Student Metacognition in Problem Solving and Biologically Inspired Cognitive Architectures

Alexei V. Samsonovich[1], Anastasia Kitsantas[2], Ellen O'Brien[3], and Kenneth A. De Jong[1]

[1]Krasnow Institute for Advanced Study, George Mason University, MS 2A1; [2]College of Education and Human Development, George Mason University, MS 4B3; [3]Department of Mathematical Sciences, George Mason University, MS 3F2

4400 University Drive, Fairfax, Virginia 22030, USA
{asamsono, akitsant, eobrien, kdejong}@gmu.edu

Abstract

To study the role of metacognition and self-regulation in student problem solving, college students were asked to solve problems in mathematics using a software tool assisting metacognition at the forethought stage. Student activity involved selecting relevant facts and strategies represented on the screen and connecting them by arrows. It is found that patterns of arrow drawing have significant correlates in student performance and cognitive states, specifically: (i) forward chaining is significantly more predominant, and backward chaining is significantly less frequent, during problem solving than during initial exploration that is not goal-driven; (ii) students scoring in the middle are more likely to enter convergent pairs of arrows compared to students who scored low or high. These findings enable diagnosing student problem solving and imply constraints on selection of cognitive architectures used for modeling student preparation for problem solving.

Keywords: SRL; metacognition; ITS; mind reading; BICA

Background and Motivation

In recent years considerable progress was made in machine learning, as well as in intelligent tutoring systems (ITS) that help students to improve their learning and self-regulation skills. Nevertheless, successful development and application of ITS is limited by the lack of understanding at a computational level of how students think about mathematical problems. It is known that in most cases students employ self-regulated learning (SRL: Zimmerman 2000; Figure 1): a complex of meta-cognitive, motivational and behavioral control techniques to support their own problem solving process. SRL is the essence and hallmark of human top-level cognition and learning, and is vital for academic achievement (Pashler et al. 2007; Zimmerman 2008). In particular, it includes metacognition during the forethought stage. Yet, this element has no well-known counterparts in traditional methods of machine learning or problem solving (e.g., Russell & Norvig 2010).

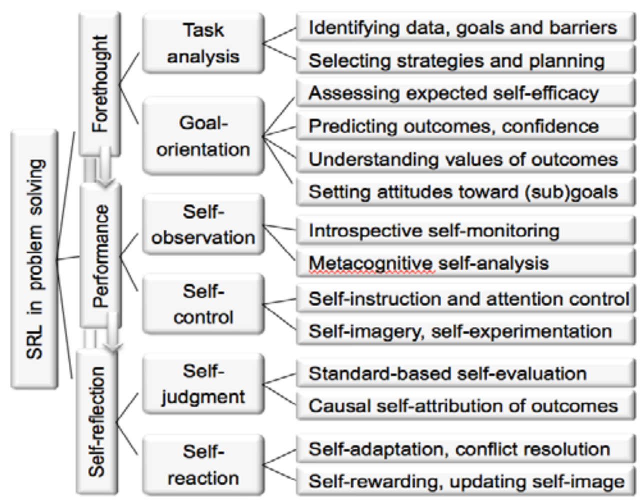

Figure 1. Model of SRL in problem solving.

On the other hand, the human ability to self-regulate is a learned skill on its own, and therefore can be acquired by students as well as by artifacts through a learning process. In education, the most effective forms of self-regulatory training occur through one-on-one instruction between the teacher and student. This approach, though effective, is difficult to implement broadly today, given large class sizes and the limited number of SRL-trained instructors. The problem could have a general solution in the form of an ITS. Yet, current ITSs provide limited SRL support, leaving the problem open (Winne & Nesbit 2009).

In order to make progress, it is necessary to have a precise theory of SRL, preferably in the form of a cognitive architecture, that goes beyond phenomenology available today. In order to develop and validate this theory, it is necessary to have a measuring device that can make student SRL visible and available for precise analysis at a detailed level. Then the creation of efficient SRL assistants, as well as SRL-based human-level artificial learners could be made possible.

Idea of our approach is that, in order to register student SRL processes in detail, a medium in the form of a cognitive architecture can be used, where elements of metacognition are formally represented as objects. Students interact with this medium using a graphical user interface (GUI), thereby performing SRL processes outside of their brain and making them available for recording. Among related approaches are studies involving concept maps constructed by students (e.g., Ritchhart et al., 2009; Hymel et al., 2011).

For example, consider the process of construction of knowledge how to solve a given mathematical problem. This process usually starts at a higher level, when student decides which strategies to use and which intermediate goals to set (the forethought phase of SRL in Figure 1: based on Zimmerman and Kitsantas 2006). The ability to record this process unnoticeably for the student opens a window into the student mind at a metacognitive level. Measuring behavioral, cognitive and metacognitive correlates of the recorded elements of SRL allows us to consider likely constraints on the cognitive architecture models intended to describe student problem solving.

The present study is aimed at a long-term goal to build a metacognitive ITS based on a biologically-inspired cognitive architecture developed at George Mason University (GMU BICA). This architecture allows for the replication of student SRL processes outside of the brain, using a simple graphical interface (Samsonovich et al. 2009). In this sense the architecture allows one to “read” and diagnose student’s mind during problem solving, and therefore to be able to provide a necessary feedback to the student at a metacognitive level.

Methods

The approach taken in this study is based on a simplified version of the Cognitive Constructor architecture (CC: Figure 2, Samsonovich et al. 2008, Samsonovich 2009) implemented with GUI in Matlab. In particular, this architecture includes a higher-level symbolic virtual environment as a learning environment for construction of new knowledge (e.g. a solution to a given problem). Elements of knowledge are represented by graphical objects that can be manually connected to each other by arrows. In the complete CC architecture, these objects would correspond to schemas, and their connection would result in the binding of schemas (Samsonovich et al. 2009). In this study, however, CC was used primarily as a “measuring tool” to observe student metacognition at a microscopic level; therefore, the CC architecture was only partially implemented (Figure 2).

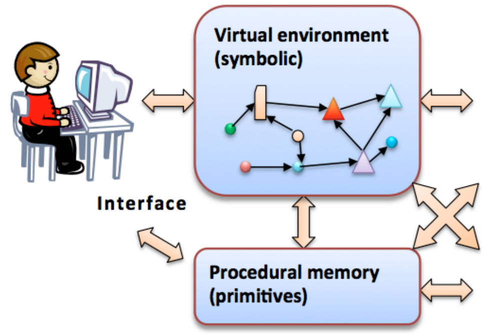

Figure 2. Implemented part of CC architecture used in this study. Arrows on the right would lead to architecture components that were truncated (for complete architecture see Samsonovich 2009).

Intervention

In the present study, CC was used by students during the forethought phase of SRL only, for constructing an approximate design of a solution of the given mathematical problem. Specifically, this was done by (i) selecting the relevant cognitive elements that may be necessary for constructing the solution and (ii) connecting selected elements by arrows. The list of cognitive elements given to students on a separate sheet included facts, rules, strategies, intermediate goals, techniques, observations, commonsense considerations, etc., labeled by letters of the alphabet. Elements were entered into, or removed from the virtual learning environment by pressing the corresponding keys on the keyboard (students also had an option to enter all available elements to the virtual environment before reading the list). After that, students started connecting entered elements to each other by arrows. Arrows were drawn on the screen by dragging the mouse. Removal of entered arrows was allowed and was done by clicking on the arrow. The outcome of this process is further referred to as “a diagram”. Instructions given to students implied that this diagram was expected to be a connected digraph representing a design of the intended solution process conceived to the best of the student’s ability before solving the problem. An example of a diagram is shown in Figure 3. Only after finalizing the diagram, students working in the experimental mode were allowed to start solving the problem on paper (see Procedures below).

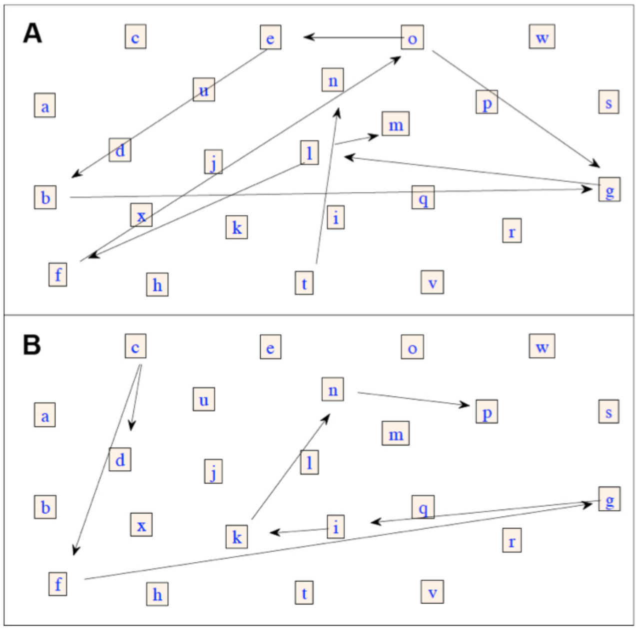

Figure 3. Two examples of diagrams constructed by one participant. A: diagram constructed during warm-up exercise; B: diagram constructed during forethought phase preceding solution of Problem 2.

Study Materials

An example of a test problem and the accompanying it list of potentially useful cognitive elements quoted from the study materials is given below. Not all elements of the list were useful and correct: some of them were intended to be misleading.

Problem 3. *Two teams with 10 members each play a game. Each player of the winning team shakes the hands of all his teammates as well as of each player of the loosing team. How many handshakes take place?*

(a) ***Data:*** *There are Two teams with 10 members each.*
(b) ***Data:*** *Each player of the winning team shakes the hands of all his teammates as well as of each player of the loosing team.*
(c) ***Goal:*** *Determine how many handshakes take place.*
(d) Everybody will shake everybody's hand.
(e) Divide the problem in two parts.
(f) Multiply the number of handshakes one winner does by the number of winners.
(g) Divide the number by two.
(h) Use the formula for the number of combinations.
(i) Use the formula for the number of permutations.
(j) Use Venn diagram.
(k) Consider all winners one by one.
(l) Count all possible handshakes one winner can make

Students working in experimental mode had sufficient time to example this list, then selected relevant elements and put them in the virtual environment, then started connecting selected elements by arrows. This was done before students were allowed to start solving the problem.

Participants

A total of 41 students participated in the study. Seventy three percent or 30 of the students were female and 85% or 35 students reported that English is their native language. The ethnic breakdown was as follows: 24% White, 9.8% Black, 14.6% Hispanic, 9.8% Asian, and 4.9% reported other. One student did not report her ethnicity. A total of 48.8% of the students were from Northern Virginia while 26.8% reported that they were from a different part of Virginia. A total of 9 students (22%) reported that they were from out of state. In terms of their student status, 53.7% were freshmen, 24.4% were sophomore, 19.5% were juniors, and 2.4% were seniors. In terms of transfer status, 88% were not transfer students while 12% were transfer students. All of the students were full time students and most (63.4%) of the students reported that they were planning on staying at the university for 7-8 semesters. Most of the students had graduated from public high schools (93%) while 2.4% graduated from private, non-religious high schools and 2.4% were home-schooled. In terms of prior performance, the students reported an average of 1291.90 score on the SATs overall, with an average of 537 and 647 on math and verbal portions, respectively. There were seven students who completed the ACT and reported an average of 19 on the ACT composite score. In terms of high school GPA, students reported an average of 3.38. Students were also asked what their target semester GPA was and the average reported score was 3.34.

Procedures

Undergraduate students enrolled in a mathematics course were recruited to participate in this study. The instructor of the course distributed fliers to the students to advertise the study. Students received 10 points extra credit in their math course. After signing the informed consent, each student was randomly assigned to one of the 32 groups used in this study. The groups differed from each other by the order in which problems were solved and also by the mode in which each problem was solved. Two modes were used: experimental (e) and control (c). In the experimental mode, participants worked on the problem using CC on an individual desktop computer. They were instructed to represent on the computer screen how they are planning to solve the problem, before actually solving it. In the control mode, participants worked on the same problem on their own, using the same printed materials, but were not instructed to construct a plan of solution before solving the problem, and did not use the computer. Each student solved 4 problems: 2 in experimental mode and 2 in control mode. The order of problems in different groups was: 1234, 1243, 2143, 2134, 3412, 4321, 4321, 4312; the sequences of modes used in different groups were eecc, ccee, ecec, cece. All possible combinations of the above problem orders and mode sequences were used, resulting in 32 groups. Groups were merged in various combinations during data analysis. In addition to experimental and control modes, all participants were asked to do a warm-up exercise before the study. The exercise consisted in connecting letters by arrows on the computer screen without any specific goal or problem in mind associated with letters on the screen. In addition to the above, students completed two of the identical SELF scale forms, one before and another after the experiment, and two reflection forms after the test, all described below.

Measures

Three kinds of measures were used in this study: introspective surveys, achievement scores, and CC recordings made during construction of the diagram.

Self-Efficacy for Learning Form (SELF)

The form known as SELF (Zimmerman & Kitsantas, 2007) was used in this study. This questionnaire is a 19 item measure that attempts to measure student self-efficacy for using various task strategies for learning. Students responses ranged on a 0 ("Definitely cannot do it") to 100 ("Definitely can do it") Likert scale in 10-unit increments. An example question is, "*When you are feeling depressed about a forthcoming test, can you find a way to motivate yourself to do well?*" The reliability coefficient for this sample suggest strong internal consistency $\alpha = .91$.

Reflection 1. Students were asked to hand write out a series of reflections. The reflections included four different writing prompts which included thoughts about their experience, what they found challenging, any addition feedback for researchers, what they thought was helpful, and any additional comments about the study.

Reflection 2. Students were asked to reflect and comment on the outcomes of their assessment. Students were provided with three writing prompts which included: "Explain what mistakes you made and why, and what would you do differently next time with similar problem"; "What did you think about the diagram and its implementation? Was it helpful?"; and "What did you learn in this session about the process of solving problems in mathematics, and how did the tool help you in this learning?"

Achievement

Achievement in problem solving was measured with performance scores (grades) given for each of the four problems that students had to solve in this study. The grading was done by one of the authors (E.O.), a GMU Mathematics instructor who is teaching the classes from which the students were recruited. Only the problem solution sheets were used for grading; the grader did not access to surveys and diagrams drawn by students on paper or entered in the computer and stored electronically. Student identities were encoded and were unavailable to the grader. Grades were given on a scale of integers from 0 to 3 interpreted as follows. 0: no response or obviously incorrect solution with no work, calculations or steps of any kind shown; 1: some attempt at finding a solution and/or true statements are written, but would not lead to the correct solution; 2: some understanding is demonstrated with correct steps or calculations, but the final answer is incorrect; 3: correct solution.

CC Recordings of Metacognitive Choices

All automatically recorded events of arrow addition for each constructed diagram, in preparation for problem solving and during warm-up exercises, were classified into five motives, or categories of added arrows, depending on how and whether the entered arrow was adjacent to the last previously entered arrow (Figure 4). The motives analyzed in this study include M1: forward chaining, when the next arrow starts from the end of the previous one, M2: backward chaining, when the next arrow ends at the origin of the previous one, M3: fan-out, when the next arrow starts from the same node as the previous one, M4: convergence, when the next arrow ends on the same node as the previous one, and M5: disjoined (non-adjacent) entry, when the next arrow does not share any node with the previous one.

Figure 4. Motives in arrow entering to the diagram in cases when the new arrow (red) is adjacent to the previous arrow (light-blue). 1: forward chaining, 2: backward chaining, 3: fan-out, 4: convergence.

Duplicate arrows were discarded. An arrow was regarded as disjoined if it was the first arrow added at the beginning of the process or the first arrow entered immediately following arrow removal, unless the removed arrow was the last previously entered arrow.

In addition, similar motives in the complete diagram (digraph) were analyzed, including, M6: an edge starts at a node that has at least one incoming edge, M7: an edge ends at a node that has at least one outgoing edge, M8: an edge starts at a node that has at least one other outgoing edge, M9: an edge ends at a node that has at least one other incoming edge, and M10: an edge is isolated. For further validation of results, shuffling was used. Shuffling was done as follows: all arrow entries within one diagram constructed by a student were randomly permuted, then the same calculation of motif probabilities was performed.

Results and Analysis

Effects of the Intervention

Effects of the intervention on student learning were estimated using achievement grades, and effects on student SRL were estimated using SELF, which is traditionally used as a measure of SRL confidence (Zimmerman & Kitsantas, 2007).

An independent t-test comparing experimental and control conditions on post SELF showed no significant differences between the groups ($t(39) = .97$, $p = .33$). Thus, no significant effect of the intervention on the perceived self-efficacy in problem solving was found in this study.

In addition, independent t-test comparing differences on achievement as measured by the three problems indicated that students in the control group achieved no different than students in the experimental group in any of the three problems: problem one ($t(39) = -1.37$, $p = .18$); problem two ($t(39) = -.33$, $p = .74$); and problem three ($t(31) = .12$, $p = .90$). The three separate problems were also averaged to examine potential overall differences. An independent t-test on the overall achievement indicates no significant differences between experimental and control modes of problem solving ($t(39) = -.87$, $p = .39$).

Similar analysis was performed for transfer learning effects measured by comparison of the control mode achievement before and after a CC experience. No significant difference was found between these conditions. Finally, it was found that the order in which a given problem was solved during the session (1st, 2nd, 3rd or 4th) also had no significant effect on achievement.

Thus, no significant effect of the intervention on student achievement in problem solving, as well as on student SRL confidence, was found during this short-term pilot study.

The absence of significant effects of the intervention on student problem solving and SRL abilities allows us to consider CC as a non-perturbing measuring device, based on the findings described in the following subsection.

Detection of Metacognitive Processes Using CC

Probabilities of motives characterizing patterns of adjacent arrow entering (Figure 4) were measured separately for diagrams constructed during problems solved in experimental mode (using CC) and for diagrams constructed during warm-up exercises, before starting the test. To compute probabilities, motif frequencies were normalized by the total counts of arrows in each diagram.

First, MANOVA was applied to the set of 4-vectors of frequencies characterizing the first 4 motives: forward chaining, backward chaining, fan-out, and convergence. Differences between the first and the second problem solved by the student with CC were found not significant in this analysis (P~1). However, significant differences (D=1, P<1.5e-6) were found between warm-up exercises and problem solving. Then, one-way ANOVA and Kruskal-Wallis (KW) test, along with other statistical measures, were applied to each motif frequency separately. Results are summarized in Figure 5 and in Table 1.

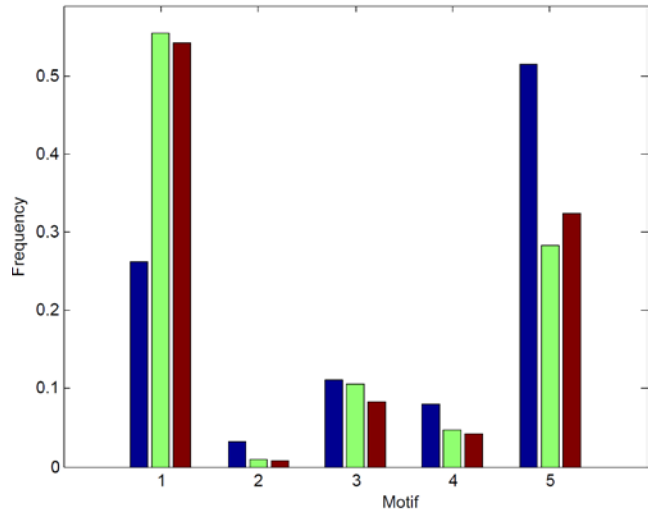

Figure 5. Probabilities of motives in arrow entering during warm-up exercises (blue), first problem solved with CC (light green), and second problem solved with CC (dark red). 1: forward chaining, 2: backward chaining, 3: fan-out, 4: convergence, 5: disjoined arrows.

Among motives in arrow sequencing (M1-M5), significant differences between warm-up exercises and problem solving were found for forward chaining (M1) and for backward chaining (M2). Probabilities of fan-out (M3) and convergence (M4) were found not significantly different under the two conditions. The predominance of forward chaining was stronger during problem solving than during warm-up exercises, while the probability of backward chaining was significantly higher during warm-up exercises. All results are summarized in Table 1.

Thus, probabilities of most motives are significantly different during warm-up exercises and during preparation for problem solving.

Prediction of Achievement Using CC Measures

Next, interaction between achievement grades and motif frequencies was analyzed. Among frequencies of the four motives of adjacent arrow addition, the only significant correlation was found with the frequency of convergence (motif M4: ANOVA P<0.0003, Table 2, Figure 6; KW P<0.025). Specifically, high probability of convergent arrow entering predicts an intermediate grade, while the lowest and the highest grades are associated with the low probability of convergent arrows entered to the diagram.

	Motives in sequences of actions					Motives in complete diagrams				
Motif ID (Figure 4)	1	2	3	4	5	6	7	8	9	10
Motif description	forw.c.	back.c.	fan-out	conv.	disj.	forw.	backw.	fan-out	conv.	disj.
Av. freq. - exercise (N=44)	0.26	0.032	0.11	0.079	0.52	0.25	0.24	0.21	0.21	0.08
Standard deviation - exercise	0.25	0.068	0.15	0.16	0.26	0.1	0.1	0.12	0.16	0.2
Av. freq. - problems (N=74)	0.55	0.0085	0.1	0.045	0.18	0.35	0.36	0.16	0.13	0.01
Standard deviation - problems	0.28	0.028	0.18	0.12	0.14	0.13	0.13	0.19	0.14	0.07
Change from exercise to problems	0.29	-0.028	-0.015	-0.034	-0.21	0.1	0.11	-0.06	-0.09	-0.07
Qualitative significant change	incr.	decr.	none	none	decr.	incr.	incr.	decr.	decr.	decr.
Pearson correlation coefficient R	0.46	-0.24	-0.04	-0.12	-0.47	0.38	0.43	-0.16	-0.28	-0.25
ANOVA P-value < (> for insign.)	1.E-07	0.01	0.63	0.18	6.E-08	2.E-05	2.E-06	0.002	2.E-03	7.E-03
Kruskal-Wallis P-value	1.E-06	0.0006	0.043	0.002	2.E-06	2.E-06	5.E-07	4.E-04	5.E-05	1.E-04

Table 1. Results of motif frequency analysis.

Prediction of Achievement Using CC Measures

Next, interaction between achievement grades and motif frequencies was analyzed. Among frequencies of the four motives of adjacent arrow addition, the only significant correlation was found with the frequency of convergence (motif M4: ANOVA P<0.0003, Table 2, Figure 6; Kruskal-Wallis P<0.025). Specifically, high probability of convergent arrow entering predicts an intermediate grade, while the lowest and the highest grades are associated with the low probability of convergent arrows entered to the diagram. Confirming this result, polynomial regression analysis gives no significant linear dependence between grades and probabilities of convergent arrows, while the first two coefficients in the quadratic fit are non-zero at the level of significance 0.05.

Table 2. ANOVA for M4 – achievement grade interaction

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	0.231	3	0.077	7.162	0.0003
Within Groups	0.740	69	0.011		
Total	0.971	72			

Interaction Between SRL and CC Measures

Next, interaction between student SELF scores and other recorded variables was analyzed. No significant linear (Pearson) correlation was found between any variable and SELF scores. Next, KW test was used to detect interaction. SELF scores were treated as group variables (in some trials rounding to discrete values was used, the number of values varied). In all explored cases, no significant interaction was detected at the level of significance alpha = 0.01. Significance at the level 0.05 was sensitive to rounding.

Qualitative Analysis of Reflection Forms

Reflection 1. Qualitative statements responding to the question about participants overall experience with the study revealed that many of the students (n = 16) reported that they liked the different strategies that were listed that they could choose from to help they solve the problem. Example statements include: “Having a list to choose possible problem solving methods” and “Being able to write out the order in which I solve the problem with steps.”

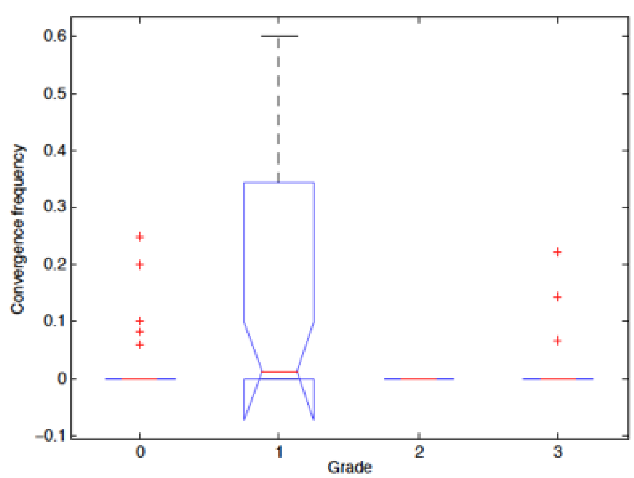

Figure 6. Interaction between achievement grades and normalized frequency of convergent arrow entering, computed over 72 diagrams constructed by students with the total of 928 added arrows (corresponds to Table 2). High frequency of convergent arrows corresponds to the intermediate grade 1.

In terms of the question that queried students perceptions of which aspects they found most challenging revealed that many of the students (n = 14) commented on how difficult the problems were. Specifically, one student said “Solving the problems without being prepared for it (was the most difficult part)” while another student said “About 2 of the problems were very challenging in which I became confused and unsure of my answers.” Another theme that emerged were the technological difficulties. Comments include: “Using the system/directions were not clear on what to do” and “Learning how to use the software (was the most difficult part).” The responses to this second question align with the responses on the third question, which asked students to provide feedback. Many of the students (n=19) replied that they would like further instructions on solving the different problems (“A little more instruction on how to solve the problem (should be provided)”).

The fourth question asked students to comment on if they thought the diagram was helpful. A total of 13 students did not think the diagram was not helpful (“No! It was rather confusing.”) while a total of 20 students felt that the diagram was helpful (Yes-the diagrams help one think about each individual step”).

A total of 20 students responded to the question that asked students to provide additional comments. Out of those 20 responses, a total of 7 students commented that this was an interesting study. For example, “(This study was) Thought-provoking, interesting” and “This was an interesting experience and I believe that it could, in fact, help someone”

Reflection 2. The first question asked students to explain the different mistakes that they made and what they would do differently next time to avoid the mistakes. Students (n=11) attributed their errors to not following the strategy. The second question asked students to describe the lessons learned from this study. A total of 26 students had responded to the question and a majority of the students (n = 18) responded that they had learned some sort of strategy to solve problems. Example quotes include “I learned that diagrams can be another tool for solving problems (I never knew that before)”; “The steps you take are very important in calculating a problem. The tool helped me organize the steps”; and “I learned that you need to consider all the different options (to solve problems).”

Discussion

Thus, we have found in this study that motives in arrow entering while working with CC have significant correlates in student performance and cognitive states. Specifically, we found the following.

(1) Forward chaining is significantly more predominant, and backward chaining is significantly less likely during metacognitive problem analysis compared to warm-up exercises. This finding suggests that forward chaining pattern correlates with metacognitive processes performed while using CC to select an approach to solve the problem. Therefore, it suggests that forward chaining of planned steps is characteristic of human metacognition, with implications for selection of cognitive architectures to describe student metacognition.

While this effect could be attributed to the bias imposed by the final topology of the diagram, our analysis shows that this is not the case: compare frequencies of M1, M2 (characterizing sequence of arrow entries) vs. M6, M7 (characterizing the final diagram topology), under the two conditions.

(2) It is found that students who score in the middle in their achievement are more likely to enter convergent pairs of arrows compared to students who scored low or high.

(3) In some cases we also observed that analysis of individual diagrams could tell us whether the student understands the problem and has a correct idea of its solution; in other words, CC is potentially capable of predicting the success (the high grade) in problem solving.

Together these findings enable diagnosing student problem solving at multiple levels: they allow us to tell whether the student is working on the problem, whether a correct solution is likely to be found, or, despite student efforts, the grade is going to be in the middle.

It is tempting to speculate that the pattern of convergent arrows indicate exploration of alternative approaches to solving the problem, and therefore potentially efficient learning. More data, however, is needed to make this conclusion.

The observed significant differences in motif frequencies imply constraints on selection of cognitive architectures for modeling student learning. Indeed, traditional architectures like Soar and ACT-R, or GMU BICA without metacognition, seem unlikely to account for predominance of forward chaining as opposed to backward chaining.

Consider an example of a problem given together with potentially useful steps and strategies as the following list.

(a) Data: $2x+3 = x+2$.
(b) Goal: Solve for x.
(c) Subgoal: Isolate x.
(d) Strategy: Use the subtraction rule.
(e) (other knowledge)

Given all these elements represented by schemas in a cognitive architecture with the rules of binding specified in detail, the process of finding a solution may amount to constructing a diagram like those constructed by students. In this sense, the model used in this study is complete, yet a very abstract one. For many practical purposes, however, it would be sufficient to build a complete model at this level of abstraction.

To solve the above example (a-e), a rational automated planner would start from the goal, then will recall that to solve a linear equation for *x*, *x* needs to be isolated. The next step will be to recall that using the subtraction rule in some cases helps to isolate *x*. After that it will remain to try several candidates for subtraction from both sides. In other words, backward chaining would be efficient in planning a process of solving this problem. In contrast, if the agent starts from the data and explores all applicable actions, then it may face a combinatorial explosion.

Why then we observed that students use forward chaining and in particular avoid using backward chaining during problem analysis? Apparently, what was observed was not the primary metacognitive process leading to finding a solution. It looks more like a validation of a candidate plan through its mental simulation: students perform actions in their mind in the order in which they will perform them in reality. What is the function of this process remains an open question: for example, it is possible to do reflection and self-judgment while watching the simulated performance from a third-person perspective.

This is the kind of findings that is needed for design of cognitive architectures. Indeed, this observation may favor metacognitive architectures that are capable of mental modeling of planned future behavior, as opposed to traditional planners and simple rule-based problem solvers.

Next, in the light of the above interpretation, it is plausible that students who enter convergent arrows are considering alternative approaches to solving the problem, without being confident in any of them. This is why they get intermediate scores (our second finding). Reproducing this phenomenon in a cognitive architecture is a challenge that may help to understand human learning and cognition.

It is also remarkable that in both cases interactions between motives and cognitive and behavioral states are stronger (and in some cases are only present) among motives characterizing sequences of arrows and not final topologies of diagrams. Frequencies of specific patterns of arrow entering characterize metacognitive processes associated with those patterns, while final topologies characterize the outcome of those processes.

At the same time, final topologies imply a bias for the sequences. For example, if the diagram is a one-dimensional chain, then there is no room for convergence in it. To check for this confounding bias, we performed similar analysis with two kinds of motives, and also used shuffled data (with the order of entered arrows randomly permuted). Shuffling resulted in the disappearance of observed effects. It is also remarkable that the dominance of forward chaining over backward chaining, as well as the significance of convergence as a correlate of medium achievement, are only present in sequence characteristics, suggesting that they characterize the process rather than the topological constraint of the diagram.

While no significant effect of the intervention on student performance was observed compared to controls, many students reported that the tool helped them to learn how to solve problems. At the same time, the absence of significant effects suggests that the tool can be used as a non-perturbing measuring device.

Having a detailed model of student SRL in the form of a cognitive architecture, plus the ability to detect student metacognitive states, will help us to provide automated intelligent assistance to students at the level of SRL and will lead to the future creation of human-level artificial learners, resulting in the long-awaited breakthrough in AI.

Acknowledgments

This research is supported by the Center for Consciousness and Transformation, George Mason University. We

References

Hymel, A, M., Levin, D. T., Barrett, J, Saylor, M.M., & Biswas, G. (2011). The interaction of childrens' concepts abut agents and their ability to use an agent-based tutoring system. *Proceedings of HCI 2011*.

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning (NCER 2007-2004).* Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Ritchhart, R., Turner, T., & Hadar, L. (2009). Uncovering student's thinking about thinking using concept maps. *Metacognition and Learning* 4: 145-159.

Russell, S. J., & Norvig, P. (2009) *Artificial Intelligence: A Modern Approach* (3rd edition). Prentice Hall.

Samsonovich, A. V. (2009). The Constructor metacognitive architecture. In Samsonovich, A. V. (Ed.). (2009). *Biologically Inspired Cognitive Architectures II: Papers from the AAAI Fall Symposium. AAAI Technical Report FS-09-01*, pp. 124-134. Menlo Park, CA: AAAI Press. ISBN 978-1-57735-435-2.

Samsonovich, A. V., De Jong, K. A., and Kitsantas, A. (2009). The mental state formalism of GMU-BICA. *International Journal of Machine Consciousness* 1 (1): 111-130

Winne, P. H., & Nesbit, J. C. (2009). Supporting self-regulated learning with cognitive tools. In Hacker, D.J., Dunlosky, J., and Graesser, A.C (Eds.). *Handbook of Metacognition in Education*. Mahwah, NJ: Erlbaum.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In Boekaerts, M., Pintrich, P. R., and Zeidner, M. (Eds.). *Handbook of Self-Regulation.* pp. 13-39. San Diego, CA: Academic Press.

Zimmerman, B. J., & Kitsantas, A. (2006). The hidden dimension of personal competence: Self-regulated learning and practice. In Elliot, A. J., & Dweck, C. S. (Eds.). *Handbook of Competence and Motivation*, pp. 509-526. New York: The Guilford Press.

Zimmerman, B. J., and Kitsantas, A. (2007). Reliability and validity of Self-efficacy for Learning Form (SELF) scores of college students. *Journal of Psychology* 215 (3): 157-163.