Exploration of Student's Use of Rule Application References in a Propositional Logic Tutor

Michael Eagle, Vinaya Polamreddi, Behrooz Mostafavi, and Tiffany Barnes
North Carolina State University
890 Oval Dr, Box 8206
Raleigh, NC 27695-8206
{mjeagle, vspolamr, bzmostaf, tmbarnes}@ncsu.edu

ABSTRACT

Many tutors offer students reference material or tips that they can access as needed. We have logged data about student use of references with Deep Thought logic tutor which to understand why and how references are used. We find evidence that students use these references in systematic ways that change over the course of the tutor, and can be predictive of rule application errors. We can use this information to increase our understanding of which concepts students find similar, what times during the tutor students feel the need to use references. Our goal is to eventually incorporate data-driven feedback based on when and how the references are accessed.

1. INTRODUCTION

Tooltips are messages that show up when a user hovers over a GUI element for a small amount of time. It is usually a small box with text that explains what the GUI element does [4]. In educational systems, these messages can contain hints or reference material that are intended to aid the student. Alonso at el. used tooltips to explain semantic relationships in a UML tutor [1]; the authors noted that students often used these tips for verification. White et al expanded upon the concept of tooltips by making them tangible within a augmented reality environment [6].

We want to further explore how these tooltips and reference material are used by students and how they affect student performance to create more effective interventions based on previously-collected student reference usage. We are inspired to add feedback and interventions based on this reference-data by the results of adding next-step hints generated from previously collected student solution-data by Stamper et al [5]. The Deep Thought logic tutor [3] provides students with logic axiom references when students hover over axiom icons within the tutor. We added logging to these hover reference actions to understand how student use of these references affect tutor performance.

We hypothesize that we will find differences in student performance metrics, such as error rate, based on their axiom-reference usage. We also expect that the way the references are used will change as students progress through the tutor.

We find that usage of references before a tutor action corresponded with a larger error rate on that action. We also find that as students progress through the tutor and as the problems increase in difficulty, they tend to use the references more often. Finally, we observe axioms that are referenced in succession indicating the rules students associate with each other and the changes in rule association as students progress through the tutor. These observations point towards trends in the collected reference data that provide better understanding of student actions and will allow us to create new, potentially better, interventions.

2. METHODS

We collect our data from the Deep Thought propositional logic tutor [3]. Each problem in Deep Thought provides the student with a set of premises and a conclusion, and asks students to prove the conclusion by applying logical axioms to the premises (see Figure 1). These logical axioms are separated into three groups based on domain concept, delineated with different colors in the tutor: logical inference rules (red), logical replacement rules (blue), and logical equivalence operations (green).

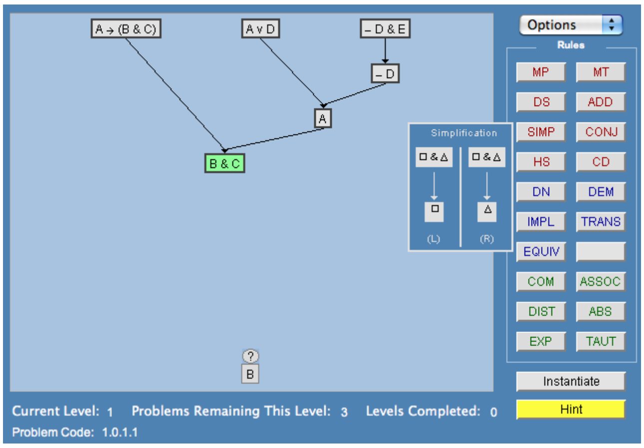

Figure 1: Example screenshot of the Deep Thought tutor.

For example, in Figure 1, a student starts with premises

$A \rightarrow (B \wedge C)$; $A \vee D$; $\neg D \wedge E$ at the top of the proof window, and conclusion B at the bottom. The student performs $SIMP(\neg D \wedge E)$, applying simplification (SIMP) to the premise $\neg D \wedge E$ and derives $\neg D$. This leads to the resulting-state of $A \rightarrow (B \wedge C)$; $A \vee D$; $\neg D \wedge E$; $\neg D$.

Errors are actions performed by students that are illegal operations of rule-application, or illegal operations of the tutor. For example: If the student were to apply simplification to premise $A \vee D$ in the above example, the system would log this interaction as a rule-application error, as simplification is not a valid rule that can be used on a disjoint expression.

2.1 Logical Axiom Reference

By default, axiom references show up when a student hovers over the GUI element (button) representing the logical axiom for two seconds with the mouse pointer. References are given as an overlay pop-up window next to the corresponding axiom button, displaying the axiom name, a visual representation of the axiom pre- and post-conditions with operands and geometric shapes as variables, and valid direction of axiom application (one-way implication or two-way equivalence). Examples of these references for the axioms HS, MP, and MT are provided in Figure 2, and Figure 1 shows how these look in the main window for simplification (SIMP).

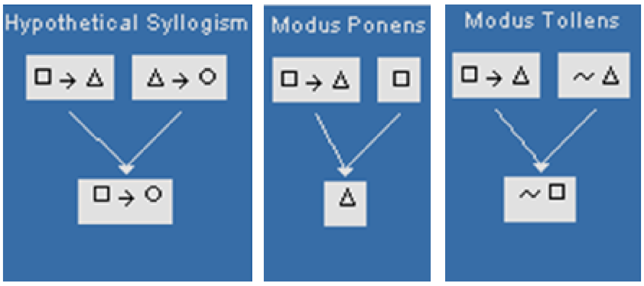

Figure 2: Hover-hints are given as an overlay pop-up window next to the corresponding axiom button, displaying the axiom name, a visual representation of the axiom pre- and post-conditions with operands and geometric shapes as variables

2.2 Data Preparation

Each row of data logged in Deep Thought represents an action performed by a student. Every reference is recorded as a separate action. For the purpose of this study, the data was pre-processed so that each transaction has a reference sequence of rules that were referenced prior to it. If there are no references preceding an action, an empty list and the corresponding 0 is recorded. In the following example (table 1, the references on row 3 and 4 are compressed down onto row 5 and the hover on row 7 is placed with the following action on row 8.

To perform analysis to understand whether hovers are correlated to student performance and errors, the number of hovers in a list and the error of the following action are correlated and tallied. Due to the desire to understand whether or not the number of rules hovered over has an impact, and not the impact of the number of hover instances, the hover lists and the corresponding count are then updated to only include unique rule instances. Due to only including number of unique rules hovered over, the number of hovers value only ranges from 0 to 19 as there are only 19 rules.

Table 1: Reference sequences are constructed from each reference made before performing an action.

User	Ref	Rule
a1	N	HS
a1	N	CD
a1	Y	DS-
a1	Y	MP-
a1	N	MT
b1	N	DS
b1	Y	CD-
b1	N	HS

Ref-Seq	# refs	User	Rule
[]	0	a1	HS
[]	0	a1	CD
[DS, MP]	2	a1	MT
[]	0	b1	DS
[CD]	1	b1	HS

A zero in the error column indicates a valid action while the code 1, 3 and 6 indicate rule-application errors. As shown in the following example, the error values for all the occurrences for each number of hovers is tallied according to whether it indicated an error or not. For example, for the number of hovers of value 1, there are three occurrences; one occurrence has an error code of 0 while two have error codes of 1 and 3. The two with the error codes of 1 and 3 are tallied under the # of errors column and the 0 is tallied in the number of non-errors (see table 2.)

Table 2: Preparation of sequence size vs. errors

Ref lists	#Refs	Err
[]	0	0
[DS, MP]	2	1
[]	0	1
[CD]	1	3
[MT,HS,CD]	3	0
[ASSOC]	1	0
[MT,MT,DS]	3	0
[ABS]	1	1

#Refs	Err	Other
0	1	1
1	2	1
2	1	0
3	0	2

3. RESULTS

Deep Thought was assigned as a mandatory assignment in a philosophy deductive logic course. Six ordered levels of problems were assigned for full completion of the tutor, with each level comprising of 2–3 problems related to specific logical rules or proof problem-solving concepts as dictated by the course curriculum. Completion of an entire level of problems is required for assignment credit. Deep Thought is run as an on-line web applet, with students allowed to work unobserved through the problem sets at their own pace throughout the semester. There are a total of 47 students who were logged as using the system for the class; students who have no log data as a result of an early course drop, or students who did not attempt the assignment are removed from the data set. Three students were removed from the data, as these students altered default system preferences in order to have the hover-hints show up after 0 seconds of hovering, which caused problems with data collection; the normal setting is 2 seconds. This results in a total of 44 students used in this data set.

To explore the connection between the use of references and rule application errors we calculated the error rate for differ-

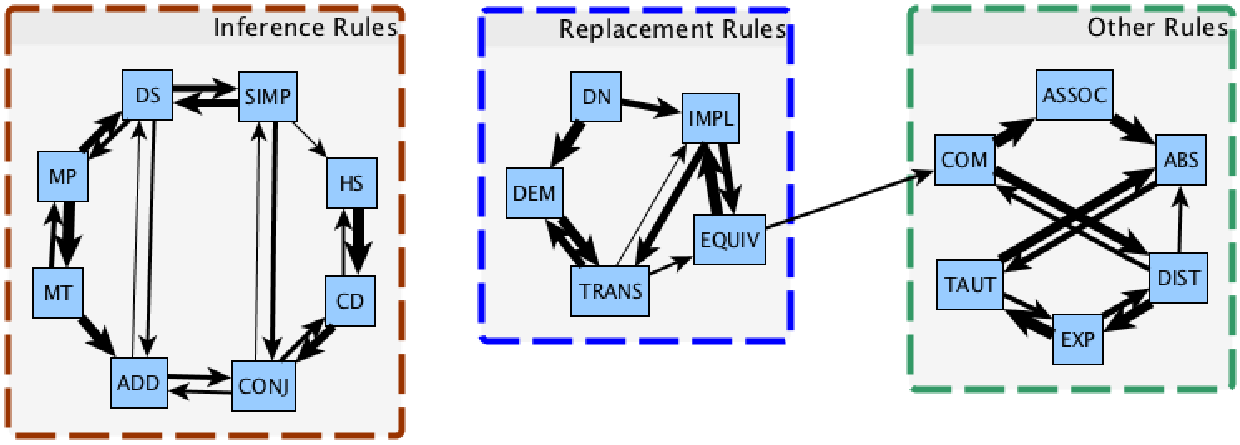

Figure 3: Graph representation of the bigrams of rule reference sequences. The edge thickness is weighted by the strength of the connection, with edges only drawn if the bigram is greater than 20%. Note how the references are generally contained within rule categories.

ing numbers of reference use. We tallied the errors and other interactions, the error rate is calculated for each number of references as follows,

$$ErrorRate = \frac{NumOfRefs}{NumOfRefs + OtherInteractions}. \quad (1)$$

The data is then binned by the number of rules within a string of consecutive hovers (0, 1, 2, 3–4, 5–19). The 0 bin is chosen to capture where references aren't used at all. The 1 bin is chosen to catch the behavior of having an action in mind, using the hover-reference as verification, and the 2 bin is chosen to catch the behavior of confusion between two rules. The 3–4 bin is chosen to catch the student with slightly more confusion, traversing rules based on spacial proximity. The >5 bin is chosen to capture referencing behavior that extended over the previous buckets. The results are presented in Table 3.

Table 3: Error rate after a number of references.

Hovers	Errors	Interactions	Error Rate
0	1201	22907	5.24%
1	104	889	11.70%
2	30	181	16.57%
3–4	14	132	10.61%
5–19	32	173	18.50%

To explore evidence of systematic reference behaviors, we analyze reference sequences as bigrams [2]. The sequences were used to generate counts of pair of rules referenced in secession which were then used as an adjacency matrix where the percentages represent edge weights (Figure 3). Edges are only drawn for bigrams greater than 20%.

In order to explore differences in how the references are used as students progress through the tutor, we generated the bigrams for all rules separated along conceptual shifts in the tutor (every two levels). We present all rules with a bigram greater than 20% during levels 1–2 in the first column of Table 4, the next two columns represent the change in bigram association from one level to the next (+ indicates that the rule is now past the 20% threshold, while - indicates that it is no longer above the threshold.)

Table 4: Rules and their associations over the course of the tutor.

Rule	1–2	3–4	5–6
ADD	CONJ, DS, MT	-DS, -MT	
CD	CONJ, HS		
CONJ	HS, SIMP	+ADD, +CD, -HS, -SIMP	-ADD
DS	ADD, MP, SIMP	-ADD	-MP
HS	CD, SIMP		-SIMP
MP	DS, MT		
MT	ADD, MP	+DS	
SIMP	CONJ, DS	+HS	
DEM	DN, TRANS	-DN	
DN	DEM, HS,	-HS, +IMPL	
EQUIV	COM, IMPL,		
IMPL	EQUIV, TRANS		
TRANS	DEM, EQUIV, IMPL		-IMPL

To analyze the use of references through the duration of the tutor, the interactions are broken up by tutor level and split into whether they were actions taken after the use of references or were actions without prior referencing. Table 5 shows the results of this analysis across the six assigned levels.

Table 5: Use of references before actions increases as student progress through the tutor.

	1	2	3	4	5	6
Actions	8256	2771	5110	5098	4424	2493
w/o ref	7862	2655	4779	4756	4082	2260
w/ ref	394	116	331	342	342	233
Percent	4.77%	4.19%	6.48%	6.71%	7.73%	9.35%

4. DISCUSSION

From Table 3 we find that usage of references before a tutor action corresponds with a larger error rate on that action. While this indicates that references alone do not seem to help students that reference perform better than those that don't reference, it does indicate that students that reference are more likely to make an error so the start of a reference sequence could be an effective place for an intervention. The number of errors after a single reference was lower than that after 2 or >5 references which could indicate that those that are using references for verification might be less likely to make an error than those that want to reference multiple rules. Investigation into the types of errors students made after two references revealed that 17 out of 30 errors contained only three of the 19 rules (DS, MP, SIMP,) this could indicate an area of confusion. The 3–4 bin has an unexpectedly lower error rate which could be because the students are thoroughly checking between the rules where their confusion lies instead of hastily taking a decision earlier. The bigram analysis supports this extrapolation because it shows that for any particular rule there are only a few strong connections, so students that reference 3–4 rules might be hovering over all the closely associated rules.

Figure 3 represents the bigrams of reference sequences with significant edge weights where the chance that any student would reference the second rule immediately after referencing the first is over 20%. The resulting graph and its clusters directly correspond to the three rule categories available in the tutor; this indicates that students tend to systematically reference rules within these groups (rather than randomly ask for references.) There are two reasons as to why they are hovering on rules in the shown order. One could be that they realize that the rules are related and are referencing them to learn the differences. Another possible reason could be that they are referencing rules in geographic proximity and the rules happen to be related since the tutor was designed to have related rules in one geographic area. The second explanation would indicate that the placement of references in a tutor is important to their utilization.

Table 4 provides evidence that students change which rules they frequently seek references for together over the course of the tutor. For example in Problems 1–2, after getting a reference on ADD the student is likely to seek reference on CONJ, DS, and MT; however in later levels the student is likely to only seek reference on CONJ. This shows some degree of learning, as the DS and MT rules are not very related to the ADD rule. The additions of rule associations can also be an indication of learning as the students are making additional connections as they progress. The rule associations shown are a promising location for new interventions where if the students are found hovering between unrelated rules, they can be shown a more effective hint that allows them to learn more about the references to clear the confusion.

We also found that as students progress through the tutor, and the problems increase in difficulty, they tend to use the references more often (Table 5). The students in the first two levels only consult the references between 4–5% of the time before choosing an action. This compares to students in the later levels using the references before 8–9% of their actions. The increase in levels 5 and 6 could be because the students must use a larger number of axioms to solve these problems, but the increase in levels 3 and 4 indicates that as the levels get more difficult, students feel the need to reference more before taking an action. The decrease for level 2 can be explained by the fact that the increase in difficulty between 1 and 2 is less steep than the increase between the other levels. From these observations, it can be extrapolated that an increase in the percentage of actions taken after referencing indicates that students feel a sense of higher difficulty which can also be useful information in making effective interventions.

5. CONCLUSIONS AND FUTURE WORK

We have found that usage of references before a tutor action corresponded with a larger error rate on that action, and as students progress through the tutor, they tend to use references more often with increased problem difficulty. Using this information we can aid students working through the tutor by using reference usage as an indicator for possible didactic intervention; offering feedback when the system determines a student having difficulty in tutor by their behavior with references.

We have also observed which axioms students tend to seek references for at the same time, revealing some changes in rule association as students become more familiar with the tutor. Using this information coupled with existing knowledge of problem parameters, we can determine which concepts students demonstrate difficulty understanding in the context of a particular problem, and provide feedback for those students accordingly.

6. REFERENCES

[1] M. Alonso, D. Py, and T. Lemeunier. A learning environment for object-oriented modeling, supporting metacognitive regulations. In *Advanced Learning Technologies, 2008. ICALT'08. Eighth IEEE International Conference on*, pages 69–73. IEEE, 2008.

[2] M. J. Collins. A new statistical parser based on bigram lexical dependencies. In *Proceedings of the 34th annual meeting on Association for Computational Linguistics*, pages 184–191. Association for Computational Linguistics, 1996.

[3] M. J. Croy. Graphic interface design and deductive proof construction. *J. Comput. Math. Sci. Teach.*, 18:371–385, December 1999.

[4] S. B. Shneiderman and C. Plaisant. Designing the user interface 4 th edition. *ed: Pearson Addison Wesley, USA*, 2005.

[5] J. Stamper, M. Eagle, T. Barnes, and M. Croy. Experimental evaluation of automatic hint generation for a logic tutor. *International Journal of Artificial Intelligence in Education*, 22(1):3–17, 2013.

[6] S. White, L. Lister, and S. Feiner. Visual hints for tangible gestures in augmented reality. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 1–4. IEEE Computer Society, 2007.