Approximate anytime inference:
Half an answer on time is better than a perfect answer too late

Frank van Harmelen
Vrije Universiteit Amsterdam

vrije Universiteit

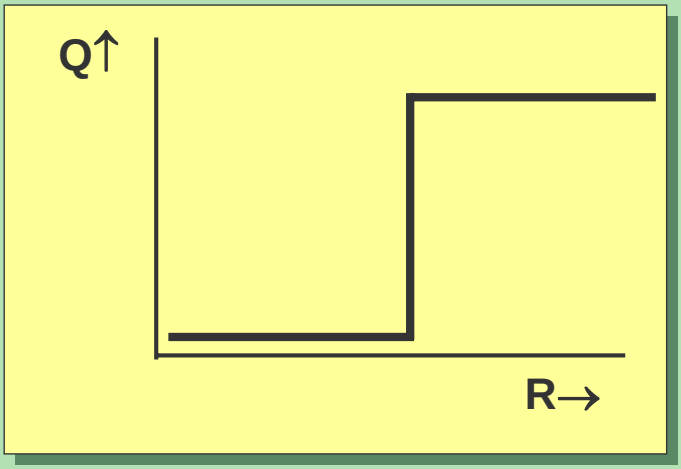
Who am I (advance warning)

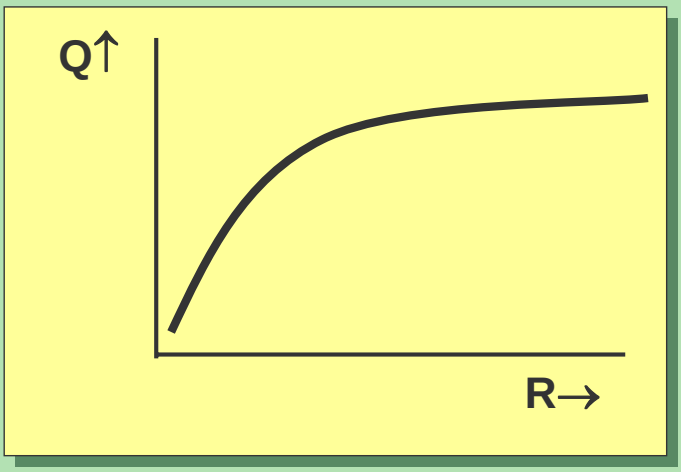
- I'm **not** a Theoretical Computer Scientist

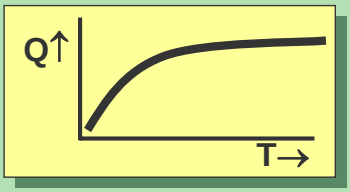
- I do Knowledge Representation & Reasoning
- often dealing with intractable problems
- **user** of theory

Robust Knowledge Representation

- Reliance on logic is a **strength**
  - Strong theoretical basis
  - Well known properties
  - Well known implementation techniques
- Reliance on logic is a **weakness**
  - **Crisp** (no approximate answers)
  - **Abrupt** (no intermediate answers)
  - **Inefficient** (no time/quality trade-off)





Why do we want 

?

time pressure, hard deadlines

perfect answer not needed

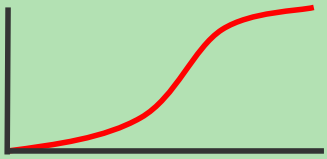
Another way of stating the problem:

Logic = a model of a perfect reasoner in idealised circumstances

- no wrong steps
- no missing knowledge
- no incorrect knowledge
- **unlimited time**

Approach: Anytime Inference

- **Quality is function of some varying resource**
  - **reasoning time,**
  - **inference accuracy,**
  - **representational precision**
- **This function is**
  - **monotonic**
  - **diminishing returns (usually)**
  - **optimal at some time T**
  - **this T not $\gg$ best abrupt algorithm**
  - **characterised by a performance profile**



How to use this?

- many problems can be stated in terms of $\vdash$:
  - planning
  - learning
  - classifying
  - **diagnosing**
  - ...
- General idea: **replace $\vdash$ with $|\!\sim$**

Two example studies

- For "reasoning" take **diagnosis**
  - replace $\vdash$ by $|\sim$
  - $|\sim$ = 1,3-**S** (Cadoli & Schaerf)
  - $|\sim$ = BCP-**k** (Dalal)

- Characterise the effects of **changing S and k**

$|\sim^1$, $|\sim^3$ (Cadoli & Schaerf)

Take subset S of propositional letters
Non-classical behaviour for letters x **outside S**:

- 1-S assignment:
  - *x* and $\neg x$ are **both false**
  - only 1 assignment possible for $x \notin S$
  - intuition on clausal form: remove part of clause

$$x \vee b, \neg b \vee c \quad \Rightarrow \quad b, \neg b \vee c$$

  - Note: theory might become $\perp$

(NB: For letters <u>in *S*</u>: classical behaviour)

$|\!\sim^1$, $|\!\sim^3$ (Cadoli & Schaerf)

Take subset S of propositional letters
Non-classical behaviour for letters x **outside S**:

- 3-S assignment:
  - x and $\neg x$ are **not both false**
  - 3 assignments possible for x $\notin$ S
  - intuition on clausal form: remove entire clause

$$a \vee b, \neg b \vee c \quad \Rightarrow \quad \neg b \vee c$$

  - Note: theory might become $\top$

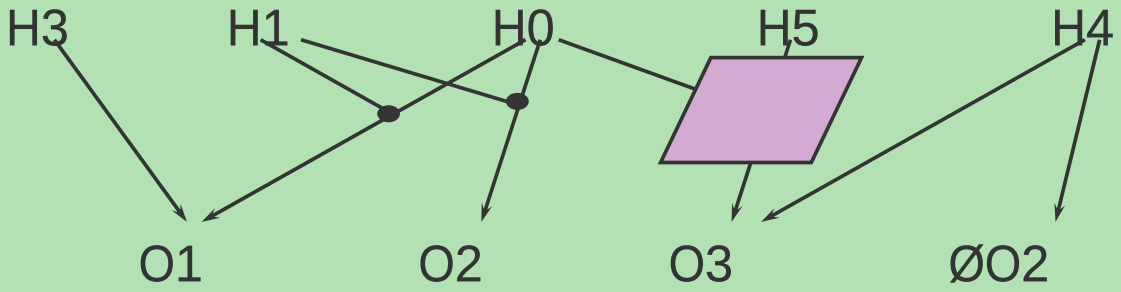
$|\!\sim^1$, $|\!\sim^3$ (Cadoli & Schaerf) (ctd)

- Properties:
- $|\!\sim^1$ is unsound but complete
- $|\!\sim^3$ is incomplete but sound
- S determines "degree of approximation":

$$\vdash_3^{\emptyset} \Rightarrow \vdash_3^{S} \Rightarrow \vdash_3^{S'} \Rightarrow \vdash_2 \Rightarrow \vdash_1^{S'} \Rightarrow \vdash_1^{S} \Rightarrow \vdash_1^{\emptyset}$$
$$\nvdash_2 \Leftarrow \nvdash_1^{S'} \Leftarrow \nvdash_1^{S} \Leftarrow \nvdash_1^{\emptyset}$$

- Efficient anytime algorithms
- total iterative cost $\leq$ cost of $\vdash_2$

Example of $|\sim^3$

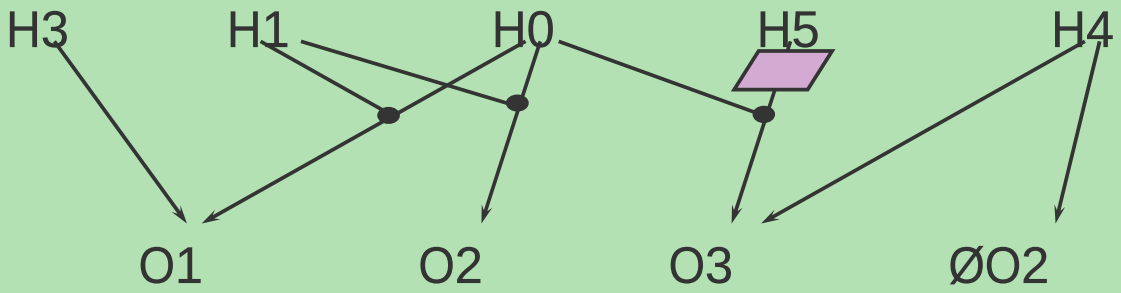


Theory T

$T \cup \{H0, H5\} \vdash O3$
if $H5 \notin S$ then
$T \cup \{H0, H5\} \not\vdash_3 O3$

incomplete

Example of $|\sim^1$

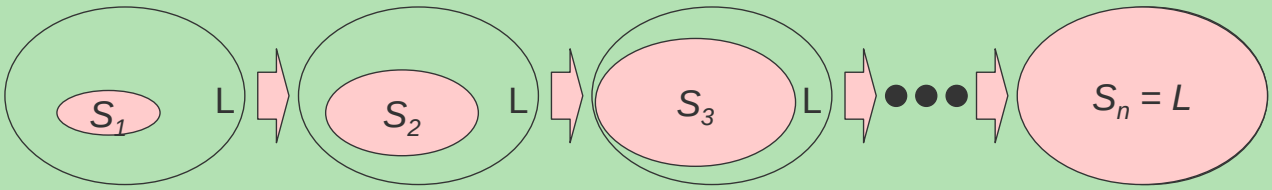


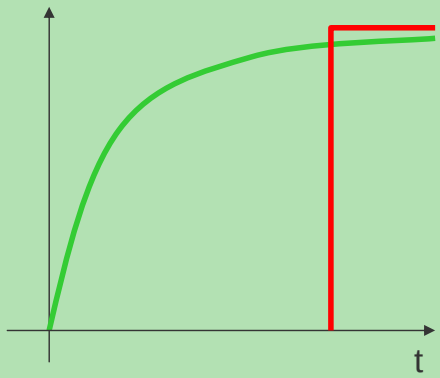
Theory T

$T \cup \{H0\} \not\vdash O3$
if $H5 \notin S$ then
$T \cup \{H0\} \vdash_1 O3$

unsound

Summary





- Anytime behaviour when S_i is increased
- Previous steps can be reused

Standard def. of diagnosis

Given:

Behaviour model: T

Observations to be explained: O^+

Observations to be consistent: O^-

Find: Explanation: E such that:

$$\left.\begin{array}{rcl} T \cup E & \vdash & O^+ \\ T \cup E \cup O^+ & \nvdash & \bot \end{array}\right\} \text{ABD}$$

$$T \cup E \cup O^- \nvdash \bot \qquad \text{CBD}$$

Theorems using $|\!\sim^{1,3}$:

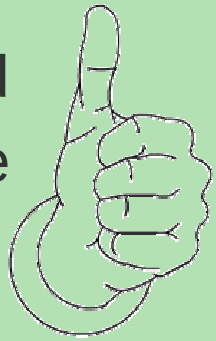
- ABD_3 **shrinks** when S increases
- CBD_1 **grows** when S increases
- Any CBD_1 **is contained in** a classical diagnosis
- Any ABD_3 **contains** classical diagnoses

$$\begin{array}{lllll} \emptyset = & \{ABD_1^{\emptyset}\} & \subseteq & \{ABD_1^{S}\} & \not\subseteq & \{ABD^2\} \\ & \{ABD^2\} & \not\subseteq & \{ABD_3^{S}\} & \not\subseteq & \{ABD_3^{\emptyset}\} = \emptyset \end{array}$$

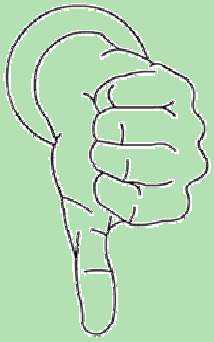
$$\begin{array}{lllll} 0 = & |\{ABD_1^{\emptyset}\}| & \leq & |\{ABD_1^{S}\}| & \leq & |\{ABD_2\}| \\ & |\{ABD_2\}| & \geq & |\{ABD_3^{S}\}| & \geq & |\{ABD_3^{\emptyset}\}| = 0 \end{array}$$

Approximate Entailment

- Semantically well-founded
- Computationally attractive
- Dual



- Parameter S is crucial for appropriate behaviour
- Almost no quantitative analysis



How to choose S?

- No good general strategy exists
- Problem-specific heuristics:
  - **S={urgent causes}**
    first diagnosis is most urgent candidate
  - **S={unreliable components}**
    first diagnosis is most likely candidate
  - **S={specific observations}**
    first diagnosis is most specific candidate

Second case study: BCP_k

BCP_0:

- **use all literals + a clause from T**
- **to derive a new literal α**
- **add α to T**

Example:

$$(P \vee Q), (P \vee \neg Q), (\neg P \vee Q), (\neg P \vee \neg Q) \quad \vdash \quad \bot$$

$$(P \vee Q), (P \vee \neg Q), (\neg P \vee Q), (\neg P \vee \neg Q) \quad \nvdash_0 \quad \bot$$

Basic notions: BCP_0 & BCP_k (ctd)

- Example:
  Th = (P v Q), (P v ¬Q), (¬P v S v T), (¬P v S v ¬T)

❶ Th $\vdash_0$ P

❷ Th, P $\vdash_0$ S **but**

❸ **not** Th $\vdash_0$ S

➔ $\vdash_0$ **cannot chain**

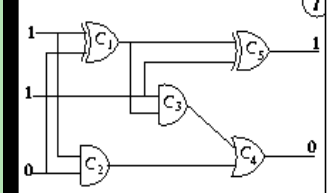
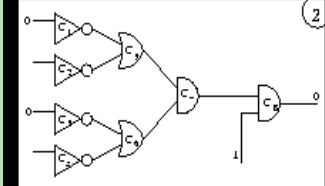
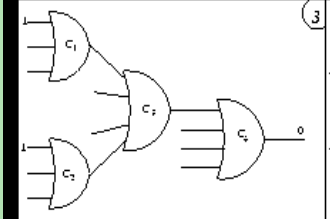
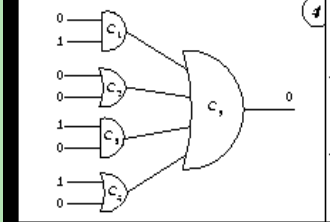
Basic notions BCP_0 & BCP_k (ctd)

- **BCP_k** = allow chaining on length $\leq k$

$$\frac{T \vdash_0 \phi}{T \vdash_k \phi} \qquad \frac{T \vdash_k \Psi \;;\; T, \Psi \vdash_k \phi}{T \vdash_k \phi}$$

- **if $|\Psi| \leq k$**
- When *k* **grows**, BCP_k is **more complete**
- For some k, BCP_k is complete

Examples of BCP_k diagnosis

System	value of k	Correct minimal k-diagnoses
1  (From [Reiter, 1987], Figure 1)	$k>=0$	{c1}, {c3,c5}, {c4,c5}
2  (From [Davis, 1988], Figure 20)	$k=0$	{ }
	$k>=1$	$\{c_1\}, \{c_3\}, \{c_5\}, \{c_6\}, \{c_7\}, \{c_8\}$
3 	$k=0, 1$	{ }
	$k=2$	$\{c_1\}, \{c_2\}, \{c_3\}, \{c_4\}$
	$k>=3$	$\{c_1,c_2\}, \{c_3\}, \{c_4\}$
4 	$k=0$	$\{c_4\}, \{c_5\}, \{c_1,c_2\}, \{c_1,c_3\}, \{c_2,c_3\}$
	$k=1$	$\{c_4\}, \{c_5\}, \{c_1,c_2,c_3\}$
	$k>=2$	$\{c_4\}, \{c_5\}$

Simple theorems

Small *k* : fewer diagnostic problems

- if ...$\vdash_k \perp$ then ... $\vdash_{k+1} \perp$

Small *k* : more diagnostic solutions

- {E | E is a BCP_{k+1} solution} $\subseteq$ {E | E is a BCP_k solution}

Thm: Upperbound on *k* (surprise)

- For a given component *c* :
  - *n* = # of unknown inputs & outputs of *c*
- For inferences on *c* : **BCP_{n-1} = classical**

- Corrollary: take *k'* = max *n-1* over all *c*
  - Then: **{$BCP_{k'}$, E} = {classical E}**
  - NB: independent of circuit size,
    independent of component complexity
    only dependent on component size!

Thm: Sharper upperbound on *k*

- Def: (*mcs* = minimal conflict set)
  - *k(mcs)* = minimal *k* needed to prove
    $\ldots \cup mcs \vdash_k \perp$
  - *k** = max *k(mcs)* over all *mcs*
- Then: **{BCP_{k*} E} = {classical E}**

Of course:

- $k^* \leq k'$
- k* not practical

Low *k* can do a lot (surprise)

Thm: for problem diagnosis

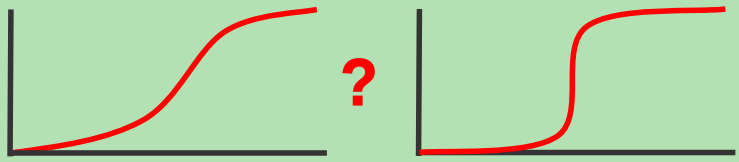
k=1 suffices for all binary circuits

Thm: for problem recognition

k=0 suffices for all known-input circuits

Open questions

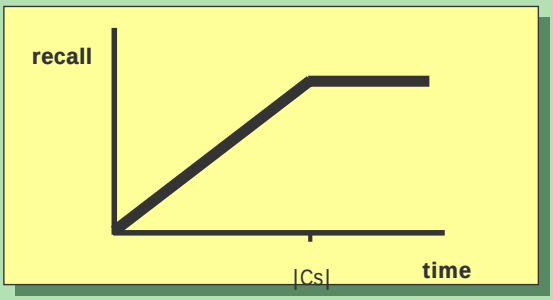
- How tight is k' (k' ≥ k*) ?
- Is k a good complexity indicator ?

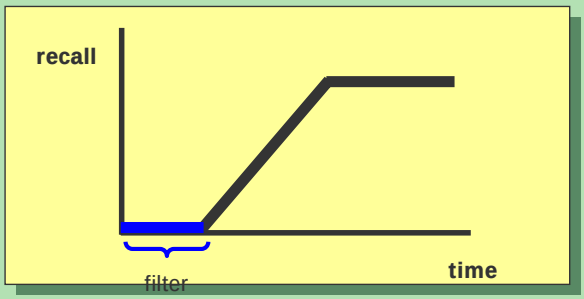


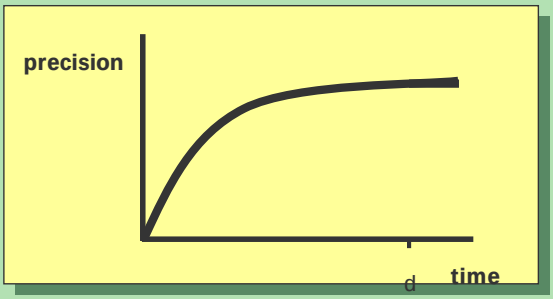
- Both are empirical questions ?

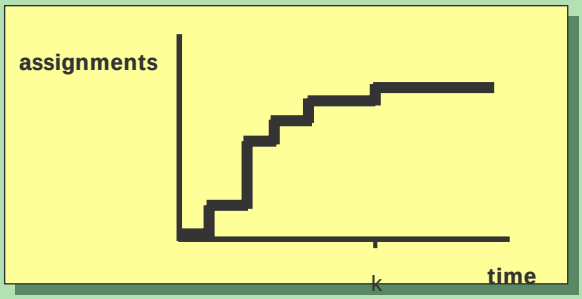
Analysis of performance profiles

- Qualitative:



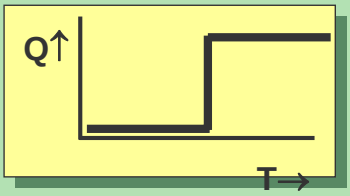




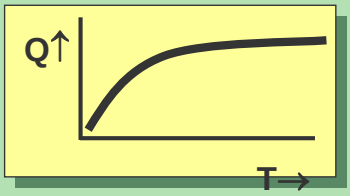


Take home message:

not



but



- **Approximate** forms of deduction
- **Anytime** algorithm for deduction
- realistic **use** of these algorithms