Bringsjord, S., Shilliday, A., Taylor, J., Bello, P., Yang, Y., & Arkoudas, K. (2006). Harnessing intelligent agent technology to "superteach" reasoning. *International Journal of Technology in Teaching and Learning, 2*(2)*, 88*-116.

Harnessing Intelligent Agent Technology to "Superteach" Reasoning

Selmer Bringsjord, Andrew Shilliday, Josh Taylor, Paul Bello, Yingrui Yang, & Konstantine Arkoudas
Rensselaer Polytechnic Institute

After briefly explaining our ultimate educational goal with respect to reasoning (to "superteach" reasoning), and our theoretical foundation, we give an overview of some of our attempts to build and harness intelligent agents in order to reach this goal. We end with coverage of the Slate system, which inherits its power from lessons learned in connection with the engineering of its more primitive predecessors.

Keywords: Reasoning, Logic, Intelligent Agents

WHAT IS SUPERTEACHING?

Our overall goal is to harness the power of intelligent agents in order to "superteach" both formal and informal reasoning. (Without descending into details that would hinder exposition, you can for now identify 'formal reasoning' with reasoning that takes place in some formal logic, and 'informal reasoning' with reasoning that is closely aligned with argumentation expressed in natural languages like English.) We define superteaching as teaching that:

1. is accessible to students physically separated from their instructors;
2. features instructors that have a high degree of "presence," which is to say that instructors can interact directly with their students, allowing them to leverage such things as direct eye contact, gesticulation, facial expression, tone of voice, rapid-fire question-and-answer, etc.;
3. maximizes the effectiveness of human instructors by augmenting or (as in the case of supertaught reasoning) replacing them with supporting intelligent agents;
4. allows a much higher instructor (human)/student ratio than conventional education; and
5. is more effective, in terms of learning, than conventional instruction.

Accordingly, we say that superteaching is a form of education defined by simultaneously high levels of five parameters, each of which is associated with a member

Selmer Bringsjord is a Professor at Rensselaer Polytechnic Institute, Andrew Shilliday and Josh Taylor are doctoral students at Rensselaer Polytechnic Institute, Paul Bello is a researcher for Air Force Research Labs, Yingrui Yang is an Associate Professor at Rensselaer Polytechnic Institute, and Konstantine Arkoudas is an Assistant Professor at Rensselaer Polytechnic Institute. Please contact Dr. Bringsjord at the Department of Cognitive Science at Rensselaer Polytechnic Institute, Troy, NY 12180-3590. E-mail: selmer@rpi.edu

of the quintet just given. Using labels that follow the quintet in an obvious way, the parameters are: Accessibility, Presence, Automation, Throughput, and Effectiveness. Our research and development can therefore be viewed as the construction of technology and techniques that, together, through time, moves from lower levels in this quintet of parameters to higher ones. If we assume that the range of values for these parameters can range from 1 (lowest) to 3 (highest), then, for the teaching of reasoning, both the current state and our goal state is approximately as shown in Table 1.

Table 1. Abstract View of Current State and Goal State for Teaching Reasoning.

Domain/Area	Accessibility	Presence	Automation	Throughput	Ped. Effectiveness	
currently:	Reasoning	1	2	1	1	2
future:	Reasoning	3	3	3	3	3

We will first give an overview of elementary logic, familiarity with which is needed in order to understand our efforts to build intelligent agents in the reasoning realm. Next, we will summarize the theoretical foundation on which our efforts are based. Following this summarization we describe some of the intelligent agents we have built in pursuit of superteaching reasoning. The focus in this section is on the Slate system, which can be viewed as the culmination of many years of research and development in our laboratory. Finally, we will describe what we see as the future of our engineering.

WHAT IS REASONING?

Logic is the science of reasoning, and to genuinely convey our attempt to superteach reasoning, it's necessary that the reader have some understanding of logic. In light of this, we provide a self-contained overview of elementary logic drawing from Bringsjord and Yang (2003c). Our coverage of logic is motivated by a desire to systematically "crack" what are called logical (or cognitive) illusions, and we move now straightaway to the first of these puzzles.

THE FIRST ILLUSION

To begin, carefully consider the following illusion, from (Johnson-Laird & Savary, 1995). Variations are presented and discussed in Johnson-Laird (1997a):

Illusion 1.

1. If there is a king in the hand, then there is an ace, or else if there isn't a king in the hand, then there is an ace.
2. There is a king in the hand.

Given these premises, what can one infer?

Almost certainly your verdict is this: One can infer that there is an ace in the hand. And you rendered this verdict despite the fact that we entitled the problem '*Illusion* 1,' which no doubt at least to some degree warned you that something unusual was in the air. Our suspicion is that good meta-reasoners, given this title, infer that the answer can't be what it seems to be, and hence don't answer with "Ace." However, this sort of meta-reasoning doesn't suffice to support the conclusion that there isn't an ace in the hand. Why do we refer to it as an illusion? Because your verdict seems correct, even perhaps *obviously* correct, and yet a little logic suffices to show that not only are you wrong, but

that in fact what you *can* infer is that there *is not* an ace in the hand!

You should not feel bad about succumbing to Illusion 1; after all, you have a lot of company. Johnson-Laird has recently reported that only one person among the many distinguished cognitive scientists to whom we have given [Illusion 1] got the right answer; and we have observed it in public lectures — several hundred individuals from Stockholm to Seattle have drawn it, and no one has ever offered any other conclusion (Johnson-Laird, 1997b). Bringsjord has time and time again, in public lectures, replicated Johnson-Laird's numbers — presented in (Johnson-Laird & Savary, 1995) — *among those not formally trained in logic*. We now provide enough coverage of symbolic logic to display its efficacy in the face of such illusions.

Modern symbolic logic has three main components: one is purely syntactic, one is semantic, and one is metatheoretical in nature. The syntactic component includes specification of the alphabet of a given logical system, the grammar for building well-formed formulas (wffs) from this alphabet, and a proof theory that precisely describes how and when one formula can be proved from a set of formulas. (Frequently the grammar of logic is distinguished from its proof theory, but in the interests of economy herein, we conflate the two.) The semantic component includes a precise account of the conditions under which a formula in a given system is true or false. The metatheoretical component includes theorems, conjectures, and hypotheses concerning the syntactic component, the semantic component, and connections between them. The two simplest and most- used logics are the propositional calculus (which goes by other names as well: e.g., 'propositional logic,' 'sentential logic') and the predicate calculus. The second of these subsumes the first, and is often called 'first-order logic,' or just 'FOL.' We now proceed to characterize the three components for both the propositional calculus and FOL, starting with the former.

PROPOSITIONAL LOGIC

Grammar. The alphabet for propositional logic is simply an infinite list $p_1, p_2, \ldots, p_n, p_{n+1}, \ldots$ of propositional variables (according to tradition p_1 is p, p_2 is q, and p_3 is r), and the five familiar truth-functional connectives $\neg$, $\rightarrow$, $\leftrightarrow$, $\wedge$ and $\vee$. The connectives can at least provisionally be read, respectively, as 'not,' 'implies' (or 'if then'), 'if and only if,' 'and,' and 'or.' In cognitive science and AI it is often convenient to use propositional variables as mnemonics that help one remember what they are intended to represent. For an example, recall Illusion 1. Instead of representing 'There is an ace in the hand' as p_i, for some $i \in N = \{0, 1, 2, \ldots\}$, it would no doubt be useful to represent this proposition as A, and we employ this representation below. Now, the grammar for propositional logic is composed of the following three rules: (1) Every propositional variable p_i is a wff; (2) If φ is a wff, then so is $\neg\varphi$; (3) If φ and ψ are wffs, then so is $(\varphi \star \psi)$, where $\star$ is one of $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$. (We allow outermost parentheses to be dropped.) This implies, for example, that $p \rightarrow (q \wedge r)$ is a wff, while $\rightarrow q$ isn't. To represent the declarative sentence 'If there is an ace in the hand, then there is a king in the hand' we can use $A \rightarrow K$.

Syntactic proofs (proof theory). A number of proof theories are possible. One such system is an elegant Fitch-style system of natural deduction, *F*, fully explained in (Barwise & Etchemendy, 1999). (Such systems are commonly referred to simply as "natural" systems.) In *F*, each of the truth-functional connectives has a pair of corresponding inference rules, one for introducing the connective, and one for eliminating the connective. Proofs in *F* proceed in sequence line by line, each line number incremented by 1. (Actually, we prefer formats that aren't linearized in this fashion (e.g., the Denotational Proof Language (DPL) known as Athena; see Arkoudas 2000; Arkoudas & Bringsjord, 2004), but we leave this issue aside here. When discussing

Slate later on, DPLs enter the picture.) Each line not only includes a line number, but also a formula (the one deduced at this line) and, in the rightmost column, a rule cited in justification for the deduction. We use a vertical ellipsis

⋮

to indicate the possible presence of 0 or more lines in the proof.

Here is the rule for eliminating a conjunction:

⋮	⋮	⋮
k	$\varphi \wedge \psi$	
⋮	⋮	⋮
m	φ	*k* ^ Elim
⋮	⋮	⋮

Intuitively, this rule says that if at line k in some derivation you have somehow obtained a conjunction $\varphi \wedge \psi$, then at a subsequent line m, one can infer to either of the conjuncts alone. Now here is the rule that allows a conjunction to be introduced; intuitively, it formalizes the fact that two propositions are the case it follows that the conjunction of these two propositions is also true.

⋮	⋮	⋮
k	φ	
⋮	⋮	⋮
l	ψ	
⋮	⋮	⋮
m	$\varphi \wedge \psi$	*k, l* ^ Intro
⋮	⋮	⋮

A key rule in *F* is supposition, according to which you are allowed to assume any wff at any point in a derivation. The catch is that you must signal your use of supposition by setting it off typographically. Here is the template for supposition:

⋮	⋮	⋮
k	φ supposition	
⋮	⋮	⋮

Often a derivation will be used to establish that from some set Φ of propositional formulas a particular formula φ can be derived. In such a case, Φ will be given as suppositions (or, as we sometimes say, *givens*), and the challenge will be to derive φ from these suppositions. To say that φ can be derived from a set of formulas Φ in *F* we follow notational custom and write.

$$\Phi \vdash_F \varphi.$$

When it is clear from context which system the deduction is to take place in, the

subscript on |-can be omitted. Here is a proof that puts to use the rules presented above and establishes that $\{(p \wedge q) \wedge r\} \vdash_F q$:

1	$(p \wedge q) \wedge r$	given
2	$(p \wedge q)$	1 ^ Elim
3	q	2 ^ Elim

Now here is a slightly more complicated rule, one for introducing a conditional. It basically says that if you can carry out a sub-derivation in which you suppose φ and derive ψ you are entitled to close this sub-derivation and infer to the conditional $\varphi \rightarrow \psi$.

⋮	⋮	
k	φ	supposition
⋮	⋮	
m	ψ	
⋮	⋮	
n	$\varphi \rightarrow \psi$	$k - m \rightarrow$ Intro

As we said, in a Fitch-style system of natural deduction, the rules come in pairs. Here is the rule in *F* for eliminating conditionals:

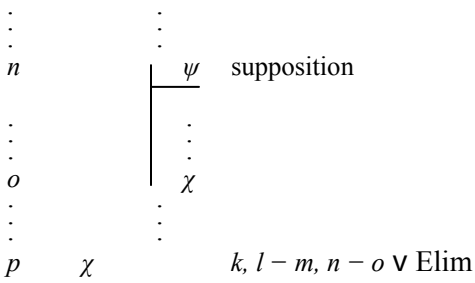
k	$\varphi \rightarrow \psi$	
⋮	⋮	⋮
l	φ	
⋮	⋮	⋮
m	ψ	$k, l \rightarrow$ Elim

Here is the rule for introducing ∨:

⋮	⋮	⋮
k	φ	
⋮	⋮	⋮
m	$\varphi \vee \varphi$	$k \vee$ Intro
⋮	⋮	⋮

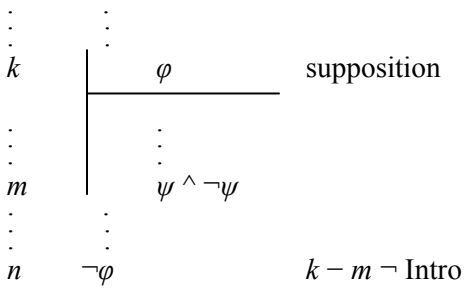
And here is the rather more elaborate rule for eliminating a disjunction:

⋮	⋮	
k	$\varphi \vee \psi$	
⋮	⋮	
l	φ	supposition
⋮	⋮	
m	χ	



The rule ∨ Elim is also known as *constructive dilemma*. The core intuition behind this rule is that if one knows that either φ or ψ is true, and if one can show that χ can be proved from φ alone, and ψ alone, then χ follows from the disjunction.

Next, here is a very powerful rule corresponding to *proof by contradiction* (sometimes called *indirect proof* or *reductio ad absurdum*). Notice that in *F* this rule is ¬*Intro*.



Sometimes a natural deduction system can be a little obnoxious, because by insisting that inference rules come exclusively in the form of pairs for each truth-functional connective, it leaves out certain rules that are exceedingly useful. Two examples are *modus tollens* and DeMorgan's Laws. The former rule allows one to infer ¬*φ* from *φ* → *ψ* and ¬*ψ*. This rule can be established through a proof in *F*, as is shown in Figure 1. This figure shows a screen shot of the completed proof as constructed in the HYPERPROOF proof construction environment, which accompanies the book by the same name, authored by Barwise & Etchemedy. The core of this proof is *reductio ad absurdum*, or ¬ Intro. DeMorgan's Laws for propositional logic sanction moving from a formula of the form ¬(*φ* ^ *ψ*) to one of the form ¬*φ* ∨ ¬*ψ*, and vice versa. The laws also allow an inference from a formula of the form ¬(*φ* ^ *ψ*) to one of the form ¬*φ* ∨ ¬*ψ*, and vice versa. When, in constructing a proof in *F*, we want to use *modus tollens* or DeMorgan's Laws, or any number of other timesaving rules, we can make the inference in question, using the rule of *tautological consequence* as a justification. This rule, abbreviated as TAUT CON in HYPERPROOF, is designed to allow the human proof constructor a way to declare that a given inference is obvious, and could with more work be fully specified using only the rules of *F*. HYPERPROOF responds with a check to indicate that an attempted inference is in fact correct. As you can see in Figure 2, HYPERPROOF approves of our use of TAUT CON, which, again, corresponds in this case not just to DeMorgan's Law in the first two occurrences of this rule, but to the useful inference of *φ* ^ ¬*ψ* from ¬(*φ* → *ψ*).

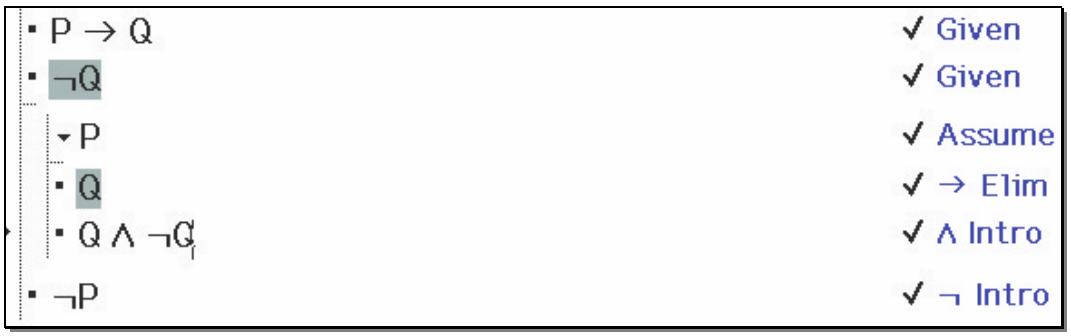


Figure 1. A poof of *modus tollens* in *F*, constructed in HYPERPROOF

We end this section with two more key concepts. A formula provable from the null set is said to be a *theorem*, and where φ is such a formula, we follow custom and write

$$|\text{-}\ \varphi$$

to express such a fact. Here are two examples that you should pause to verify in your own mind: |- $(p \wedge q) \rightarrow q$; |- $(p \wedge \neg p) \rightarrow r$. We say that a set Φ of formulas is *syntactically consistent* if and only if it's not the case that a contradiction can be derived from Φ.

*Semantics (Truth Tables).*The precise meaning of the five truth-functional connectives of the propositional calculus is given via truth-tables, which tell us what the value of a statement is, given the truth-values of its components. The simplest truth-table is that for negation, which informs us, unsurprisingly, that if φ is *T* then $\neg\varphi$ is *F* (first row below double lines), and if φ is *F* then $\neg\varphi$ is *T* (second row).

φ	$\neg\varphi$
T	*F*
F	*T*

Here are the remaining truth-tables.

φ	ψ	$\varphi \ni \psi$
T	*T*	*T*
T	*F*	*F*
F	*T*	*F*
F	*F*	*F*

φ	ψ	$\varphi (\psi$
T	*T*	*T*
T	*F*	*T*
F	*T*	*T*
F	*F*	*F*

φ	ψ	$\varphi \rightarrow \psi$
T	*T*	*T*
T	*F*	*F*
F	*T*	*T*
F	*F*	*T*

φ	ψ	$\varphi \leftrightarrow \psi$
T	*T*	*T*
T	*F*	*F*
F	*T*	*F*
F	*F*	*T*

Notice that the truth-table for disjunction says that when both disjuncts are true, the entire disjunction is true. This is called *inclusive* disjunction. In *exclusive* disjunction, it is one disjunct or another, but not both. This distinction becomes particularly important if one is attempting to symbolize parts of English (or any other *natural language*). It would not do to represent the sentence

George will either win or lose.

as

$$W \vee L,$$

because under the English meaning there is no way both possibilities can be true, whereas by the meaning of $\vee$ it would be possible that *W* and *L* are *both* true. (As we shall soon

see, inclusive versus exclusive disjunction is a key distinction in cracking Illusion 1.) We could use $\vee_x$ to denote *exclusive disjunction*, which we can define through the following truth-table.

φ	ψ	$\varphi \vee_x \psi$
T	*T*	*F*
T	*F*	*T*
F	*T*	*T*
F	*F*	*F*

Before concluding this section, it is worth mentioning another issue involving the meaning of English sentences and their corresponding symbolizations in propositional logic: the issue of the "oddity" of *material conditionals* (formulas of the form $\varphi \rightarrow \psi$). Consider the following English sentence.

> If the moon is made of green cheese, then Dan Quayle will be the next President of the United States.

Is this sentence true? If we were to ask "the man on the street," the answer would likely be "Of course not!" — or perhaps we would hear: "This isn't even a meaningful sentence; you're speaking nonsense." These responses are quite at odds with the undeniable fact that when represented in the propositional calculus, the sentence turns out true. Why? The sentence is naturally represented as

$$G \rightarrow Q.$$

Since *G* is false, the truth-table for $\rightarrow$ classifies the conditional as true. Results such as these have encouraged some to devise better (but much more complicated) accounts of the "if – then's" seen in natural languages (e.g., see Goble 2001). These accounts will be beyond our purview here, however: we will not embark on such a search, and will be content with the conditional as defined by the customary truth-table for $\rightarrow$ presented above.

Given a truth-value assignment *v* (i.e., an assignment of T or F to each propositional variable p_i), we can say that *v* "makes true" or "models" or "satisfies" a given formula φ; this is typically written:

$$v \models \varphi.$$

Some formulas are true on all models. For example, the formula $((p \vee q) \wedge \neg q) \rightarrow p$ is in this category. Such formulas are said to be valid and are sometimes referred to as validities. To indicate that a formula φ is valid we write

$$\models \varphi.$$

Another important semantic notion is *consequence*. An individual formula φ is said to be a consequence of a set Φ of formulas provided that all the truth-value assignments on which all of Φ are true is also one on which φ is true; this is customarily written

$$\Phi \models \varphi.$$

The final concept in the semantic component of the propositional calculus is the concept of consistency once again: we say that a set Φ of formulas is *semantically consistent* if and only if there is a truth-value assignment on which all of Φ are true.

Metatheoretical Results, in Brief. At this point it's easy enough to describe some key metatheory for the propositional calculus. In general, metatheory would deploy logical and mathematical techniques in order to answer such questions as whether or not provability implies consequence, and whether or not the reverse holds. When the first direction holds, a logical system is said to be *sound*, and this fact can be expressed in our notation as

If $\Phi \vdash \varphi$ then $\Phi \models \varphi$.

Roughly put, a logical system is sound if it is guaranteed that true formulas can only yield (through proofs) true formulas; one cannot pass from the true to the false. When the "other direction" is true of a system it is said to be *complete*; in our notation this is expressed by

If $\Phi \models \varphi$ then $\Phi \vdash \varphi$.

The propositional calculus is both provably sound and complete. One consequence of this is that all theorems in the propositional calculus are valid, and all validities are theorems. This last fact is expressed more formally as:

$\models \varphi$ if and only if $\vdash \varphi$

CRACKING THE FIRST ILLUSION WITH PROPOSITIONAL LOGIC

At this point we have at our disposal enough logic to "crack" Illusion 1, to which you should now refer. In this illusion, 'or else' is to be understood as *exclusive* disjunction, even when you make the exclusive disjunction explicit, the results are the same (e.g., you still have an illusion of you use Illusion 1). Therefore using obvious symbolization the two premises become

(1′) $((K \rightarrow A) \vee (\neg K \rightarrow A)) \wedge \neg((K \rightarrow A) \wedge (\neg K \rightarrow A))$
(2′) K

Figure 2 shows a proof in *F*, constructed in HYPERPROOF that demonstrates that from these two givens one can correctly conclude $\neg A$.

Illusion 1

1. If there is a king in the hand then there is an ace, or if there isn't a king in the hand then there is an ace, but not both.

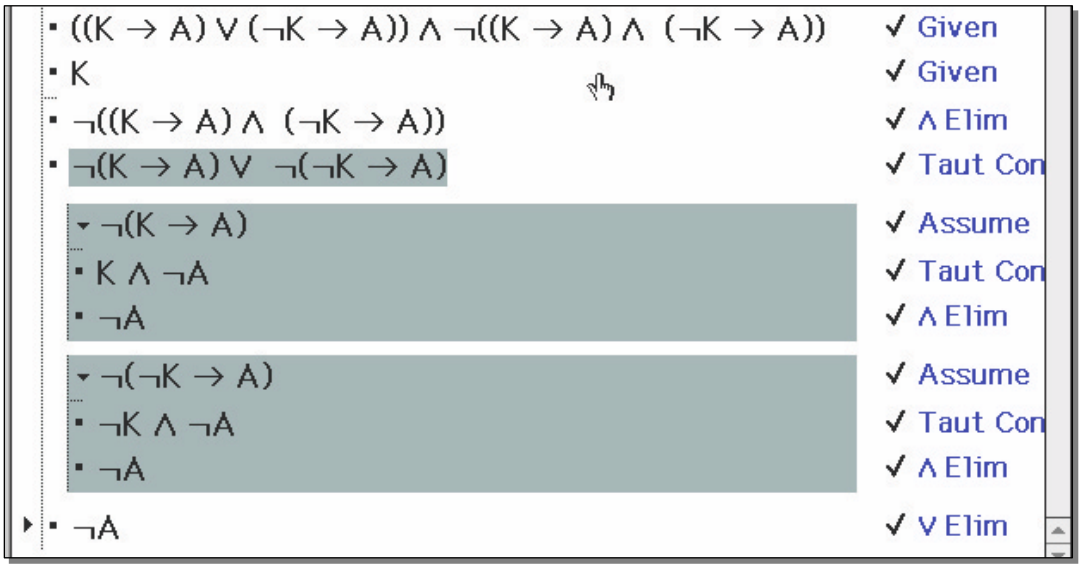


Figure 2. A proof that there is no ace in the hand in *F*

Next, consider a new illusion devised by Bringsjord (actually by his algorithm for devising logical illusions; see (Bringsjord & Yang, 2003a)).

Illusion 2

(3) The following three assertions are either all true or all false:
- If Billy is happy, Doreen is happy.
- If Doreen is happy, Frank is as well.
- If Frank is happy, so is Emma.

(4) The following assertion is definitely true: Billy is happy.

Can it be inferred from (3) and (4) that Emma is happy?

Most subjects answer "Yes," but get the problem wrong — because their reasons for answering with an affirmative are incorrect. They say "Yes" because they notice that since Billy is happy, if the three conditionals are true, one can "chain" through them to arrive at the conclusion that Emma is happy. But this is only part of the story, and the other part has been ignored: viz., that it could be that all three conditionals are false. Some subjects realize that there are two cases to consider (conditionals all true, conditionals all false), and because they believe that when the conditionals are all false one cannot prove that Emma is happy, they respond with "No." But this response is also wrong. The correct response is "Yes," because *in both cases it can be proved that Emma is happy*. This can be shown using propositional logic; the proof, once again constructed in HYPERPROOF, is shown in Figure 3. This proof establishes

$$\{\neg(B \rightarrow D), \neg(D \rightarrow F)\} \vdash E$$

Note that the trick is exploiting the inconsistency of the set $\{\neg(B \rightarrow D), \neg(D \rightarrow F)\}$ in order to get a contradiction. Since everything follows from a contradiction, E can then be derived.

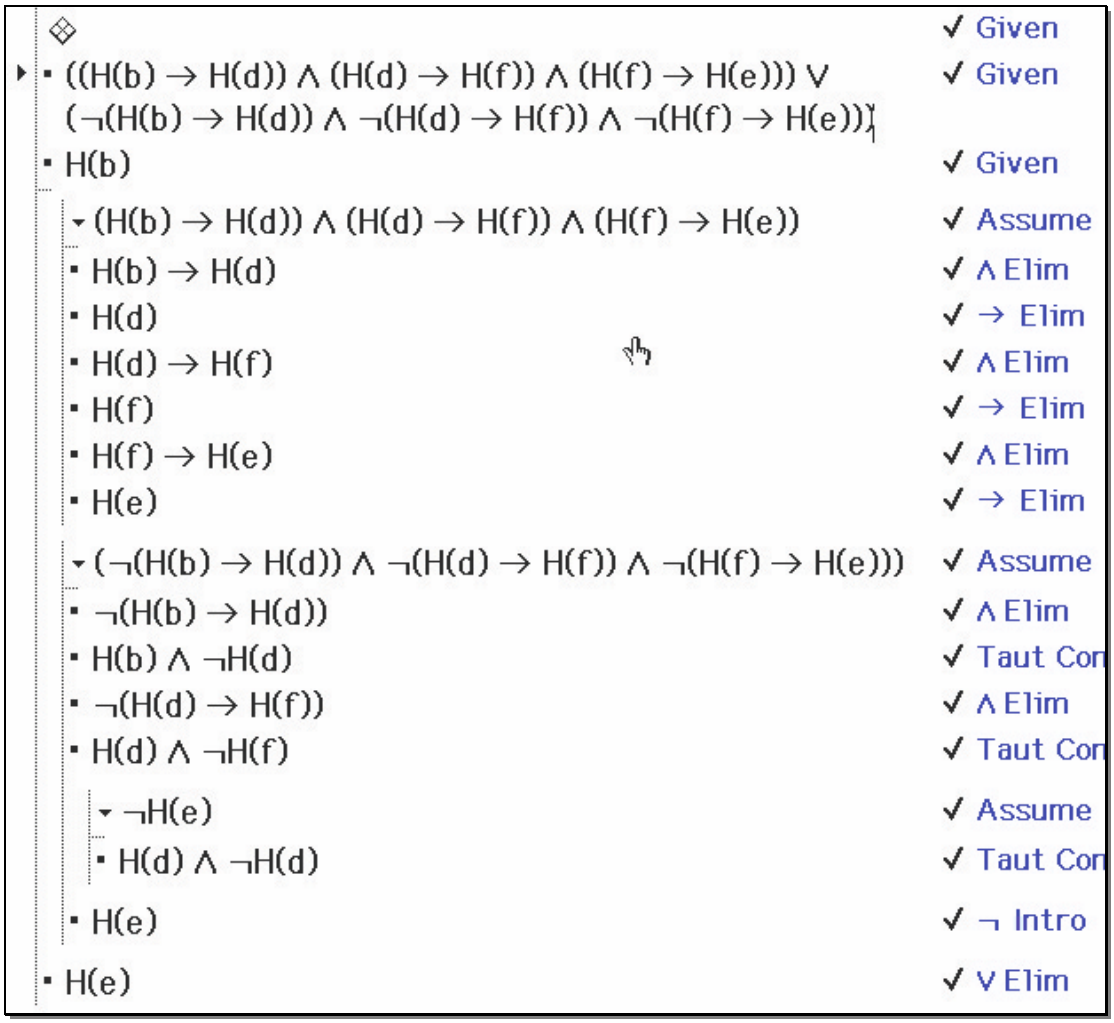


Figure 3. A proof that 'Emma is happy' in *F*

A SECOND ILLUSION; FOL TO THE RESCUE

Now for a second illusion (from: Johnson-Laird & Savary, 1995). What can be inferred from the following two propositions?

(2´´) There is a king in the hand.

Given these premises, what can you infer?

(4) All the Frenchmen in the room are wine-drinkers.
(5) Some of the wine-drinkers in the room are gourmets.

Most subjects respond with

Therefore: (6) Some of the Frenchmen in the room are gourmets.

Alas, (6) cannot be derived from (4) and (5), but the propositional calculus provides insufficiently expressive machinery to reveal why. The best we can do to represent (4) in this logic is to pick and use some propositional variable; *F*, say. But this does nothing to capture the internal structure of the proposition, which is captured by the claim to the effect that, for every *x*, if *x* is a Frenchman, then *x* is a wine-drinker. However, FOL can come to the rescue. We provide now an exceedingly brief overview of FOL, after which we show how FOL can be deployed to crack the second illusion.

FIRST-ORDER LOGIC: AN OVERVIEW

For FOL, our alphabet will now be augmented to include

=	the identity or equality symbol
variables x, y, . . .	like variables in elementary algebra, except they can range of anything, not just numbers
constants c_1 , c_2 , . . .	you can think of these as proper names for objects
relation symbols R, G, . . .	used to denote properties, e.g., W for being a wine-drinker
functors f_1 , f_2 , . . .	used to refer to functions
quantifiers ∃, ∀	the first (existential) quantifier says that "there exists at least one . . .," the second (universal) quantifier that "for all..."
truth-functional connectives (¬, ∨, ^, →, ↔)	now familiar to you, same as in the propositional calculus

Predictable *formation rules* are introduced to allow us to represent propositions like (4), (5), and (6). For example, with these rules, we can now represent (4) as

$$\forall x(Fx \rightarrow Wx),$$

which says that for every thing *x*, if it has property *F* (is a Frenchman), then it has property *W* (is a wine-drinker). Proposition (5) becomes

$$\exists x(Wx \wedge Gx)$$

and (6) is represented as

$$\exists x(Fx \wedge Gx)$$

As in propositional logic, sets of formulas (say Φ), given certain *rules of inference*, can be used to prove individual formulas (say φ); such a situation is expressed by meta-expressions having exactly the same form as those introduced above, e.g., $\Phi \vdash \varphi$. The rules of inference for FOL in such systems as *F* include those we saw for the propositional level, and new ones: two corresponding to the existential quantifier ∃, and two corresponding to the universal quantifier ∀. For example, one of the rules associated with ∀ says, intuitively, that if you know that everything has a certain property, then any particular thing *a* has that property. This rule, known as *universal elimination* (or, sometimes, *universal introduction*) allows us to move from some formula $\forall x\varphi$ to a formula with $\forall x$ dropped, and the variable *x* in φ replaced with the constant of choice. For example, from 'All Frenchman in the room are wine-drinkers,' that is, again,

$$\forall x(F\,x \rightarrow W\,x),$$

we can infer by ∀ Elim that, where a names some particular object,

$$F\,a \rightarrow W\,a,$$

and if we happen to know that in fact *F a* we could then infer by familiar propositional reasoning that Ra. The rule ∀ Elim in *F*, when set out more carefully, is

k	∀*xφ*	
.	.	.
.	.	.
.	.	.
l	*φ*(*a*/*x*)	*k* ∀ Elim

where *φ*(*a*/*x*) denotes the result of replacing occurrences of *x* in *φ* with *a*.

Semantics (Interpretations). FOL includes a semantic side, which systematically provides meaning (i.e., truth or falsity) for formulas. Unfortunately, the formal semantics of FOL gets quite a bit trickier than the truth table-based scheme sufficient for the propositional level. The central concept is that in FOL formulas are said to be true (or false) on *interpretations*; that some formula *φ* is true on an interpretation is often written as *I* |= *φ*. (This is often read, "*I* satisfies, or models, *φ*.") For example, the formula ∀*x*∃*yGyx* might mean, on the standard interpretation for arithmetic, that for every natural number *n*, there is a natural number *m* such that *m* > *n*. In this case, the *domain* is the set of natural numbers, that is, N; and *G* symbolizes 'greater than.' Much more could of course be said about the formal semantics (or *model theory*) for FOL — but this is an advanced topic beyond the scope of the present, brief treatment. For a fuller but succinct discussion using the traditional notation of model theory see Ebbinghaus, Flum and Thomas (1984). The scope of the present discussion does allow us to report that FOL, like the propositional calculus, is both sound and complete; proofs can be found in (Ebbinghaus et al. 1984). This fact entails a proposition that will prove useful momentarily: that if φ isn't a consequence of Φ, then φ cannot be proved from Φ. In the notation introduced earlier, this is expressed as:

$$\Phi \mid\!= \varphi \text{ then } \Phi \not\vdash \varphi$$

CRACKING ILLUSION 2

How does FOL allow us to solve Illusion 2? The simplest solution is to note that we can find an interpretation in which (4) and (5) are true, but (6) is not. This will show that (6) isn't a consequence of (4) and (5), from which it will immediately follow that (6) cannot be proved from (4) and (5). Here is the interpretation we need: Alvin is a wine-drinker and a gourmet, and not a Frenchman. Bertrand is a Frenchman and a wine-drinker, but not a gourmet. No one else, in this imaginary scenario, exists. In this situation, all Frenchmen are wine-drinkers, and there exists someone who is a wine-drinker and a gourmet. This ensures that both (4) and (5) are true. But it's not true that there exists someone who is both a Frenchman and a gourmet. This means that (6) is false; more generally, it means that (6) isn't a consequence of (4) and (5), which in turn means that {(4), (5)} |- (6), and this result solves Illusion 2.

We could also crack Illusion 2 in FOL by showing that all the possibilities for deducing the formula for (6) from those representing (4) and (5) will fail. We could note

first that there is no way, in *F*, at the level of FOL, to deduce (6) from {(4), (5)}; i.e., using the notation we now have at our disposal.

$$\{\ \forall x(Fx \rightarrow Wx),\ \exists x(Wx \wedge Gx)\ \vdash \exists x(Fx \wedge Gx)$$

One way to see this is to note that if we were able to correctly deduce the purported conclusion in FOL, then we would need to be able to first deduce $Fa \wedge Ga$ somehow from the formal versions of (4) and (5), and then deduce $\exists x(Fx \wedge Gx)$ by $\exists$ Intro from this intermediary formula. But how would we be able to infer $Fa \wedge Ga$ in the first place? There is no way to obtain this conjunction.

This concludes our overview of FOL. Now to the theoretical underpinnings of our research.

THEORETICAL UNDERPINNINGS

There are two parts to the theoretical foundation for our attempt to superteach reasoning. We begin by explaining the first part: that we are neo-Piagetians in regard to reasoning. We then briefly describe the second part: our particular theory of human reasoning: *mental metalogic.*

A NEO-PIAGETIAN POSITION ON HUMAN REASONING

While some elements of Piaget's thought remain very much alive today, the consensus seems to be that at least one part has long been reduced to a carcass: the part according to which *F* Humans naturally develop a context-free deductive reasoning scheme at the level of elementary first-order logic.

More specifically, the level here is that of the propositional calculus plus command over some set of simple operations involving the quantifiers 'some' ($\exists$ in FOL) and 'all' ($\forall$ in FOL)). The proposition *F*, or at least a thesis very close to it (more about variants on *F* below), is articulated and defended by Inhelder and Piaget (1958).

As evidence that *F* is generally regarded to be stone cold dead, one can do no better than Peter Wason's (1995) relaxed remarks in his contribution to a book (Newstead & Evans, 1995) written in his honor. Wason is credited with devising the problems that led to the rejection of *F*, and the remarks in question arise from his retrospection on experiments in which subjects received four of these problems as stimuli.

For example, we read: "The first formal experiments, done partly in Scotland, met with grave looks from dedicated neo-Piagetians; the subjects' were clearly incompatible with 'formal operations'" (Wason, 1995; p. 296).

What kind of experiments is Wason referring to? We don't have space to revisit all the puzzles given in these experiments. However, we note one of them, certainly the most famous one: the Wason selection task:

> Suppose that I have a pack of cards each of which has a capital Roman letter written on one side, and a digit from 1 to 9 written on the other side. Suppose in addition that I claim the following rule is true:
>
> • If a card has a vowel on one side, then it has an even number on the other side.
>
> Imagine that I now show you four cards from the pack:
>
> E T 4 7

Which card or cards should you turn over in order to decide whether the rule is true or false?

Only about 5% of the college (!) educated population gives the correct answer, which is E *and* 7. If you said (only) E, you saw that the rule in question would be overthrown were

there to be an odd number on the other side of this card — but you failed to note that if the 7 card has a vowel on the other side, this too is a case that shoots down the rule.

Wason (1995) writes here and elsewhere as if *F* has been long buried; most others in the psychology of reasoning follow suit. For example, the other contributors to the volume in question, each and every one of them, is likeminded: they either explicitly reject or presuppose the falsity of Piaget's *F*. For example, as Johnson-Laird (1995) states "It seems that adult subjects in the selection task have not reached the Piagetian level of formal operations. Yet they are supposed to have attained it around the age of 12." (p. 133)

Put brutally, our response is this. The reason that subjects perform poorly on problems like those given by Wason is that their education is defective; and because their education is defective, they haven't reached Piaget's stage of formal operations. We realize, of course, that *F* uses the term 'naturally,' and we realize as well that this connotes that people, *without special training*, will reach the competence in question. We realize also that part of what we've been calling the 'received view' is that some such term is part of the thesis at stake. For example, in their discussion of the psychology of reasoning, Stillings. Weisler, Chase, Feinstein, Garfield & Rissland (1995) opine that a proposition virtually identical to *F* is "obviously" overthrown by the fact that the vast majority of subjects fail to solve problems like those devised by Wason and Johnson-Laird.

But this is a bluff we are quite willing to call. What, precisely, does 'naturally' mean? We all know that without special training humans are not able to solve even simple arithmetic problems. For example, consider this problem:

> • John is given 3/4 of a chocolate chip cookie. Each of his ten friends will be content if they receive 1/8 of such a cookie. If John is willing to keep none for himself, and he can divide his cookie-part precisely, how many friends can he satisfy?

Even educated adults do poorly on this problem. (At a recent talk at a major university, Bringsjord found, upon presenting this problem, that a goodly number of *professors* had completely forgotten how to divide fractions.) Does poor performance of many subjects on a puzzle like this imply the falsity of some such proposition as the following one?

F_A Humans naturally develop a context-free scheme at the level of elementary arithmetic.

If not, then why should *F* fall? Perhaps, again, the problem pertains not to underlying cognitive development, but rather to education, pure and simple. If someone insists on a rather strict reading of 'naturally,' according to which only a bare minimum of "official" education is required to support the ascription of the adverb naturally, and therefore according to which F_A is indeed taken to be false, then we will be quite content to settle for defending the view that

F' If educated in logic as they are in arithmetic, humans develop a context-free deductive reasoning scheme at the level of elementary first-order logic — a scheme that will allow for the solving of problems like those famously pressed against Piagetian by Wason, Johnson-Laird, and others, and for the solving of significantly harder problems as well.

The intelligent agents we have built are designed to provide the kind of appropriate education in reasoning referred to by F'. We thus see our intelligent agents are vindicating Piaget.

MENTAL METALOGIC

There is now overwhelming empirical evidence for the view that while some human knowledge does seem to be accurately modeled in purely syntactic or symbolic form (the theory of *mental logic* proposes such knowledge (see Rips, 1994; Yang, Braine, & O'Brien 1998). Some knowledge is represented in irreducibly semantic form, or in what we call *mental models* (see Johnson-Laird, 1983; Johnson-Laird, Legrenzi, Girotto, & Legrenzi 2000). Mental models can be pictorial or imagistic in form, rather than symbolic, syntactic, or linguistic. The theory within cognitive science that posits, explains, and empirically justifies the view that human reasoning centrally involves *both* mental logic and mental models is *mental metalogic* (MML) (Rinella, Bringsjord, & Yang, 2001; Yang, Braine, & O'Brien, 1998; Yang & Bringsjord, in-press; Yang & Bringsjord 2001a, 2001b; Yang & Johnson-Laird, 2000a, 2000b). MML is the second part of our foundation for striving to superteach A pictorial overview of MML is shown in Figure 4.

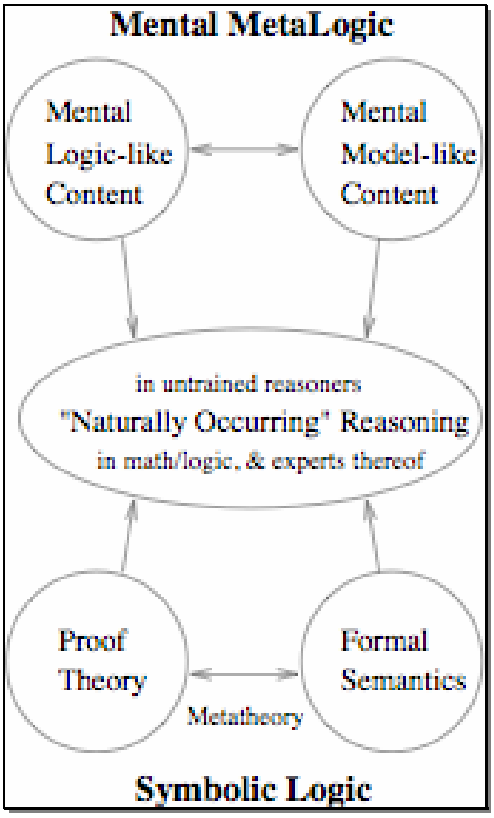


Figure 4. Overview of mental metaLogic symmetry with symbolic logic

On the efficacy of mental MetaLogic. The pedagogical effectiveness of reasoning taught from the standpoint of MML is made plain in (Rinella et al. 2001), which describes a semester-long experiment that provides strong empirical evidence against the view (advocated by Cheng and Holyoak, and many others in psychology and cognitive science) that humans have little context-independent deductive capacity, and that they can't acquire such a capacity through instruction in formal logic. This position is based, in no small part, upon Cheng, Holyoak, Nisbett, & Oliver's (1986) well-known investigation of the efficacy of an undergraduate course in logic in improving

performance on problems related to the Wason Selection Task, in which they found the benefit of such training to be minimal. Our study, involving a similar pre-test/post-test design on a logic class at RPI, suggests a rather more optimistic view on the basis of different results, including an improvement in a group of 100 subjects from 29 correct to 84 correct responses on the tasks most similar to Cheng et al.'s (1986) questions and from seven to 58 on a question involving the use of *reductio ad absurdum*. The hypothesized reasons for this improvement in the effectiveness of logic instruction include the marriage of our new theory of reasoning MML with techniques suggested by this theory (e.g., the technique of disproof via visual counterexamples made possible by HYPERPROOF). Detailed examples of this technique are presented in Rinella et al. (2001). This technique has inspired part of the functionality of the Slate system, an intelligent agent for learning both informal and formal reasoning. (The main purpose of Slate is actually to function as an intelligent agent that assists "professional" informal (e.g., intelligence analysts) and formal (e.g., logicians, mathematicians, and theoretical computer scientists) reasoners, but we don't discuss this side of Slate in the present paper.)

ATTEMPTS AT HARNESSING INTELLIGENT AGENTS

There are two stages in our attempt to superteach reasoning. In the first stage, glimpses of which we now provide in this section, we built agents dedicated to specific topics, and assessed the pedagogical efficacy of these agents. In the second stage, described above, we designed and have now built Slate, a "super"-agent, that is, an intelligent that is an ensemble of many agents working in concert. After describing Slate, we end the paper with a look toward the future.

AN AGENT THAT TEACHES INDIRECT PROOF

We constructed our own serviceable proof construction environment: RIP, the Rensselaer Intelligent Prover. RIP is a simple environment for building natural deduction-style proofs, at the level of the propositional calculus, and without the graphical capability of Barwise and Etchemendy's (1994) HYPERPROOF. We represented the tree-like structure of natural deduction-style proofs with the skeleton code for a directory browser. The rule window of RIP directly parallels the proof structure window and is a view of the rules selected at each step of the way in the proof. In the middle of the main program window is a panel of buttons that provides for all the functionality that the user needs in constructing proofs. On the right hand side, there is a browser window in which lessons are presented in the form of either web pages, or PowerPoint presentations annotated by the agent. See Figure 5 for a visual layout of the interface. The "guts" of RIP is the "industrial strength" theorem prover OTTER (Wos, 1996; Wos, Overbeek, Lusk, & Boyle 1992), which is distributed freely for unrestricted use, and has been used by Bringsjord for many applications. OTTER is a refutation-resolution based theorem proving utility, and has performed exceptionally well in the past in various competitions. OTTER has now been eclipsed by much more powerful ATPs, (see Voronkov, 1995).

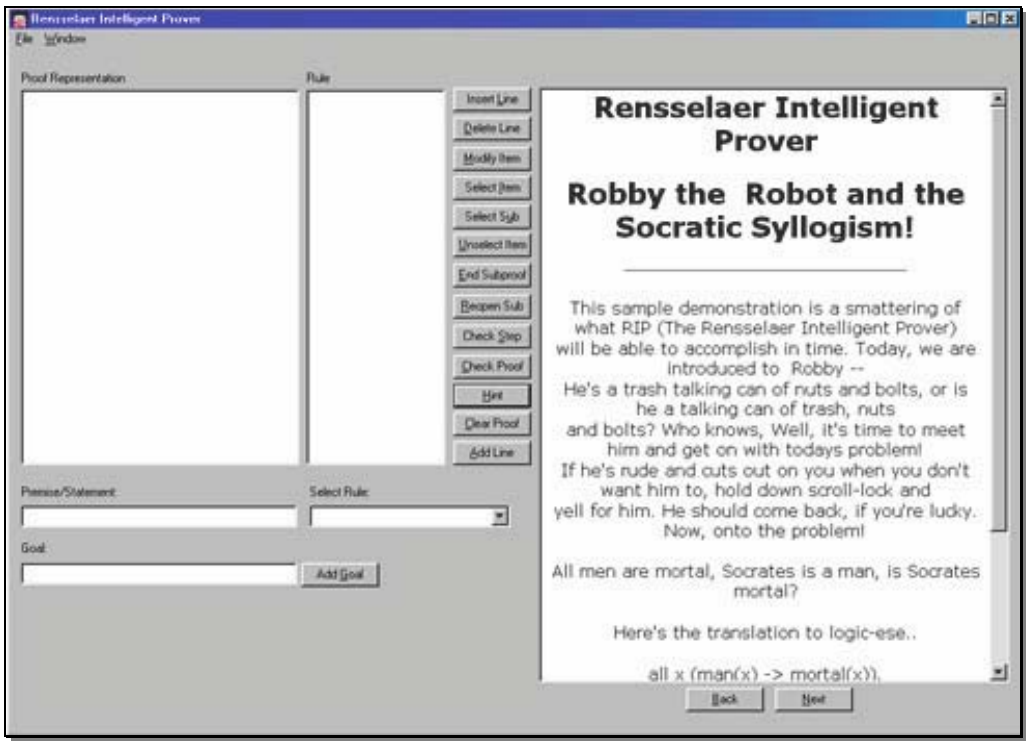


Figure 5. User interface — RIP

In order to augment RIP with intelligent agents that could appear on screen in visual form and give advice and present material that would ordinarily be imparted by a human in lecture or lab sessions, we turned to Microsoft Agent. Microsoft Agent is a programmable desktop caricature that is capable of speaking, moving, gesturing, and of responding to voice input from users. This relatively rich set of behaviors was just what we were looking for, and conveniently enough, all the code required to make the agent take high-level actions is embedded in the web pages that the system displays. Long gone are the days of the Microsoft Office Assistant (the annoying paperclip), who bothers you whenever he possibly can. Robby the robot is our agent of choice, and as of now, he's rather polite. He only speaks when spoken to, and when you tell him you've had enough, he steps out for a while. Robby is shown in Figure 6.

Once RIP was sufficiently mature, we carried out a simple pilot study to begin to assess the efficacy of an artificial agent in teaching students who do not have a physically present instructor. The agent we designed (see Figure 7) was used to teach a specific topic, namely, proof by contradiction (or *reductio ad absurdum*), to students from the RPI *Introduction to Logic* class. Note that this topic is a hard one. As an indication of this, note that while many psychologists of reasoning in the mental logic camp (championed by Braine (1998) and Rips (1994)) concede that untrained humans understand such rules as that from $\varphi \rightarrow \psi$ and φ one can conclude ψ, nearly all psychologists of reasoning believe that untrained humans have absolutely no conception of proof by contradiction (in which, if assuming φ leads to a contradiction ψ э $\neg\psi$, one can infer $\neg\varphi$).

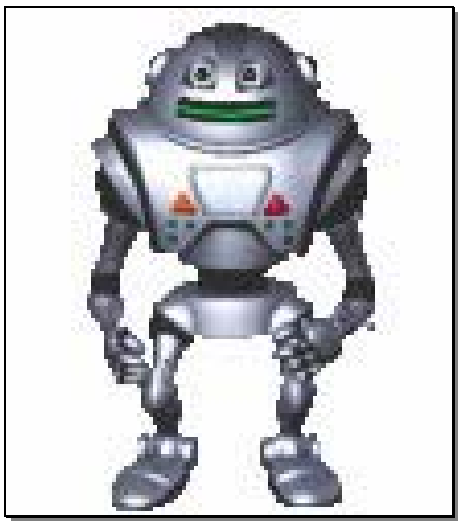


Figure 6. Robby the robot

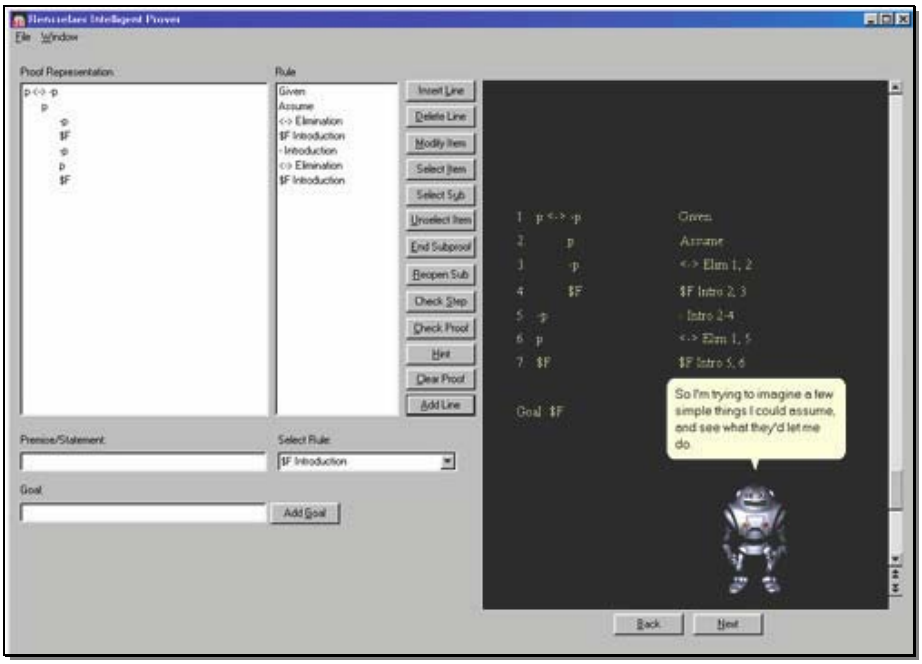


Figure 7. The intelligent proving engine in action

After randomly dividing subjects into two groups matched for reasoning ability by a pre-test, (the 1998 version, differing only in date, was used in our large study: (Rinella et al., 2001). We gave students in the experimental group a brief session in the use of the software, without presenting to them any content. The purpose of this short training was to ensure that the hour they would have later for instruction would not be compromised by a lack of familiarity with the mechanics of using the interface, which was in some ways unlike the other programs used in the course. We then gave each group simultaneous instruction — the experimental group viewed the interactive software; the control received instruction as normal from their professor.

Following the instruction, we gave each participant a post-test consisting of a logic proof that required understanding of the concept discussed in the lessons (again, proof by contradiction). Three of six students in the experimental group received full credit, only one of seven did so from the control, with another earning partial credit. Relatively low attendance rates certainly complicates the statistical interpretation, but the difference was reported at a significance of .092, indicating that it was quite unlikely that in-person instruction was better than instruction by the intelligent agent.

Though we did expect that students would not do significantly better when taught by a physically present human instructor, it came as a welcome surprise that even this early version of a small part of our envisaged system was able to substantially outperform an experienced full professor with over fifteen years of experience teaching this subject (achieving twice the success rate). Needless to say, our sample size is very small. We are progressing now to experiments based on dividing an introductory logic class of approximately 100 students in half (control and experimental), and running the two modes for an entire semester.

COMPARING HUMAN- AND AGENT-BASED TEACHING OF SYLLOGISMS

We first devised a new seven-step algorithm for solving all syllogisms, in the sense that, where *A* is our algorithm and *S* an arbitrary syllogism, *A* takes *S* as input and produces a corresponding formal proof (natural deduction-style, expressible in HYPERPROOF and FITCH) if *S* is valid, and a *dis*proof if *S* is invalid. (A version of the algorithm is given in (Bringsjord & Yang, 2003b)). The disproof includes a diagrammatic situation that serves as a counterexample, that is, a situation in which the premises of *S* are true, while its conclusion is false. (While syllogisms, in the context of even elementary symbolic logic, are easy, to our knowledge no such algorithm has ever been devised, and it is impossible to create such an algorithm on the basis of either purely syntactic schemes (e.g., the mental logic of Rips, 1994) or purely semantic schemes (e.g., the mental models approach of Johnson-Laird, (1983), which by definition cannot offer a syntactic progression of steps in a normal-style proof).

Next, we built a complete agent-based "ecumenical" module on syllogisms, and on our algorithm. (We refer to our modules as 'ecumenical' because they are on topics that are agreed to be part of the canon of introductory logic.) The information in this module is delivered by an intelligent agent who explains syllogisms and associated concepts, as well as the algorithm, and offers interactive exercises and quizzes (see Figure 8). Periodic exercises are given by the agent to students as checkpoints to ensure that they have been paying attention to the material. The lecture will not proceed till assurance is given to the artificial agent that the user has assimilated the material. During the course of the lecture, the intelligent agent adds key terms to a vocabulary list which the student can reference at any point in the lecture (see Figure 9).

At the conclusion of each section of the lecture, a quiz is given to assess student performance and provide feedback as to which steps of the algorithm needed reinforcement (see Figure 10).

With the module built, we proceeded with our between-subjects pilot study. It was a matched design, with random assignments; we used the Raven's (Advanced) Progressive Matrices (Raven, 1962) to match students for reasoning ability. We expected that training provided by the agent would produce performance at least on par with that produced by the human lecturer. Sixteen RPI undergraduates participated in the experiment to fulfill the requirements for the course, Methods of Reasoning. The 16 participants were randomly assigned to two groups (n = 8 for each group). Group 1 learned about syllogisms and the seven-step algorithm from our agent-based module, Group 2 from a

traditional human-delivered lecture.

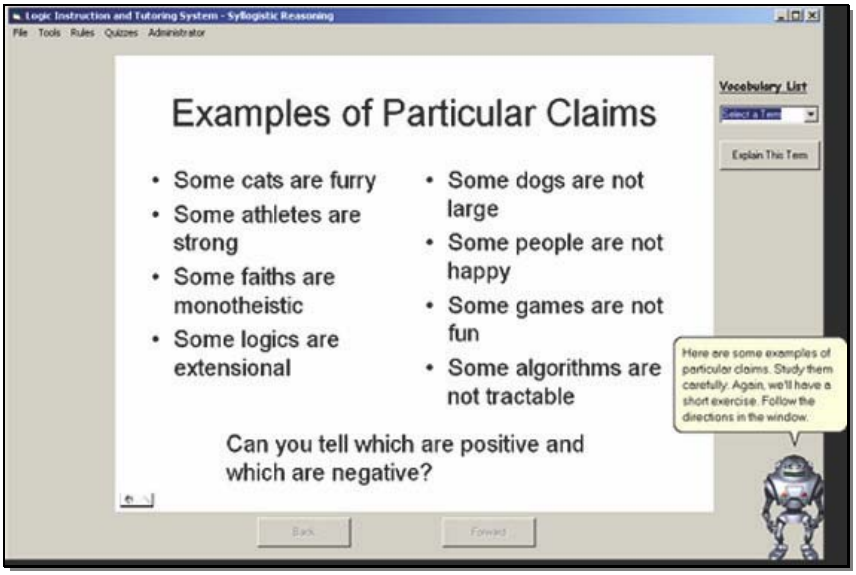


Figure 8. Ecumenical module – lecture

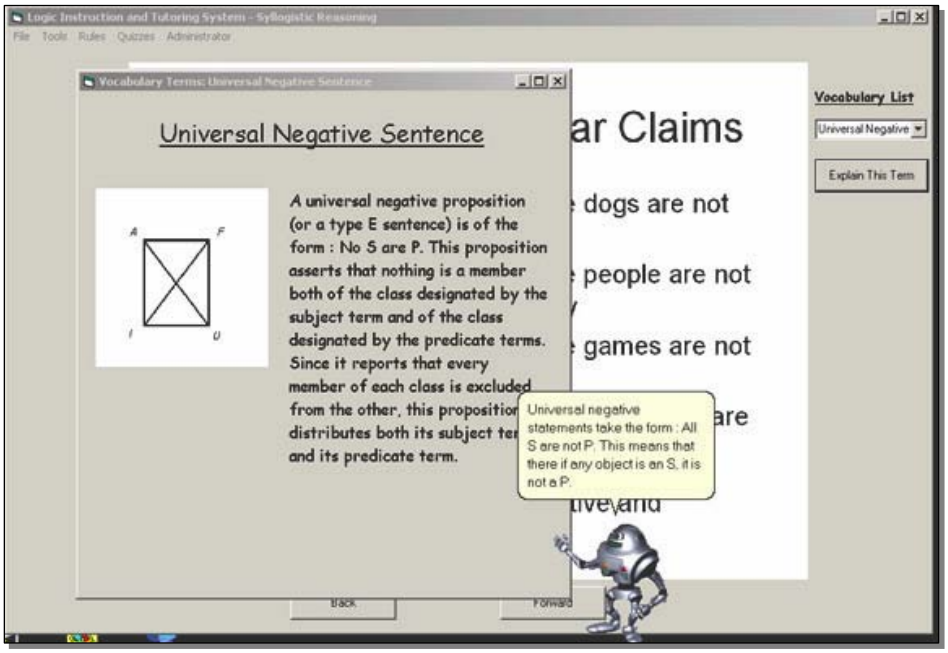


Figure 9. Ecumenical module – vocabulary

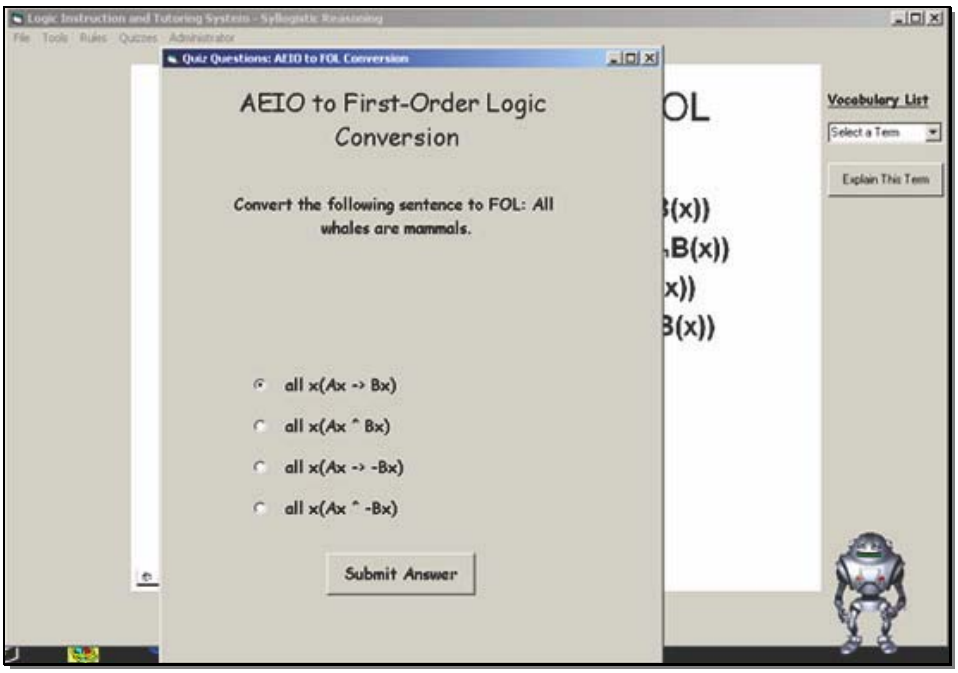


Figure 10. Ecumenical syllogism module – quiz

The same set of four syllogisms was used to test both groups after the training was completed. The post-test included one syllogism of a notoriously difficult type: [Some boats are whales. All boats can walk. Therefore some whales can walk.] The seven-step algorithm, if followed correctly, produces a proof of $\llcorner x(Wh(x) \ni Wk(x))$, which is a first-order formalization of 'Some whales can walk,' from the first two of these three propositions.

Three parameters were considered in evaluating the reasoning capacity: (i) Translation of a problem from English into the formal language of first-order logic; the quality of this translation was scored as 0 (neither premises nor conclusion were presented properly), 1 (only premises presented properly), or 2 (all the three sentences were presented properly). (ii) Intermediate steps, for which the quality was scored as either 0 (no instantiation), 1 (partial instantiation), or 2 (complete instantiation). (iii) Inference steps, for which the quality was scored as 0 if a final answer was incorrect, 1 if a reasonable counterexample was given in the case of an invalid syllogism, or 2 if everything was correct. Thus, two dependent variables were used: the accuracy and the total score of qualities.

The results show that there was no difference in the accuracy between the two groups (the overall means are 3.63 and 3.38). The mean composite scores for the group trained by the lecture was 12.6 and by the automatic agent was 21.5. The difference was significant (Mann-Whiney test, U=14, $p < 0.05$). The results indicate that in problems of the syllogism type, the two kinds of training made no difference in participant's performances (accuracy) but the training by the agent did produce reasoning capacities of a higher quality level (quality scores) than those produced by human lecture. Obviously, the sample size is small; this is only a pilot study. But the result does seem to justify pushing ahead to develop a complete set of ecumenical modules. In particular it will be interesting to see how performance differs in the two modes of training when not only sample sizes of 100 or so are used, but when more complex problem types are used as well.

SLATE, AN INTELLIGENT AGENT FOR TEACHING VERTICALLY INTEGRATED REASONING

We now provide an encapsulated description of a comprehensive, sophisticated intelligent "super"- agent designed to facilitate the learning of formal and informal reasoning. This agent is Slate. Slate is the culmination of many years of preparatory effort designing and building agents of smaller scope, such as those briefly described above, and also more mature systems (e.g., Bello & Bringsjord, 2003).

We should point out that Slate can also perform not merely as a teaching assistant, but also as a *bona fide* on-the-job assistant to professional reasoners in both formal and informal domains. For example, Slate can assist logicians and mathematicians in their work, and can assist intelligence analysts on the job as well. However, since the emphasis in the present paper is on education, we leave aside the professional use of Slate. We provide an overview of Slate by going through, in broad strokes, the progression a student of formal reasoning might go through in order to produce a natural deduction-style proof of theorem in the propositional calculus.

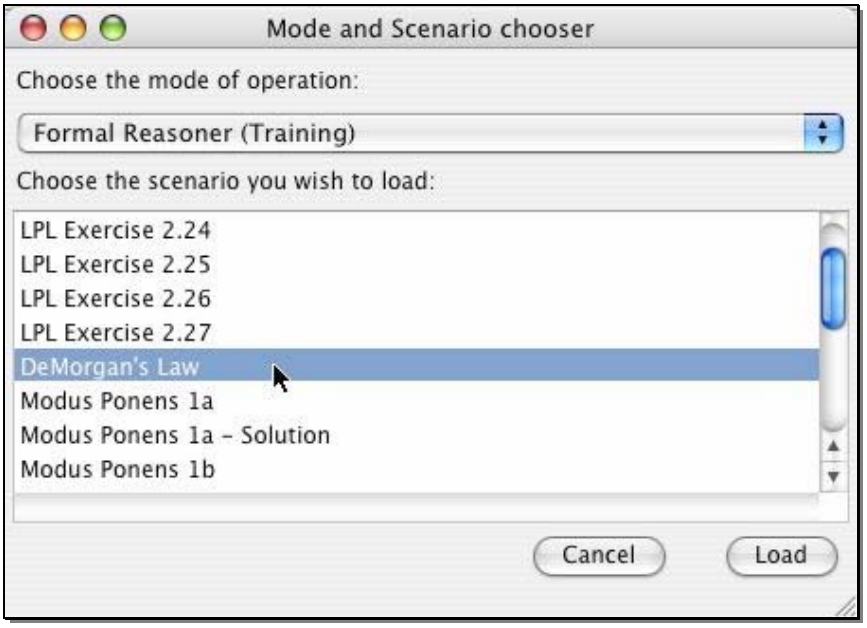


Figure 11. The student can select from library of exercises (formal or informal), or enter professional mode

Using slate to learn formal reasoning. Slate offers students a library of exercises, partitioned between those designed to help students learn formal reasoning, and those designed to help students learn informal reasoning (specifically in connection with intelligence analysis) (see Figure 11). In this section we concentrate on what happens after a student selects a problem from the formal reasoning side. Specifically, assume that the student in this case has selected an exercise in which she is challenged to determine whether or not it's true that

$$\{\neg(A \supset B)\} \vdash \neg A \vee \neg B.$$

If this equation holds, the student must produce a natural deduction-style proof of this theorem. On the other hand, if this provability result cannot be obtained, the student must

produce a counter- example: an interpretation *I* on which ¬(*A* ə *B*) is true, but ¬*A* ∨ ¬*B* is false. If the proof can be found, it must be expressed in a system known as NDL, which is a simplified version of Athena. Both NDL and Athena, designed and built by Arkoudas, are Denotational Proof Languages (Arkoudas, 2000), languages for presenting, checking, and discovering formal proofs that conform to a high-assurance style of programming called "certified computation" (Arkoudas & Rinard, 2004). Astute readers will realize that in this case, a proof can indeed be found: DeMorgan's Law allows a negated conjunction to be transformed into a disjunction, where each disjunct is negated. For present purposes, the reader can understand the student's ultimate objective to be that of presenting a proof of this law in *F*, the rules for which we introduced above. Now let's continue.

Figure 12 shows the situation after the student has selected the exercise. Notice that the relevant propositions appear in Slate's workspace in the form of icons, ready to be inspected and manipulated. Clicking on an icon show, in symbolic form, what proposition the icon represents.

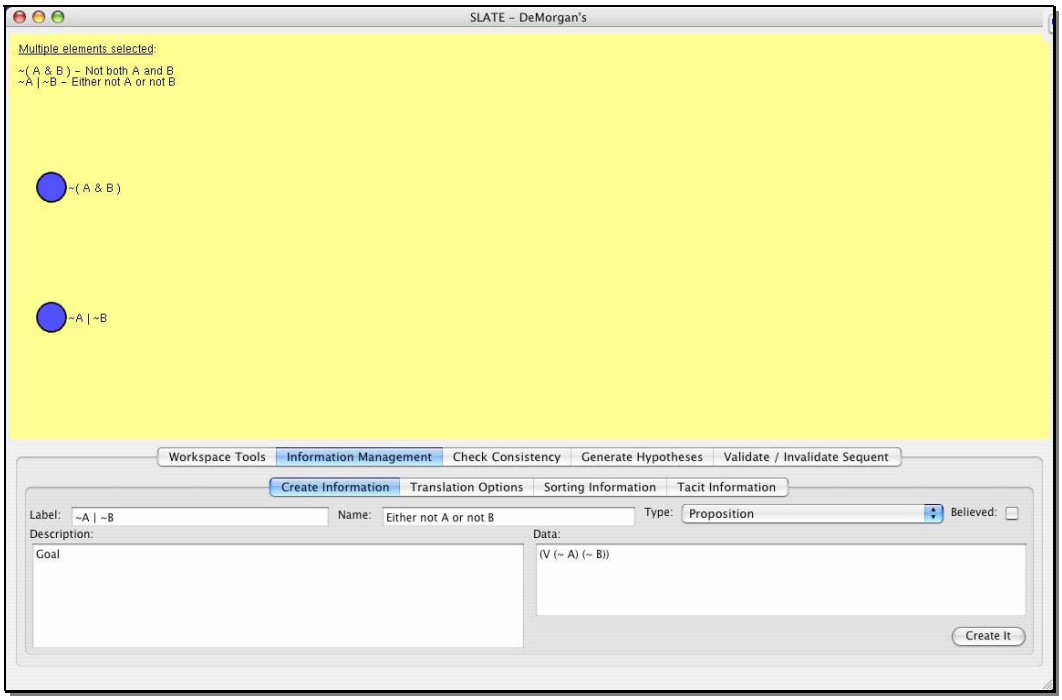


Figure 12. Propositions appear in slate for inspection and manipulation

Slate offers the student an array of systems that can be consulted for suggestions as to how the exercise can be productively tackled. In the example at hand, the student has decided to ask the Oscar system (Pollok, 1995) to try and prove the result. Figure 13 shows a screenshot after Oscar returns an affirmative answer, and a corresponding proof in its own format. The student will use this information when laying out, in diagrammatic form, a proof of her own.

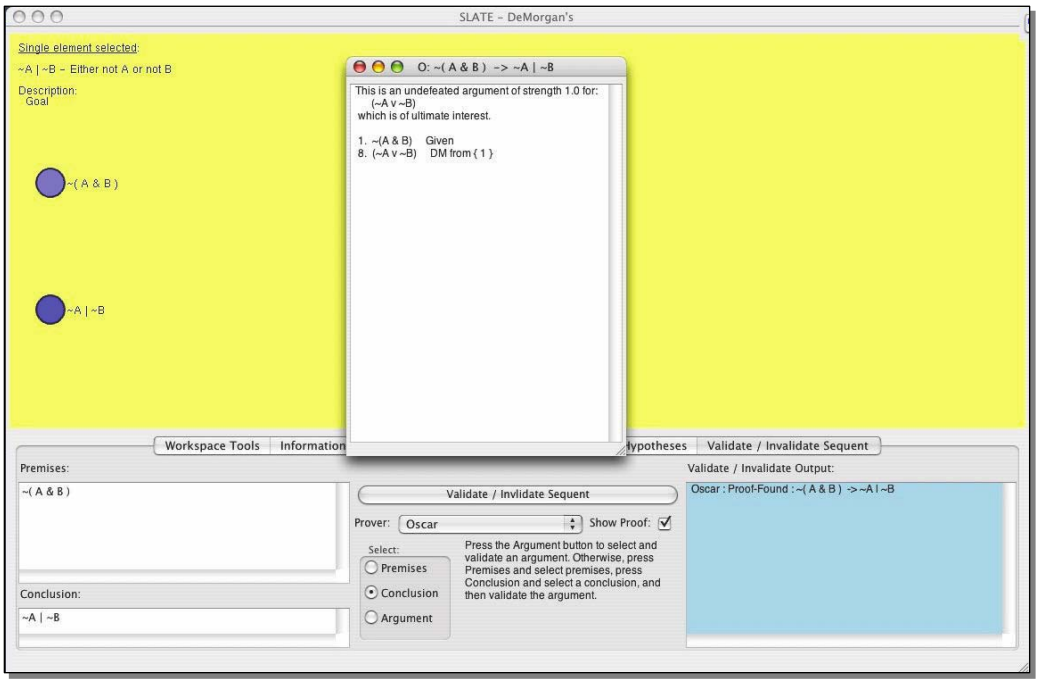


Figure 13. Oscar declares that this is a theorem, and returns a proof in its own format

Having studied the proof generated by Oscar, the student sets out to build her own proof in Slate's diagrammatic proof construction environment. Figure 14 shows a proof constructed by the student. Notice that the proof is indeed diagrammatic in nature.

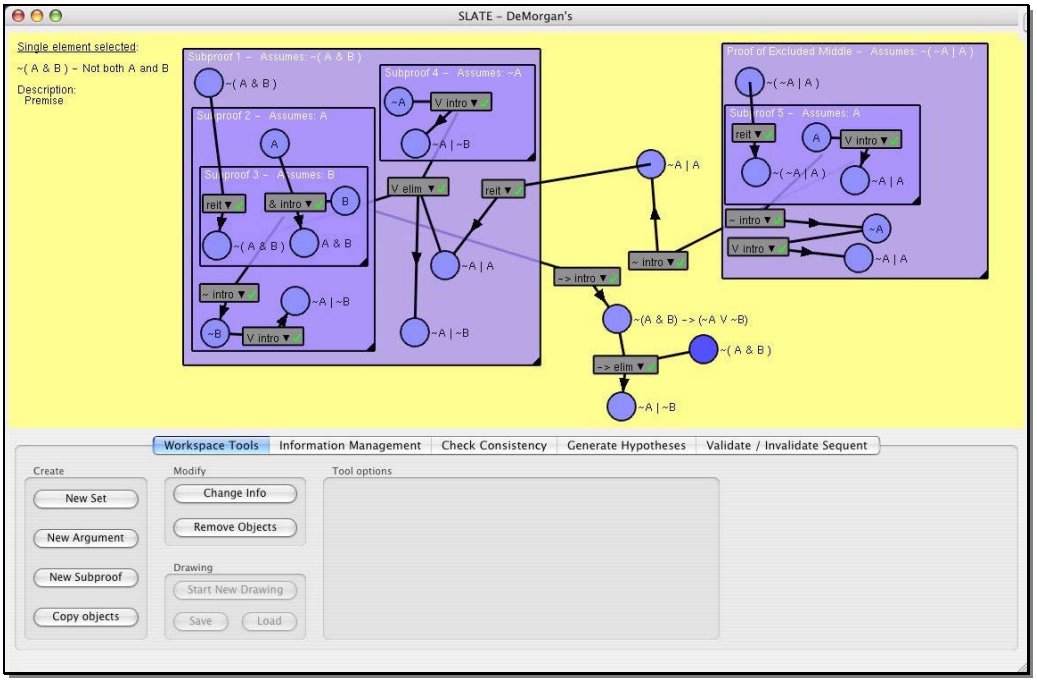


Figure 14. The student constructs a proof of her own in Slate's diagrammatic proof

CONSTRUCTION SYSTEM

Now comes the final challenge: the student must convert the diagrammatic proof into a symbolic proof, expressed in NDL. Figure 15 shows the situation after the NDL proof has been constructed. At this point, Slate's automatic grading utility can check the proof, diagnose any errors, return the diagnosis to the student, and issue a grade. In this example, the student is 100% correct. Finally, though we don't show it, the student can ask Slate to convert the NDL proof to English.

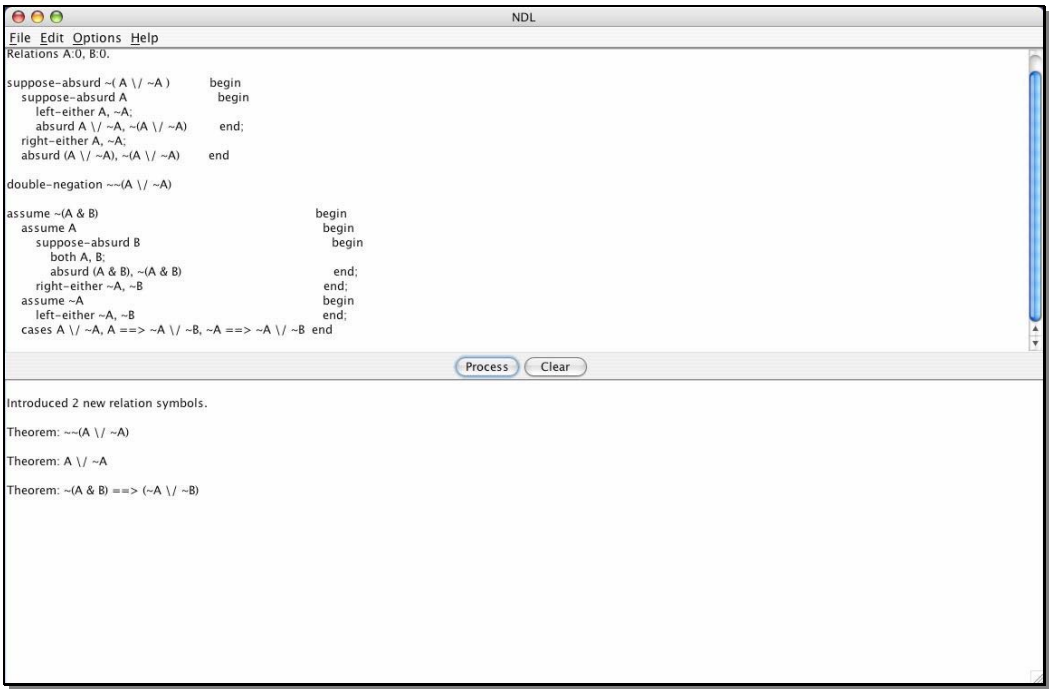


Figure 15. The student builds the proof in NDL.

Using slate to learn informal reasoning. What about informal reasoning? We have space enough only to say a few words about Slate's capacity in this area. First, for informal reasoning, an appropriate exercise would be selected. At present, all of these exercises are case studies designed to train intelligence analysts. Some of these case studies have been designed by others (see Hughes, 2003). Some of those designed by Bringsjord are so-called isomrophs of problems in (Pelletier, 1986).

THE FUTURE

In Fall 2005, Slate was deployed at Rensselaer in the course *Introduction to Logic*, with great success. In addition, plans are being laid to have Slate used by intelligence analysts who desire to be reasoners immune from well-known pitfalls (see Heuer, 1999). We are currently analyzing data returned by deployment at Rensselaer, and by other experiments outside the scope of the present paper (e.g., experiments designed to see if humans empowered by Slate *qua* assistant are immune from logical illusions). We stand ready to refine the system in an attempt to move ever closer to a time when Slate will indeed be able to superteach both formal and informal reasoning.

REFERENCES

Arkoudas, K. (2000). *Denotational proof languages*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Arkoudas, K., & Bringsjord, S. (2004). Metareasoning for multi-agent epistemic logics. *In Proceedings of the Fifth International Conference on Computational Logic in Multi-Agent Systems*, (pp. 50–65). Lisbon, Portugal.

Arkoudas, K., & Rinard, M. (2004). *Deductive runtime certification.* Paper presented at the 2004 Workshop on Runtime Verification, Barcelona, Spain.

Barwise, J., & Etchemendy, J. (1994). *Hyperproof*. Stanford, CA: CSLI.

Barwise, J., & Etchemendy, J. (1999). *Language, proof, and logic*. New York: Seven Bridges.

Bello, P., & Bringsjord, S. (2003). HILBERT and PATRIC: Hybrid intelligent agent technology for teaching context-independent reasoning. In V. Formichov (Ed.), *Educational Technology and Society*, *6*, 30–42.

Braine, M. (1998). Steps toward a mental predicate-logic. In M. Braine & D. O'Brien (Eds.), *Mental Logic*, 273–331. Mahwah, NJ: Lawrence Erlbaum.

Bringsjord, S., & Yang, Y. (2003a). Logical illusions and the welcome psychologism of logicist artificial intelligence. In D. Jacquette (Ed.), *Philosophy, psychology, and psychologism: Critical and historical essays on the psychological turn in philosophy*, 289–312. Netherlands: Kluwer, Dordrecht.

Bringsjord, S., & Yang, Y. (2003b). Problems used in psychology of reasoning are too easy, given what our economy demands. *Behavioral and Brain Sciences*, *26*(4), 528–530.

Bringsjord, S., & Yang, Y. (2003c). Representations using formal logics. In L. Nadel (Ed.), *Encyclopedia of cognitive science, 3*, 940–950. London, UK: Nature Publishing Group.

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, R. M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology*, *18*, 293–328.

Ebbinghaus, H. D., Flum, J., & Thomas, W. (1984). *Mathematical logic*. New York: Springer-Verlag.

Goble, L., (Ed.) (2001). *The Blackwell guide to philosophical logic*. Oxford, UK: Blackwell Publishers.

Heuer, R. (1999). *Psychology of intelligence analysis*. Pittsburgh, PA: United States Government Printing Office.

Hughes, F. J. (April, 2003). The art and science of the process of intelligence analysis: Case study #2 (The Case of Wigmore vs. Al-Qaeda). Technical report, Joint Military Intelligence College, Washington, DC.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press,

Johnson-Laird, P. N. (1995). Inference and mental models. In S. E. Newstead & J. S. T. Evans (Eds.), *Perspectives on thinking and reasoning*, (pp. 115–146). Hillsdale NJ: Lawrence Erlbaum.

Johnson-Laird, P. (1997a). Rules and illusions: A critical study of Rips's, *The Psychology of Proof, Minds and Machines*, *7*(3), 387–407.

Johnson-Laird, P. N. (1997b). And end to the controversy? A reply to Rips. *Minds and Machines*, *7*, 425–432.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., & Legrenzi, M. S. (2000). Illusions in reasoning about consistency. *Science*, *288,* 531–532.

Johnson-Laird, P., & Savary, F. (1995). How to make the impossible seem probable. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, pages 381–384. Hillsdale, NJ: Lawrence Erlbaum.

Newstead, S. E., & Evans, J. S. T., (Eds.). (1995). *Perspectives on thinking and reasoning*. Englewood Cliffs, NJ: Lawrence Erlbaum.

Pelletier, F. J. (1986). Seventy five problems for testing automatic theorem provers. *Bell System Technical Journal*, *2*, 191–216.

Pollok, J. (1995). *Cognitive carpentry: A blueprint for how to build a person*. Cambridge, MA: MIT Press.

Raven, J. C. (1962). *Advanced progressive matrices set II*. London, UK: H. K. Lewis.

Rinella, K., Bringsjord, S., & Yang, Y. (2001). Efficacious logic instruction: People are not irremediably poor deductive reasoners. In J. D. Moore & K. Stenning (Eds.). *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*, (pp. 851–856). Mahwah, NJ: Lawrence Erlbaum.

Rips, L. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.

Stillings, N., Weisler, S., Chase, C., Feinstein, M., Garfield, J., & Rissland, E. (1995). *Cognitive science*. Cambridge, MA: MIT Press.

Voronkov, A. (1995). The anatomy of vampire: Implementing bottom-up procedures with code trees. *Journal of Automated Reasoning*, *15*(2), 237-265.

Wason, P. (1995). Creativity in research. In *Perspectives on thinking and reasoning*, (pp. 287–301). Hillsdale, NY: Lawrence Erlbaum.

Wos, L. (1996). *The automation of reasoning: An experimenter's notebook with* OTTER *tutorial*. San Diego, CA: Academic Press.

Wos, L., Overbeek, R., Lusk, E., & Boyle, J. (1992). *Automated reasoning: Introduction and applications*. New York: McGraw Hill.

Yang, Y., Braine, M., & O'Brien, D. (1998). Some empirical justification of one predicate-logic model. In M. Braine & D. O'Brien (Eds.). *Mental logic*, (pp. 333–365). Mahwah, NJ: Lawrence Erlbaum.

Yang, Y., & Bringsjord, S. (2001a). Mental metalogic: A new paradigm for psychology of reasoning. In *Proceedings of the Third International Conference on Cognitive Science,*, (pp. 199–204). Hefei, China: University of Science and Technology of China Press.

Yang, Y., & Bringsjord, S. (2001b). The mental possible world's mechanism: A new method for analyzing logical reasoning problems on the GRE. In *Proceedings of the Third International Conference on Cognitive Science*, (pp. 205–210). Hefei, China: Press of the University of Science and Technology of China.

Yang, Y., & Bringsjord, S. (in-press). *Mental metalogic: A new, unifying theory of human and machine reasoning*. Mahwah, NJ: Erlbaum.

Yang, Y., & Johnson-Laird, P. N. (2000a). How to eliminate illusions in quantified reasoning. *Memory and Cognition*, *28*(6), 1050–1059.

Yang, Y., & Johnson-Laird, P. N. (2000b). Illusory inferences with quantified assertions. *Memory and Cognition*, *28*(3), 452–46.

ACKNOWLEDGEMENT: With much gratitude we acknowledge the financial support provided by Advanced Research and Development Activity (ARDA), contract MDA-904-03-C-0408 to Booz Allen Hamilton (general contractor) and RPI (sub-contractor), Novel Intelligence from Massive Data Program.