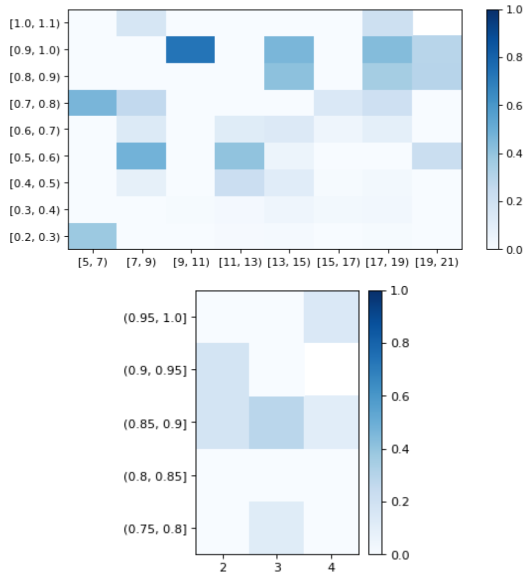


Figure 10: *Inspired strategy heat-map for generalists (above) and specialists (below).*

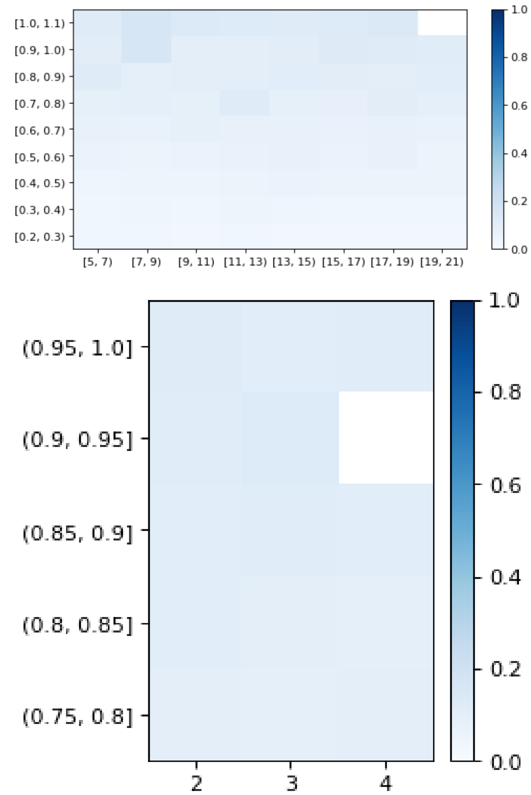


Figure 12: *Feedback repeat strategy heat-map for generalists (above) and specialists (below).*

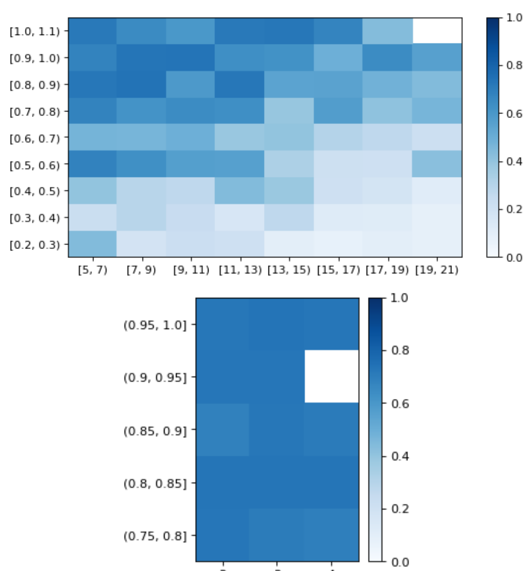


Figure 11: *Inspired repeat strategy heat-map for generalists (above) and specialists (below).*

V. Discussion

The repeat-inspired strategy has shown to be more effective than the fixed, baseline strategy not only across students but also over students of different characteristics. Even among students who have less knowledge of the target topic or less connectivity surrounding it, the repeat-inspired strategy still prevails over the fixed strategy when we qualitatively examine the trajectories of student learning over time. Below, we show example graphs of this general pattern for the student attributes of (1) topic value, (2) topic sub-graph size, (3) topic sub-graph average value. The x-axis represents time and the y-axis the amount learned when the target topic was chosen. For time-steps where the target is not chosen, we plot the y-axis as being the same as the value of the prior, most recent time-step where the topic chosen was the target. For a student value of [0.8,1.0), there is a divergence from the trend, but otherwise the inspired-repeat strategy helps students learn more effectively than the fixed strategy.

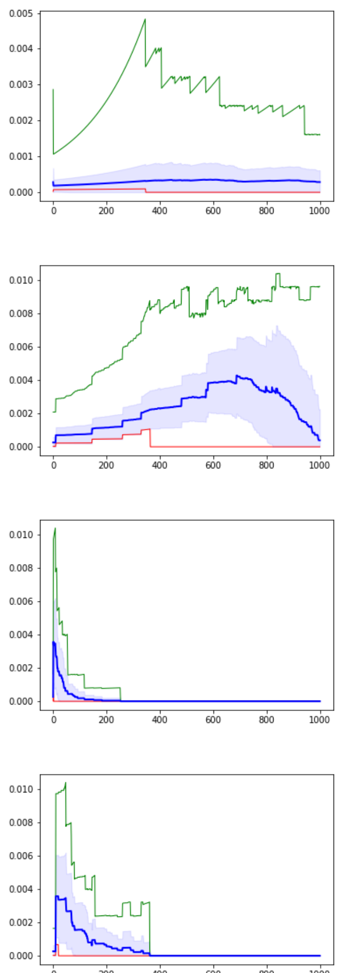


Figure 13: *Comparing fixed and inspired-repeat with different original student topic values. Here, we use teacher-pair index 13. The first two images show the trajectory for students with a value of [0.2,0.4) while the latter images for the value of [0.8,1.0). The first and third images show the use of the fixed strategy while the second and fourth show the use of the inspired-repeat strategy. For the value of [0.8,1.0), the fixed strategy helps students converge faster than inspired-repeat.*

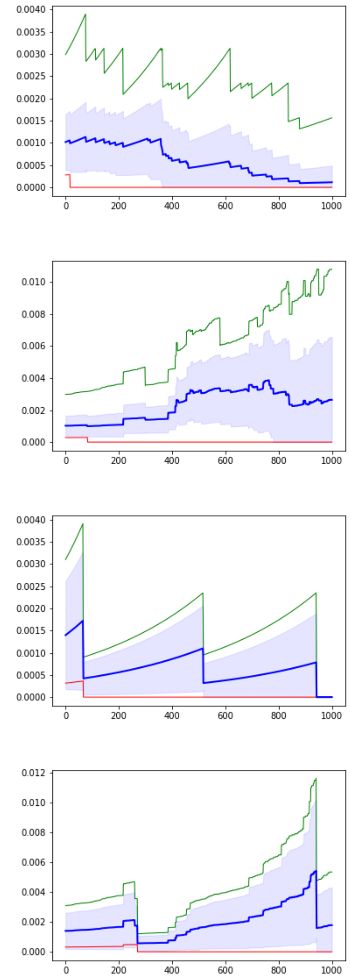
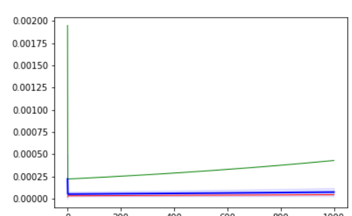


Figure 14: *Comparing fixed and inspired-repeat with different original student sub-graph sizes. Here, we use teacher-pair index 10. The first two images show the trajectory for students with a sub-graph size of 5 while the latter images for the size of 19. The first and third images show the use of the fixed strategy while the second and fourth show the use of the inspired-repeat strategy.*



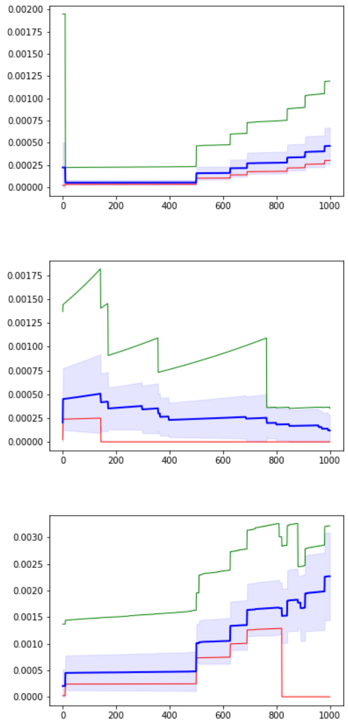


Figure 15: *Comparing fixed and inspired-repeat with different original student sub-graph average values. Here, we use teacher-pair index 0. The first two images show the trajectory for students with an average value of [0.0,0.2) while the latter images for the average value of [0.6,0.8). The first and third images show the use of the fixed strategy while the second and fourth show the use of the inspired-repeat strategy.*

It is possible that as the teacher is "rambling" by using the inspired-repeat strategy to discuss topics other than the target, the students will learn new or stronger connections between the topics surrounding the target, thereby better learning these topics and then learning the topic in question.

In some of the trajectories, we see that the average student amount of learning is actually decreasing with time. This is likely due to some students who can no longer learn because they have already reached the maximum amount of topic knowledge (1.0). In order to better clarify the results, it would help if this restriction did not exist. Another potential change is a more carefully considered way of updating links. Currently, links between nodes are updated the same amount each time, no matter the increased value of the nodes or the increased connectivity. The addition of the *forgetting* of knowledge without continued stimulus could also be potentially interesting. It would likely impact the performance of the inspired-repeat strategy, which has intervals between when the target topic is taught again.

The use of a teacher that was "specialist" or "generalist" seem to matter less than how that teacher ended up presenting information to the students.

Because the model presented here is theoretical, with its structure based on prior research and personal observations, it suffers from a lack of clear practical use. For example, though we may optimize the value of the repeat parameter for this model, that value has little meaning for real teachers teaching real students. It would be interesting to conduct further *quantitative* investigation into how different long-term strategies for teaching individuals lead to their eventual gains in a particular domain. However, there have been some observations that a more open approach to learning is generally better than a more specialized approach.[22] Some domains, such as playing chess and often sports may benefit from intense specialization–however most domains appear to require a variety of different experiences in order to develop a deeper understanding of the domain itself.

Beyond just the individual, there have been increasing calls within domains; from collecting upon the experience of a vast array of organizations to address plastic waste;[23] to converg-

[22] Hoog. *Why hard work and specialising early is not a recipe for success*. 2020. URL: `https://thecorrespondent.com/337/why-hard-work-and-specialising-early-is-not-a-recipe-for-success/44615161074-4a7c370a`.

[23] Ellen MacArthur Foundation. *New Plastics Economy*. 2017. URL: `https://www.ellenmacarthurfoundation.org/our-work/activities/new-plastics-economy`.

ing research in the social sciences to think more "interdisciplinary";[24] to approaching the teaching of mathematics, traditionally regarded as a dry and "stuffy" subject, with a more intuitive and visual perspective.[25] All of these suggest the worth of reflecting more upon the interconnectivity of knowledge. Furthermore, it suggests the need of creating structure to help people confront the ever-growing array of diverse knowledge, so we learn to work together instead of come to grow apart.

VI. Conclusion

We show how using the assumption that concepts are interrelated within the mind, at least as presented here, lead to certain patterns of knowledge interaction being more effective than others in learning a particular concept better. Specifically, we show that a strategy that uses a combination of focusing on the chosen concept and "related" concepts works best in helping students of varying prior understandings of the concept learn the concept more effectively.

[24] Madsbjerg McNamara Sieck Brashears Konner.

[25] Crane. *Discrete Differential Geometry: An Applied Introduction*. 2019. URL: `https://www.cs.cmu.edu/~kmcrane/Projects/DDG/`.

References

Anderson, John Robert. "Retrieval of propositional information from long-term memory". In: *Cognitive psychology* 6.4 (1974), pp. 451–474.

Bock, Kathryn and Zenzi M Griffin. "The persistence of structural priming: Transient activation or implicit learning?" In: *Journal of experimental psychology: General* 129.2 (2000), p. 177.

Brashears Konner, Madsbjerg McNamara Sieck.

Brown McDaniel, Roediger. *Make it stick: the science of successful learning*. Belknap press of Harvard university press, 2014.

Bunting, Michael F, Andrew RA Conway, and Richard P Heitz. "Individual differences in the fan effect and working memory capacity". In: *Journal of memory and language* 51.4 (2004), pp. 604–622.

Caruana, Rich. "Multitask learning". In: *Machine learning* 28.1 (1997), pp. 41–75.

Ching, Terence HW et al. "Association Splitting for Obsessive-Compulsive Disorder: A Systematic Review". In: *Current Psychiatry Research and Reviews Formerly: Current Psychiatry Reviews* 15.4 (2019), pp. 248–260.

Crane. *Discrete Differential Geometry: An Applied Introduction*. 2019. URL: `https://www.cs.cmu.edu/~kmcrane/Projects/DDG/`.

Evgeniou, Theodoros and Massimiliano Pontil. "Regularized multi–task learning". In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. 2004, pp. 109–117.

Finn, Chelsea, Pieter Abbeel, and Sergey Levine. "Model-agnostic meta-learning for fast adaptation of deep networks". In: *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org. 2017, pp. 1126–1135.

Foundation, Ellen MacArthur. *New Plastics Economy*. 2017. URL: `https://www.ellenmacarthurfoundation.org/our-work/activities/new-plastics-economy`.

Garcia, Victor and Joan Bruna. "Few-shot learning with graph neural networks". In: *arXiv preprint arXiv:1711.04043* (2017).

Hoog. *Why hard work and specialising early is not a recipe for success*. 2020. URL: https://thecorrespondent.com/337/why-hard-work-and-specialising-early-is-not-a-recipe-for-success/44615161074-4a7c370a.

Malhotra, Gaurav et al. "On the persistence of structural priming: Mechanisms of decay and influence of word-forms". In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 30. 30. 2008.

McLeod, Saul. *Lev Vygotsky*. 2018. URL: https://www.simplypsychology.org/vygotsky.html.

Moeser, Shannon D. "The role of experimental design in investigations of the fan effect." In: *Journal of Experimental Psychology: Human Learning and Memory* 5.2 (1979), p. 125.

Munakata. *CCNBook/Memory - Computational Cognitive Neuroscience Wiki*. 2014. URL: https://grey.colorado.edu/CompCogNeuro/index.php/CCNBook/Memory.

Pan, Sinno Jialin and Qiang Yang. "A survey on transfer learning". In: *IEEE Transactions on knowledge and data engineering* 22.10 (2009), pp. 1345–1359.

Priming. URL: https://www.psychologytoday.com/us/basics/priming.

Radvansky, Gabriel A and Rose T Zacks. "Mental models and the fan effect." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17.5 (1991), p. 940.

Rohrer, Doug. "Interleaving helps students distinguish among similar concepts". In: *Educational Psychology Review* 24.3 (2012), pp. 355–367.

Snell, Jake, Kevin Swersky, and Richard Zemel. "Prototypical networks for few-shot learning". In: *Advances in neural information processing systems*. 2017, pp. 4077–4087.

Tulving, Endel, Daniel L Schacter, and Heather A Stark. "Priming effects in word-fragment completion are independent of recognition memory." In: *Journal of experimental psychology: learning, memory, and cognition* 8.4 (1982), p. 336.

Vilalta, Ricardo and Youssef Drissi. "A perspective view and survey of meta-learning". In: *Artificial intelligence review* 18.2 (2002), pp. 77–95.

Weingarten, Evan et al. "From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words." In: *Psychological Bulletin* 142.5 (2016), p. 472.