Slate: An Argument-Centered Intelligent Assistant to Human Reasoners

Selmer Bringsjord, Joshua Taylor, Andrew Shilliday, Micah Clark, Konstantine Arkoudas

Rensselaer AI and Reasoning Lab
Department of Cognitive Science
Department of Computer Science
Rensselaer Polytechnic Institute
Troy, NY USA

Based on slides for CMNA 08
8th Workshop on Computational Models of Natural Argument
In conjunction with ECAI 2008, Patras, Greece
21 July 2008
Slate was designed and developed by:
Selmer Bringsjord
Andrew Shilliday
Joshua Taylor

With valuable suggestions from:
Micah Clark, Wayne Gray,
Michael Schoelles, Marc Destefano,
and Jason Wodicka.

Slate is the property of Rensselaer Polytechnic Institute (RPI) and the Rensselaer Artificial Intelligence and Reasoning (RAIR) Lab. When officially released, sponsors and general contractors enjoy an unrestricted license to the system.

Copyright (c) 2003-2008 Rensselaer Polytechnic Institute. All rights reserved.
Someone who lives in Dreadsbury Mansion killed Aunt Agatha. Agatha, the butler, and Charles live in Dreadsbury Mansion, and are the only people who live therein. A killer always hates his victim, and is never richer than his victim. Charles hates no one that Aunt Agatha hates. Agatha hates everyone except the butler. The butler hates everyone not richer than Aunt Agatha. The butler hates everyone Agatha hates. No one hates everyone. Agatha is not the butler.

Who killed Agatha?
Charles hates no one that Aunt Agatha hates.
Agatha hates everyone except the butler.
The butler hates everyone Agatha hates.
A killer always hates his victim and is never richer than his victim.

Someone who lives in Dreadsbury Mansion killed Aunt Agatha.
The butler hates everyone not richer than Aunt Agatha.
Agatha is not the butler.
No one hates everyone.

Agatha, the butler, and Charles live in Dreadsbury Mansion, and are the only people who live therein.

Ready.
Selected information is inconsistent.

A killer always hates his victim, and is never richer than his victim. Agatha hates everyone except the butler. Charles hates no one that Aunt Agatha hates. Agatha, the butler, and Charles live in Dreadsbury Mansion, and are the only people who live therein. The Butler killed Agatha.

Hypothetical The Butler killed Agatha. – The Butler killed Agatha.
A killer always hates his victim, and is never richer than his victim.

The butler hates everyone Agatha hates.

Someone who lives in Dreadsbury Mansion killed Aunt Agatha.

No one hates everyone.

Agatha is not the butler.

The butler hates everyone not richer than Aunt Agatha.

Agatha hates everyone except the butler.

Agatha, the butler, and Charles live in Dreadsbury Mansion, and are the only people who live therein.

The Butler killed Agatha.

Charles killed Agatha.

Agatha killed herself.

Charles hates no one that Aunt Agatha hates.

Deduction ▼

Deduction ▼

Deduction ▼

Deduction ▼
Theoretical Foundations

- Robust, multi-faceted theory of heterogeneous human and machine reasoning
- Affirming deductive, inductive, abductive, analogical and visual reasoning
- Using arguments and counter-arguments; proofs and disproofs; models and countermodels; and strength factors
“System 2” Cognition

- “System 1” cognition bound to concrete contexts; prone to error
- “System 2” cognition “[abstracts] complex situations into canonical representations … stripped of context”
- Slate facilitates and amplifies users’ “System 2” cognitive abilities.
• Slate must support the processes that “System 2” cognition depends on, such as:
  
  • mental logic — manipulation of linguistic and symbolic entities
  
  • mental models — imagining and exploring possible situations
  
  • meta-reasoning, and mental meta-logic
• Slate is based on a cognitively plausible epistemic theory.

• In Slate, “justified beliefs” are grounded in justifications, not Kripke semantics.
Visual Interface

• Slate, though logically grounded, is inherently visual.

• To assist in argumentation, Slate adopts a representation that subsumes and exceeds earlier argument-mapping technology.
Some of the Frenchmen in the room are gourmets.

All the Frenchmen in the room are wine-drinkers.

Some of the wine-drinkers in the room are gourmets.
• Slate adds formal, mechanically processable, semantics to the argument-mapping workspace.

• Propositions, hypotheses, and arguments are depicted graphically.
• Models are also depicted graphically.

• In the future, more specialized visualizations will be available through the use of “visual-logics,” such as Vivid.
Agatha hates The Butler, who is richer than Charles, who hates Agatha.

Charles hates Agatha, and The Butler has a cycle of hating himself.
Linguistic Interface

• There are benefits to working with formal logic directly, but it’s usually easier to work with natural languages.

• Furthermore, many of Slate’s intended users don’t have a mastery of formal logic.
Slate can process information expressed in _logically-controlled English_—a subset of English that can be unambiguously translated into a formal logic.
Process: Intelligence Reports → Multi-Sorted Logic
Phase One: Intelligence Reports $\rightarrow$ Controlled English

Phase One:
- Intelligence Reports
  $\rightarrow$ Transcription (informed by Lexicon Database)
  $\rightarrow$ Controlled English

Phase Two:
- Controlled English
  $\rightarrow$ Translation (informed by Lexicon Database)
  $\rightarrow$ Discourse Representation Structures

Phase Three:
- Discourse Representation Structures
  $\rightarrow$ Translation (informed by Semantic Ontology)
  $\rightarrow$ Multi-Sorted Logic
Phase Two: Controlled English $\rightarrow$ DRS
Phase Three: DRS $\rightarrow$ Multi-Sorted Logic
Reading Process Implementation

Process: Intelligence Reports → Multi-Sorted Logic
Reasoning Technology

• Argument checking
• Proof search
• Model finding
Strength Factors & Uncertainty

- Humans use probabilistic language, but not in accordance with probability theory.
- Slate’s strength factors are based on Roderick Chisholm’s.
- Users apply strength factors to certain parts of strength factors, and Slate propagates strength factors to the rest.
Slate employs a set of 9 discrete strength factors:

4 Certain
3 Evident
2 Beyond Reasonable Doubt
1 Probable
0 Counterbalanced
-1 Probably False
-2 Reasonable to Disbelieve
-3 Evidently False
-4 Certainly False
Deduction

(2) Beyond Reasonable Doubt

Ind (3)

(1) Probable
(1) Probable

Ind (2)

(3) Evident
(4) Certain
(0) Counterbalanced

Ready.
Interoperability

• Standardized representational schemes (e.g., IKL, Common Logic) are used for syntactic interoperability between Slate and other KB systems.

• We use PBSI$^+$ for semantic interoperability.

• PBSI$^+$ is a formal technique for developing bridging axioms for ontology translation that provably preserve semantic meaning.
Interoperability
Gödelian Example

• Gödel’s first incompleteness result is one of the most celebrated results in mathematical logic, stating that

• any consistent, formal, recursively enumerable theory that proves basic arithmetical truths cannot prove all arithmetical truths.
• The proof's results can be difficult to accept, perhaps from the novelty and ingenuity of Gödel's proof, and

• the counter-intuitiveness of his result.

• A Gödelian logic puzzle can quicken acceptance of the result, and makes for a good demonstration of Slate.
1. The constants $\sim$, $\cdot$, $P$, and $M$ are each distinct.
2. The constants $\sim$, $\cdot$, $P$, and $M$ are the only glyphs.
3. The concatenation of two terms is an expression if and only if both terms are themselves expressions.
4. Concatenation is associative.
5. The term $\varphi$ is an expression if and only if $\varphi$ is a glyph or is the concatenation of two expressions.
6. The mirror of an expression $\varphi$ is defined as the concatenation of $\varphi$, $\cdot$, and $\varphi$ (i.e., $\varphi \cdot \varphi$).
7. If $\varphi$ is an expression, then $P \cdot \varphi$, $P \cdot M \cdot \varphi$, $\sim P \cdot \varphi$, and $\sim P \cdot M \cdot \varphi$ are sentences.
8. If $\varphi$ is an expression then the sentence $P \cdot \varphi$ is true if and only if $\varphi$ is provable.
9. If $\varphi$ is an expression, then the sentence $P \cdot M \cdot \varphi$ is true if and only if $\varphi$ is not provable.
10. If $\varphi$ is an expression, then the sentence $\sim P \cdot M \cdot \varphi$ is true if and only if $\varphi$ is not provable.
11. If $\varphi$ is an expression, then the sentence $\sim P \cdot M \cdot \varphi$ is true if and only if the mirror of $\varphi$ is not provable.
12. Every sentence $\varphi$ that is provable is also true.
13. \( \sim PM \) is an expression.
14. \( \sim PM \ast \sim PM \) is a sentence.
15. \( \sim PM \ast \sim PM \) is true.
16. \( \sim PM \ast \sim PM \) is not provable.
~ P M is an expression

Ready.
~ P \land M \text{ is a sentence}

~ P \land \sim P \land M \text{ is true}
\[ \sim \ P \ M \text{ is an expression} \]

\[ \sim \ P \ M \text{ is a sentence} \]

\[ \sim \ P \ M \text{ is true} \]

\[ \sim \ P \ M \text{ is not provable} \]

Proposition P-12 – P-12
Slate supports informal reasoning.

Recreating an example from Doyle’s *The Adventure of the Blue Carbuncle* shows strength-factor-based and non-deductive arguments in Slate.
Description of hat.

- Flat brim, curled at the edges
- Lined with red silk.
- Brim pierced for a hat-securer.
- Elastic for hat-securer missing.

Ready.
Deduction

Lined with red silk.

Abd (2)

Owner was well off when hat was purchased.
(2) Beyond Reasonable Doubt

The hat is three years old.
(1) Probable

Owner is no longer well off.
(1) Probable

Description of hat.

Flat brims with curled edges were in style three years ago.

Deduction

Flat brim, curled at the edges

Deduction

Elastic for hat-securer missing.

Deduction

Brim pierced for a hat-securer.
Efficacy and Evaluation

- Slate’s theory is good.
- Highly trained individuals can use Slate effectively.
- Can the typical undergraduate student?
**Methodology**

- **Subjects**: 40 undergraduate RPI students
  - 20 performed the task using only pencil and paper.
  - 20 performed the task with the option to use Slate.
Design

• Univariate between-subject design

• Subjects complete several reasoning problems

• Subjects solved problems using pencil and paper (the control condition), or were allowed to use Slate to help.

• Salient dependent variable, and metric for good performance, was subject’s solution accuracy.
Procedure

- Each problem was of the following form:
  - Premises $a_1, a_2, \ldots, a_n$ enumerated
  - Prompt $p$, “What logically follows?”
  - Subjects instructed to provide and justify an answer
  - Subjects allowed as much time as necessary, but no outside help
• P&P subjects were talked through an example inference problem and then allowed to begin work on real problems.

• Slate subjects were shown a tutorial video that included solving a sample problem. Once comfortable, subjects were allowed to begin.
Results

• Exactly one of the following statements is true:

  • If there is a king in the hand, then there is an ace in the hand.

  • If there isn’t a king in the hand, then there is an ace in the hand.
Assume the following proposition is true:

1. ‘If there is a king in the hand, then there is an ace in the hand’, or, ‘If there is not a king in the hand, then there is an ace in the hand’, but not both.

What can you infer from the above premise?

Answer & Justification:

That there is always an ace in the hand. This is true because either way with a king in the hand or without there is always an ace. Having the ace does not infer anything about the king however.

Incorrect (P&P)
Assume the following proposition is true:

1. ‘If there is a king in the hand, then there is an ace in the hand’, or, ‘If there is not a king in the hand, then there is an ace in the hand’, but not both.

What can you infer from the above premise?
Answer & Justification:

Answer — There must not be an ace in the hand.

Justification — See Report.

We are told that exactly one of the following statements is true:
1) If there is a king in the hand then there is an ace in the hand.
2) If there is not a king in the hand then there is an ace in the hand, but not both.

If exactly one of the above statements is true, then it follows that exactly one of the above statements must be false. More specifically, it is either the case that statement 1 is false, or that statement 2 is false.

Imagine that statement 1 is the statement which is false. From this assumption, it follows that there is a king in the hand and that there is not an ace in the hand. So, if we assume that statement 1 is false, we can conclude that there is not an ace in the hand.

Now, imagine that statement 2 is the one which is false. Along a similar line of reasoning, it follows that there is not a king in the hand, and that once again there is not an ace in the hand.

No matter which of the above statements is false, we can still infer that there is not an ace in the hand, and because we know that one of the statements must be false, we can conclude that there must not be an ace in the hand.
Refactorization

- A new Java UI and implementation that will
- Allow reasoning components to run on remote, high-performance servers.
- Support multi-user shared workspaces.
We are currently exploring ways to exploit Semantic Web technology including:

- Information exchange via RDF
- Ontology/signature sharing using RDFS and OWL
Reasoning Technology

- Continued work on integrating reasoning technology into Slate.
- For instance, currently Slate only employs *finite* model-finders, but researchers in the RAIR Lab are working to construct *infinite* model finders.
Thank You
Ευχαριστώ