Artificial Intelligence II

CSC 720—Spring 2010

Jon Doyle

Department of Computer Science
North Carolina State University

January 17, 2011

Logical rationality

Logic and rationality

Logic as instrumental

- An analytical tool
- A representational tool
- A computational tool
- *Logic proper*

Logic as a norm for thinking

- Prescriptions for representation
- Prescriptions for reasoning
- *Logicism*

Logic provides an analytical tool

- Logical analysis can apply to all types of representations
- Precise concepts for expressing meanings
  - How to describe things with words
  - Language and syntax
  - Meaning and semantics
- Precise concepts for critiquing meanings
  - How to tell if your words meant what you intended
  - Inference and proof
  - Analytical concepts
- Formal inference systems which "preserve truth"
  - *Not a guarantee of correct answers!*
  - *Not a guarantee of useful answers!*
  - *Not a guarantee of intelligible answers!*

Formal representations and computational mechanisms

- Expressing logical descriptions and inferential connections
  - Predicate, modal, and model-theoretic logics
  - Description logics
  - Microtheories and metatheories
- Computing logical inference and learning
  - Resolution-based logic programming systems
  - Answer-set programming systems
- Computing theory revision and identification
  - Reason maintenance systems
  - Inductive logic programming and PAC learning

There are many different logics

Propositional and first-order logics

- Subject-independent building blocks

Other logical languages “build in” different concepts

- Higher-order logics (set-theoretic, categorical, . . .)
- Modal logics (necessitative, dynamic, . . .)
- Model-theoretic logics (probability, topology, . . .)
- Philosophical logics (deontic, epistemic, . . .)
- Commonsense logics (nonmontonic, fuzzy, . . .)
- Computational logics (temporal, functional, . . .)

Logicism

Normative logical rationality

- Beliefs must be consistent
- Inferences must be sound
- Inferences must be complete
  - *If it's sound, do it!*
- Knowledge of logic must be complete
  - No insistence that general knowledge is complete
- Learning and action must preserve consistency

Consistency

Emerson's thesis

- "A foolish consistency is the hobgoblin of little minds"

Minsky's corollary

- Only small minds can be consistent for very long

Niven's corollary

- "Some mistakes we must carry with us"

Soundness

Soundness is not sufficient

- $A \vdash A \wedge (A \vee A)$

Minsky: adding control axioms does not help

$$\frac{\begin{array}{c} A \\ A \rightarrow B \\ \text{“}\mathtt{Avoid\ concluding}\ B\text{”} \end{array}}{B \wedge \text{“}\mathtt{Avoid\ concluding}\ B\text{”}}$$

Soundness is not necessary

- `At(door)` $\vdash$ `[Walk-to(desk)]At(desk)`

Logical omniscience

Full knowledge of logical consequences

- If P is known and $P \models Q$, then Q is known too
- If P and Q are inconsistent, the inconsistency is known
- To ensure consistency, one must have full knowledge of consequences

Necessary for logical rationality

- To ensure consistency, one must have full knowledge of consequences

Modal logics of knowledge and belief

Suppose that $\Box\phi$ means "I know that ϕ"

- Then $\Box\phi \rightarrow \phi$ seems right

Suppose that $\Box\phi$ means "I believe that ϕ"

- Then $\Box\phi \rightarrow \phi$ seems wrong

Kripke semantics

A form of possible world semantics for necessity

- $\Box\phi$ is true if ϕ is true in all possible worlds
- $\Box\phi$ is true if $w \models \phi$ for all $w \in W$

Insight: what is possible varies across logics

- Capture this variation in an *accessibility relation* S between possible worlds in W
- Accessibility = what is possible with respect to a possible world
- $\Box\phi$ is true in world w if $w' \models \phi$ for all $w' \in S(w)$
- Disputed axioms correspond to different accessibility relations

Example

Modal logic axiom schemata

- *K*: $\Box(\alpha \rightarrow \beta) \rightarrow (\Box\alpha \rightarrow \Box\beta)$
  - Deductive closure of knowledge
  - All models
- *T*: $\Box\phi \rightarrow \phi$
  - Knowledge implies truth
  - Reflexive models
- *PI* (or 4): $\Box\phi \rightarrow \Box\Box\phi$
  - Positive introspection: "I know I know what I know"
  - Transitive models
- *NI* (or *E* or 5): $\neg\Box\phi \rightarrow \Box\neg\Box\phi$
  - Negative introspection: "I know I don't know what I don't know"
  - Euclidean models

Axiomatic modal logic

- Base logic K = axiomatic predicate logic
  - + Axiom scheme K : $\Box(\alpha \rightarrow \beta) \rightarrow (\Box\alpha \rightarrow \Box\beta)$
  - + Necessitation rule: From α, infer $\Box\alpha$

Epistemic logics

- Logic T = K + scheme T
- Logic $S4$ = T + scheme *PI*
- Logic $S5$ = $S4$ + scheme *NI*

Doxastic logics

- Logic $K4$ = $S4$ - T
- Logic $K5$ = $S5$ - T

Logical views of thinking

Mental states modeled as logical theories

- Consistent set of logical sentences
- *Theory* = consequentially closed set of logical sentences
  - Usually modeled as deductively closed

Reasoning modeled as theory evolution

- Distinguish change in view from logical inference
  - Add, remove, or change statements
  - Same or changing language

Monotonicity and nonmonotonicity

Temporal mappings from instants to mental states

- Monotonic: $t \leq t' \rightarrow \mathrm{Th}(t) \subseteq \mathrm{Th}(t')$
- Nonmonotonic: $t \leq t' \not\rightarrow \mathrm{Th}(t) \subseteq \mathrm{Th}(t')$
- Ordinary reasoning is temporally nonmonotonic

Logical mappings from axiom sets to theories

- Monotonic: $A \subseteq A' \rightarrow \mathrm{Th}(A) \subseteq \mathrm{Th}(A')$
- Nonmonotonic: $A \subseteq A' \not\rightarrow \mathrm{Th}(A) \subseteq \mathrm{Th}(A')$
- Ordinary logic is logically monotonic

A practical problem

Your predicament:

- You think the English test is on Tuesday
- You think the math test is on Wednesday
- You hear your classmate Alice wish that the tests were not on the same day

What do you do?

Ideal belief or theory revision

Theory $A = Cn(A)$ as the object of change

- *Addition* $A + x = Cn(A \cup \{x\})$
- *Contraction* $A \dot{-} x$ removes x from A if possible
- *Revision* $A \dot{+} x$ consistently adds x to A
- Levi identity: $A \dot{+} x \stackrel{\text{def}}{=} (A \dot{-} \neg x) + x$

Quine's *minimum mutilation* principle

- What is minimized?
- Rescher's *preferred maximal consistent subsets*
- What sort of preferences?

Logical learning

Candidates for revision

Some candidates for consideration

- $A \Downarrow x \stackrel{\text{def}}{=} \{B \subseteq A \mid B \nvdash x\}$
- $A \Downarrow^* x \stackrel{\text{def}}{=} \{Cn(B) \mid B \in A \Downarrow x\}$
- $A \downarrow x \stackrel{\text{def}}{=} \max_{\subset}(A \Downarrow x)$

Some candidates for contractions

- Maxichoice: choose one from $A \downarrow x$
- Partial meet: intersect a subset of $A \downarrow x$
- Full meet: $\bigcap A \downarrow x$

AGM contraction axioms

Alchourrón, Gärdenfors, and Makinson axioms

($\dot{-}$1) $A \dot{-} x$ is a theory whenever A is

($\dot{-}$2) $A \dot{-} x \subseteq A$

($\dot{-}$3) If $x \notin Cn(A)$, then $A \dot{-} x = A$

($\dot{-}$4) If $\nvdash x$, then $x \notin Cn(A \dot{-} x)$

($\dot{-}$5) If $\vdash x \leftrightarrow y$, then $A \dot{-} x = A \dot{-} y$

($\dot{-}$6) $A \subseteq Cn((A \dot{-} x) + x)$ whenever A is a theory

($\dot{-}$7) $(A \dot{-} x) \cap (A \dot{-} y) \subseteq A \dot{-} (x \wedge y)$ whenever A is a theory

($\dot{-}$8) If $x \notin A \dot{-} (x \wedge y)$, then $A \dot{-} (x \wedge y) \subseteq A \dot{-} x$ whenever A is a theory

Another perspective

Epistemic entrenchment

- Ordering over propositions in theories
- $x < y$ means give up x before y
    - $x \leq y$ if $x \vdash y$
    - $x \leq y$ iff $x \notin A \dot{-} (x \wedge y)$ or $\vdash x \wedge y$

Conditions on contractions

Gärdenfors and Makinson axioms

($\leq$1) If $x \leq y$ and $y \leq z$, then $x \leq z$
($\leq$2) If $x \vdash y$, then $x \leq y$
($\leq$3) Either $x \leq x \wedge y$ or $y \leq x \wedge y$
($\leq$4) If A is consistent, then $x \leq y$ for all y iff $x \notin A$
($\leq$5) If $x \leq y$ for all x, then $\vdash y$

Relating the perspectives

- Theorem: the axiom sets are equivalent

What determines entrenchment?

AGM approach based on *coherence* among beliefs

- Coherence conditions on theories (deductive closure and consistency) and on contractions
- No explicit separation between types of beliefs, only differing levels of entrenchment
- Entrenchment apparently exogenous to standard attitudes (belief, desire, intention)

Why one order and not another?

- Entrenchment from agent preferences
- Entrenchment from structure of memory

Identifying logical states of mind

How does one tell what an agent believes?

- Ask questions about what is believed?
- Infer from observed behavior?
- Ask for explanations of behavior?
- Observe in brain?

Later: economists give a different answer

Logical and other semantics

Logical semantics is based on denotation

Other conceptions

- Embodiment and causal connection
- Operational
- Pragmatic
- Red Queen semantics

Reflection versus requirement

- Necessitation rule: ϕ, therefore $\texttt{Know}(\phi)$
- Knowledge schema: $\texttt{Know}(\phi) \rightarrow \phi$
- Adding axiom $\texttt{Know}(\phi)$ produces ϕ

- Axiom $\texttt{Know}(\phi)$ acts as a specification or requirement of different character than merely adding ϕ
  - “You are getting sleepy, very sleeeeeepy”
- Force of requirement depends on constitution that enforces the knowledge schema