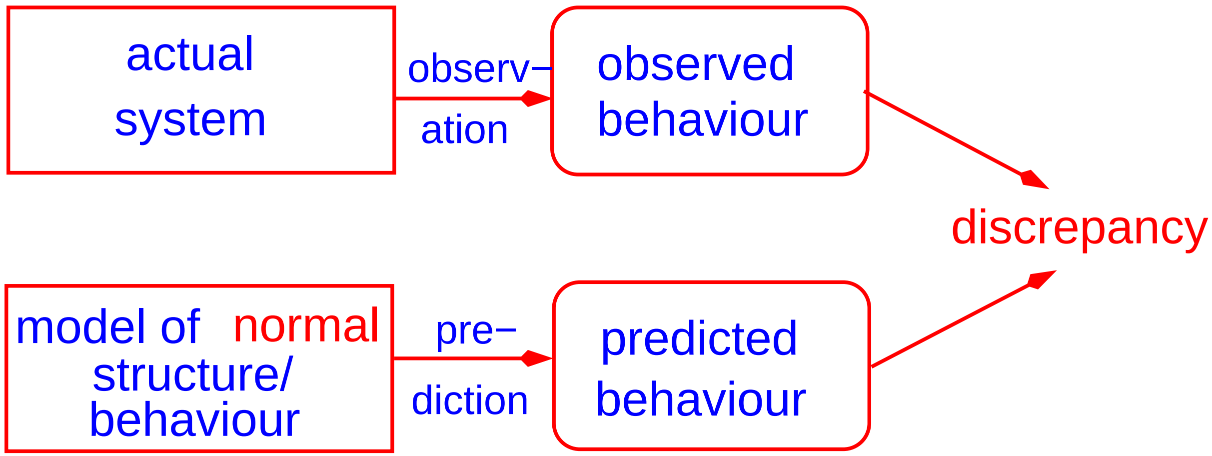
Consistency-based diagnosis (cont.)



Difference between predicted behaviour and observed behaviour ⇒ defect!

Originators:

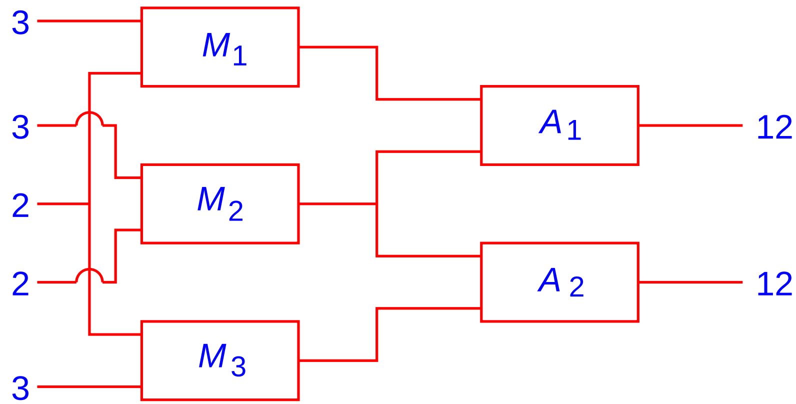
- R. Reiter, “A Theory of diagnosis from first principles”, *Artificial Intelligence*, vol. 32, 57–95, 1987.
- J. de Kleer, A.K. Macworth, and R. Reiter, “Characterising diagnoses and systems”, *Artificial Intelligence*, vol. 52, 197–222, 1992.

Plan for today

- Quick revision of basic concepts
- There are some opimisations possible for the hitting-set algorithm
- Consistency-based diagnosis is an example of non-monotonic reasoning (quite common in AI). We show why this is the case
- Finally, dagnoses can be seen as hypotheses that are revised when new observations are made. We extend the theory in this way

$\Longrightarrow$ First, revision!

System specification



Multiplier-adder

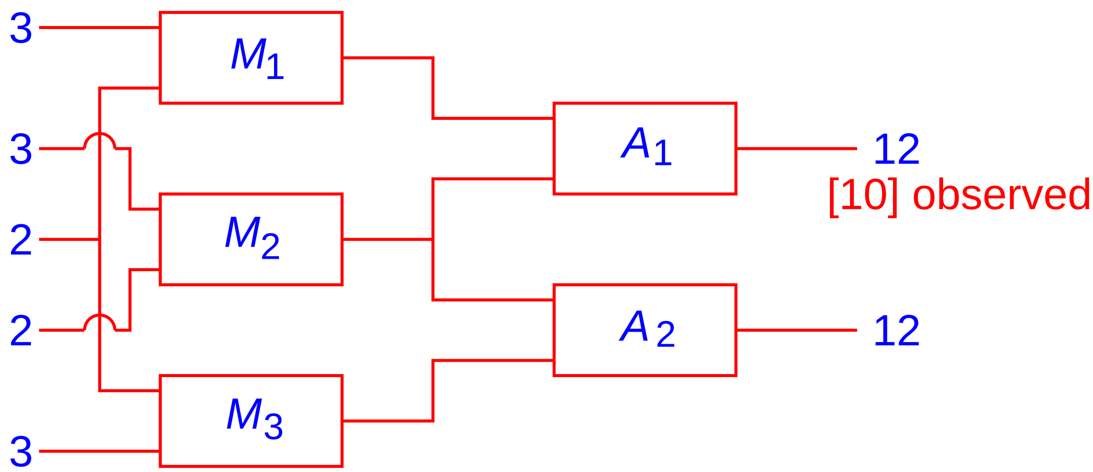
SYStem specification $\mathrm{SYS} = (\mathrm{SD}, \mathrm{COMPS})$:

- SD (System Description):

  $$\forall x((\mathsf{MUL}(x) \wedge \neg \mathrm{Ab}(x)) \rightarrow \mathrm{in}_1(x) \times \mathrm{in}_2(x) = \mathrm{out}(x))$$
  $$\forall x((\mathsf{ADD}(x) \wedge \neg \mathrm{Ab}(x)) \rightarrow \mathrm{in}_1(x) + \mathrm{in}_2(x) = \mathrm{out}(x))$$
  $$\mathrm{MUL}(M_1), \mathrm{MUL}(M_2), \mathrm{MUL}(M_3), \mathrm{ADD}(A_1), \mathrm{ADD}(A_2)$$
  $$\mathrm{in}_1(A_1) = \mathrm{out}(M_1), \mathrm{in}_2(A_1) = \mathrm{out}(M_2)$$
  $$\mathrm{in}_1(A_2) = \mathrm{out}(M_2), \mathrm{in}_2(A_2) = \mathrm{out}(M_3)$$

- $\mathrm{COMPS} = \{M_1, M_2, M_3, A_1, A_2\}$

Diagnostic problem



- System specification $\mathrm{SYS} = (\mathrm{SD}, \mathrm{COMPS})$
- Diagnostic problem $\mathrm{DP} = (\mathrm{SYS}, \mathrm{OBS})$, with OBS a set of observations

Example:

$$\mathrm{OBS} = \{\mathrm{in}_1(M_1) = 3, \mathrm{in}_2(M_1) = 2, \mathrm{in}_1(M_2) = 3, \mathrm{in}_2(M_2) = 2, \mathrm{in}_1(M_3) = 2, \mathrm{in}_2(M_3) = 3, \mathrm{out}(A_1) = 10, \mathrm{out}(A_2) = 12\}$$

Diagnosis

- Diagnostic problem $\mathrm{DP} = (\mathrm{SYS}, \mathrm{OBS})$, with OBS a set of observations
- A diagnosis D: smallest (subset minimal) set of components, such that

$$\mathrm{SD} \cup \mathrm{OBS} \cup \underbrace{\{\mathrm{Ab}(c) \mid c \in D\}}_{\text{Faulty components}} \cup \underbrace{\{\neg\mathrm{Ab}(c) \mid c \in \mathrm{COMPS} - D\}}_{\text{Nonfaulty components}}$$

  is consistent
- Remark: $\{\mathrm{Ab}(c) \mid c \in D\}$ can be omitted (why?)

For the multiplier-adder:
$D = \{A_1\}, \{M_1\}, \{M_2, M_3\}, \{A_2, M_2\}$

Conflict set and hitting set

Let $\mathrm{CS} \subseteq \mathrm{COMPS}$ be a set of components, then CS is called a conflict set iff

$$\mathrm{SD} \cup \mathrm{OBS} \cup \{\neg \mathrm{Ab}(c) \mid c \in \mathrm{CS}\}$$

is inconsistent

Proposition: For each $D \subseteq \mathrm{COMPS}$ that is a diagnosis and each conflict set CS it holds that: $D \cap \mathrm{CS} \neq \varnothing$

Theorem: D is a diagnosis for diagnostic problem $\mathrm{DP} = (\mathrm{SYS}, \mathrm{OBS})$ iff D is a minimal hitting set for all conflict sets of DP

Hitting-set tree

Let F be a set of sets

- Let $T = (V, E, l_V, l_E)$ be a labelled tree, with
- node labels

$$l_V(v) = \begin{cases} S & \text{if } S \in F,\, S \neq \varnothing \\ \checkmark & \text{otherwise} \end{cases}$$

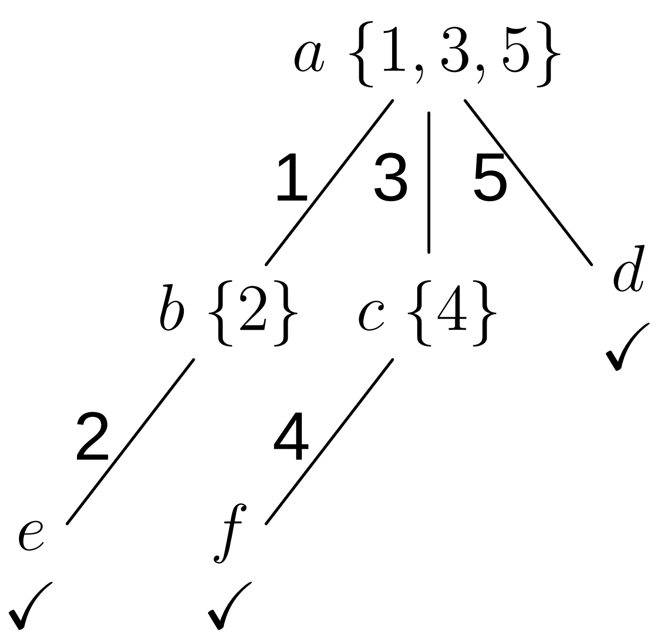
  and
- edge labels if $l_V(v) = S$ then $\forall s \in S$: $l_E(v, v_s) = s$

Hitting sets

The hitting set $H(v)$ for node v is defined as:

$$H(v) = \{l_E(u, w) \mid (u, w) \text{ is on the path from the root to } v\}$$

Example (incorrect why?):



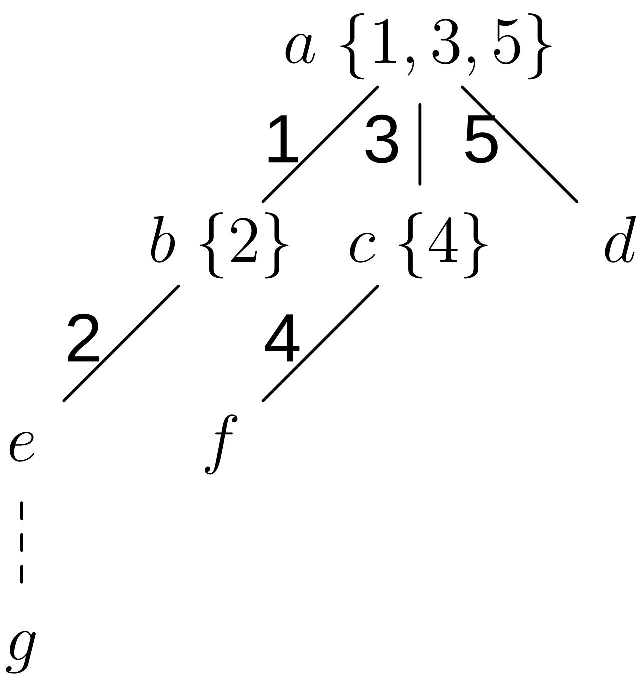
- $l_V(a) = \{1, 3, 5\}$, $l_V(b) = \{2\}$, $l_V(e) = \checkmark$, etc.
- $l_E(a, b) = 1$, $l_E(a, c) = 3$, $l_E(b, e) = 2$, etc.
- $H(a) = \emptyset$
- $H(b) = \{1\}$
- $H(e) = \{1, 2\}$
- $H(f) = \{3, 4\}$

Hitting sets

The hitting set $H(v)$ for node v is defined as:

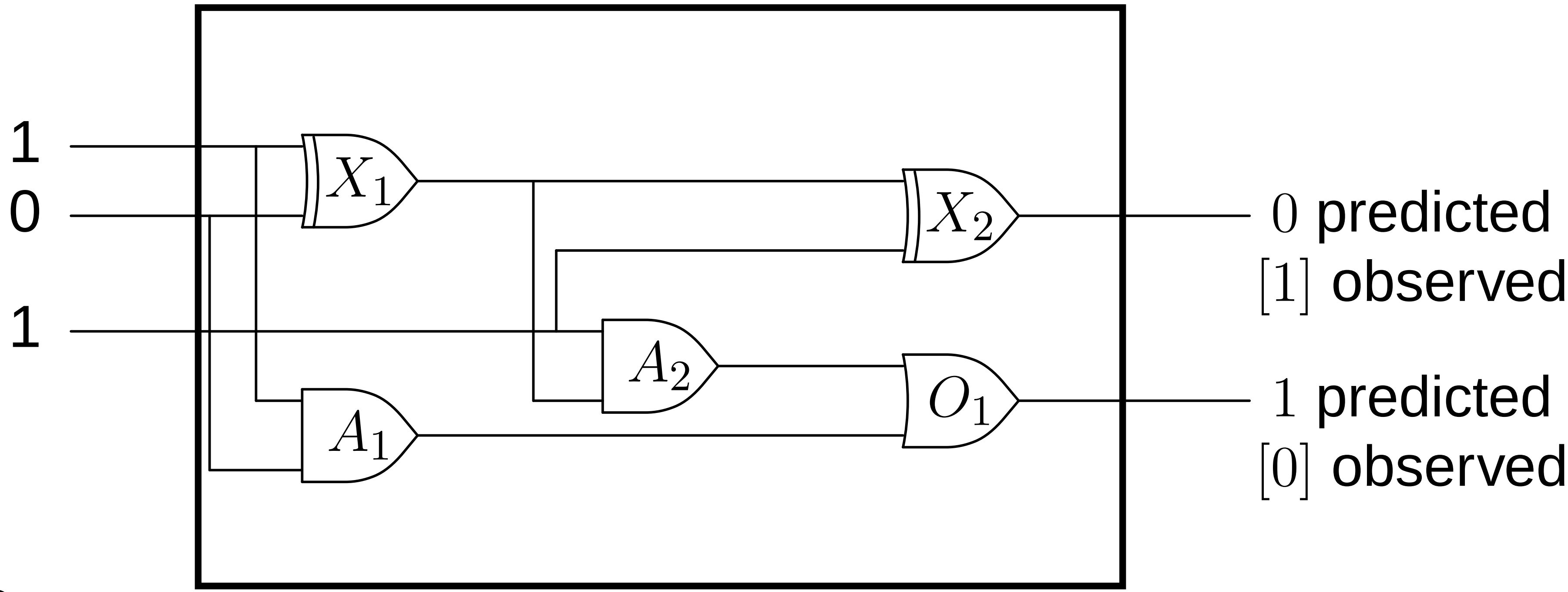
$$H(v) = \{l_E(u,w) \mid (u,w) \text{ is on the path from the root to } v\}$$

Example (correct):



- $l_V(a) = \{1,3,5\}$, $l_V(b) = \{2\}$, $l_V(e) = \checkmark$, etc.
- $l_E(a,b) = 1$, $l_E(a,c) = 3$, $l_E(b,e) = 2$, etc.
- $H(a) = \emptyset$
- $H(b) = \{1\}$
- $H(e) = \{1,2\}$
- $H(f) = \{3,4\}$

Example: full-adder



SD:

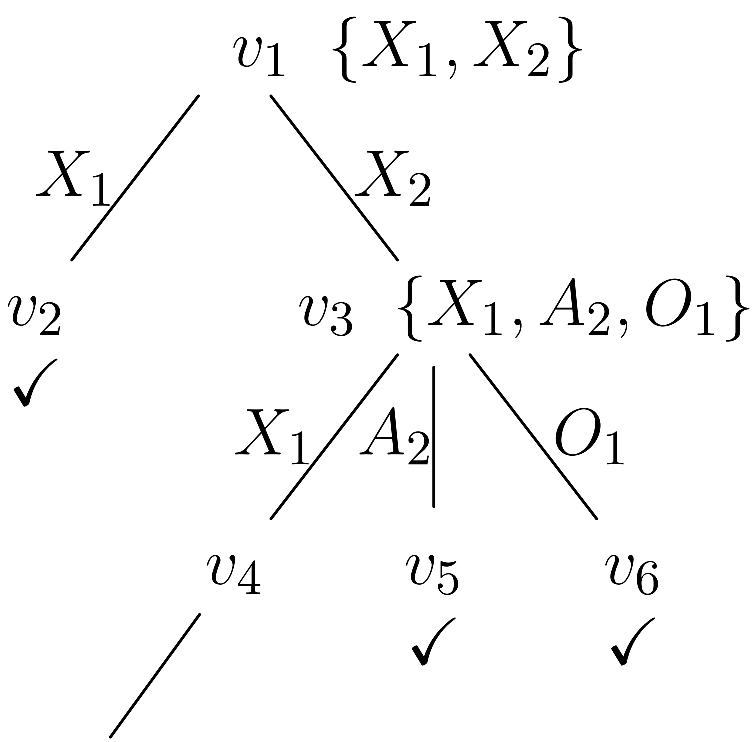
$$\forall x((\mathrm{ANDG}(x) \wedge \neg \mathrm{Ab}(x)) \rightarrow (\mathrm{out}(x) = \mathrm{in}_1(x) \wedge \mathrm{in}_2(x)))$$
$$\forall x((\mathrm{ORG}(x) \wedge \neg \mathrm{Ab}(x)) \rightarrow (\mathrm{out}(x) = \mathrm{in}_1(x) \vee \mathrm{in}_2(x)))$$

$$\vdots$$

$$\mathrm{ORG}(O_1), \mathrm{ANDG}(A_1), \mathrm{XORG}(X_1), \ldots$$

$$\mathrm{COMPS} = \{A_1, A_2, X_1, X_2, O_1\}$$

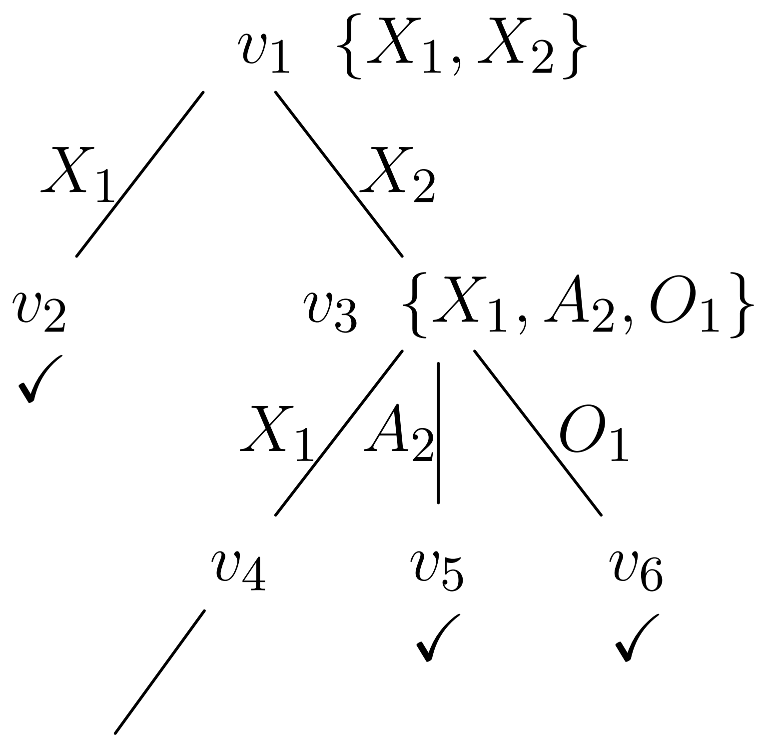
Example HS tree



1. $\mathrm{CS}_1 \leftarrow \mathrm{TP}(\mathrm{SD}, \mathrm{COMPS}, \mathrm{OBS})$;
   $\mathrm{CS}_1 \leftarrow \{X_1, X_2\}$
2. $\mathrm{CS}_2 \leftarrow \mathrm{TP}(\mathrm{SD}, \mathrm{COMPS} - \{X_1\}, \mathrm{OBS})$;
   $\mathrm{CS}_2 \leftarrow \checkmark$ (diagnosis found)
3. $\mathrm{CS}_3 \leftarrow \mathrm{TP}(\mathrm{SD}, \mathrm{COMPS} - \{X_2\}, \mathrm{OBS})$;
   $\mathrm{CS}_3 \leftarrow \{X_1, A_2, O_1\}$
4. $\vdots$

Diagnoses D: $\{X_1\}$, $\{X_2, A_2\}$, $\{X_2, O_1\}$
(note that $\{X_2, X_1\}$ not subset minimal)

Pruning of the hitting-set tree

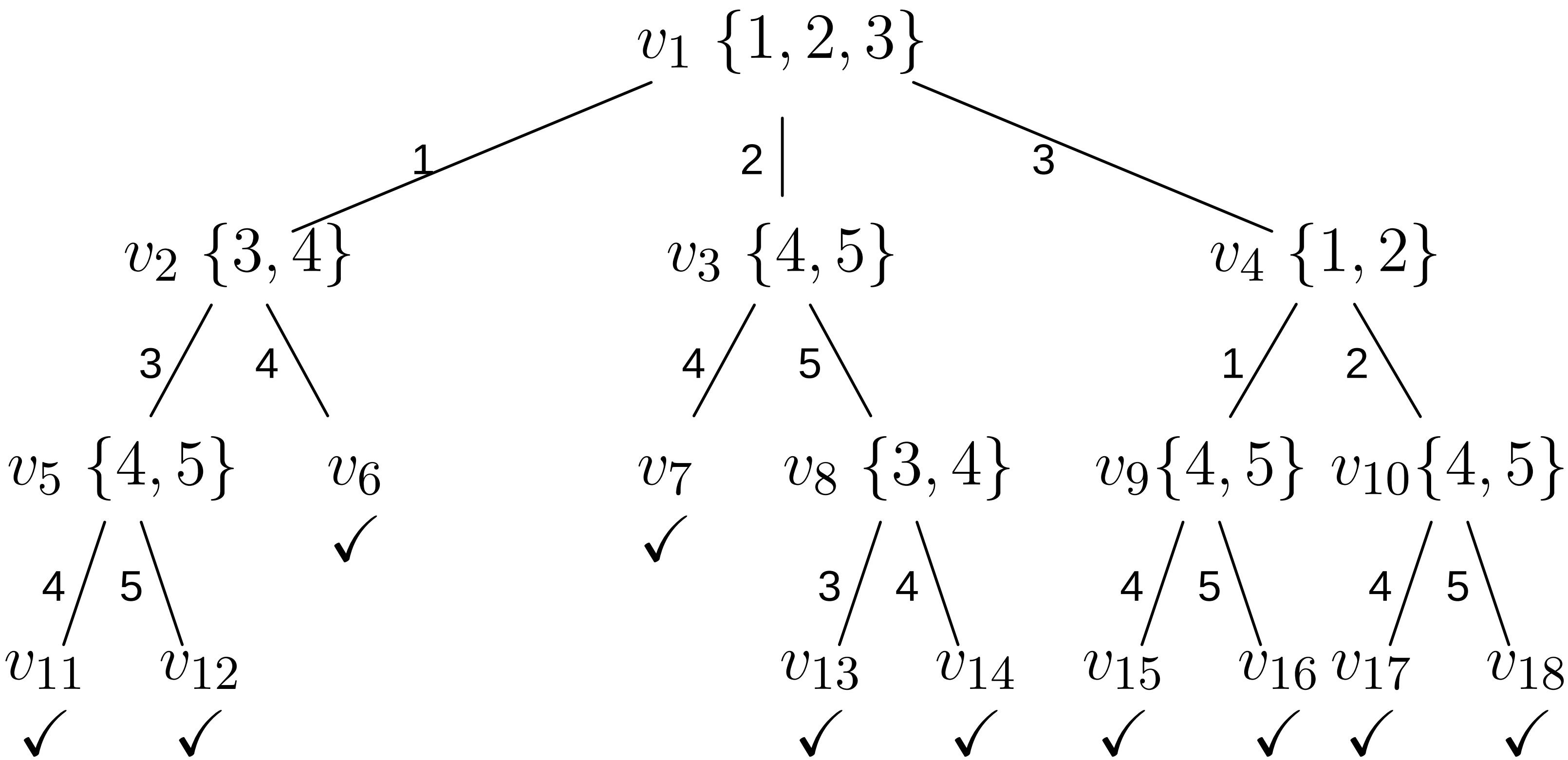


Note that there is no need to extend the hitting set $H(v_4)$ (as $\{X_2, X_1\}$ is not subset minimal)

$\Rightarrow$ pruning of the hitting-set tree

Hitting-set tree

$$F = \{\{1,2\},\{4,5\},\{1,2,3\},\{3,4\}\}$$

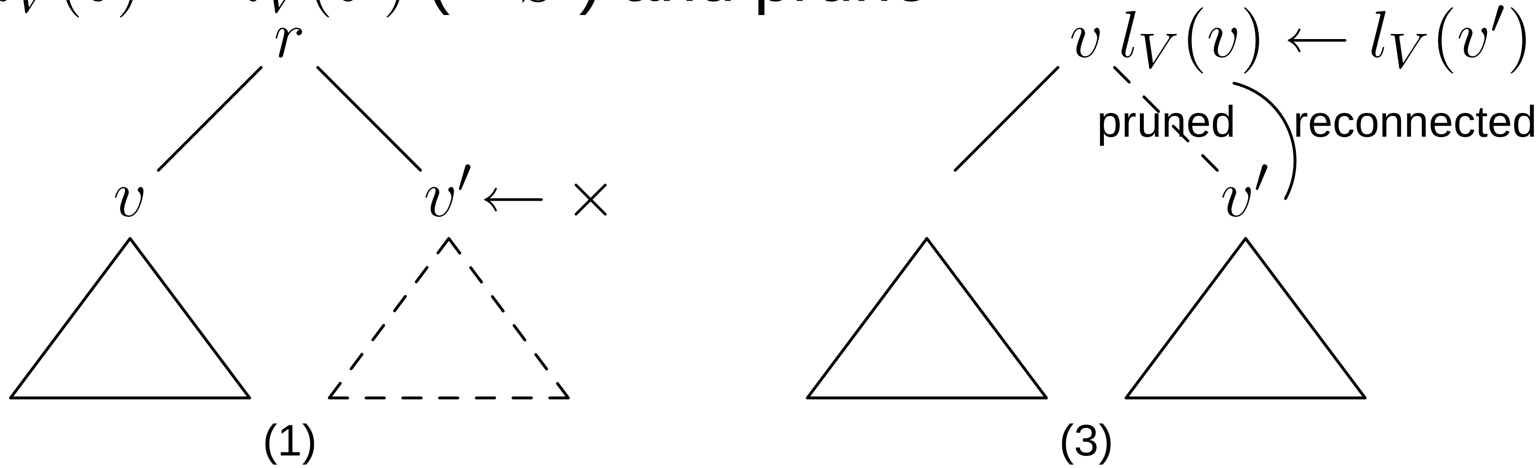


Minimal hitting set: $H(v)$ is subset-minimal and $l_V(v) = \checkmark$

Examples: $H(v_6) = \{1,4\} \subset H(v_{11})$; $H(v_7) = \{2,4\} \subset H(v_{14})$

Optimisation

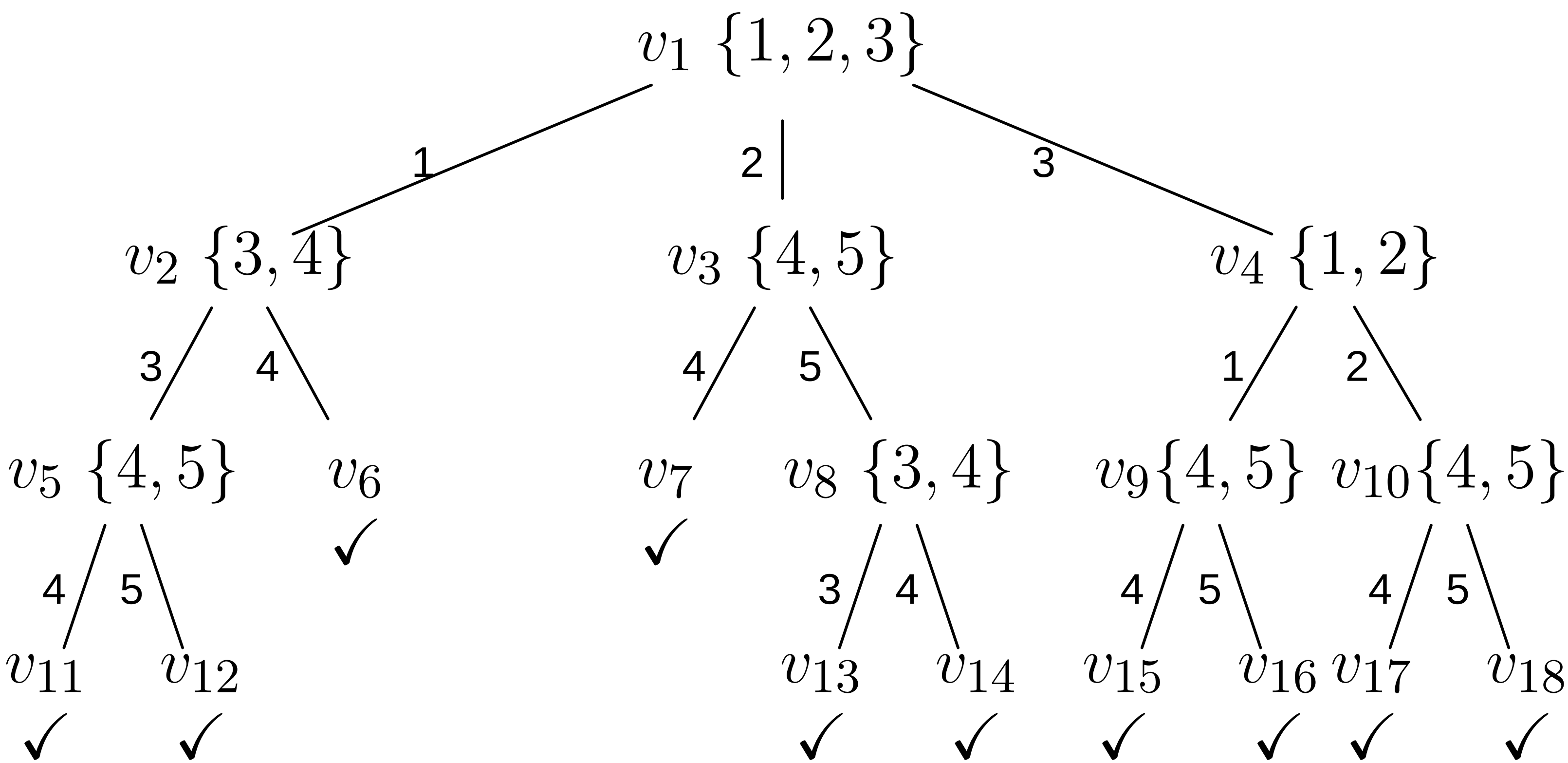
- Pruning the HS tree:
  1. $H(v) = H(v')$: prune the subtree with root v' and $l_V(v') = \times$
  2. $H(v) \subset H(v')$: ignore v', $l_V(v') = \times$
  3. for $l_V(v) = S, l_V(v') = S' \in F$ with $S' \subset S$: $l_V(v) \leftarrow l_V(v')$ ($= S'$) and prune



- Reuse of labels when F is dynamic (as in diagnosis): if $S' \in F$ and $H(v) \cap S' = \varnothing$, then $l_V(v) = S'$

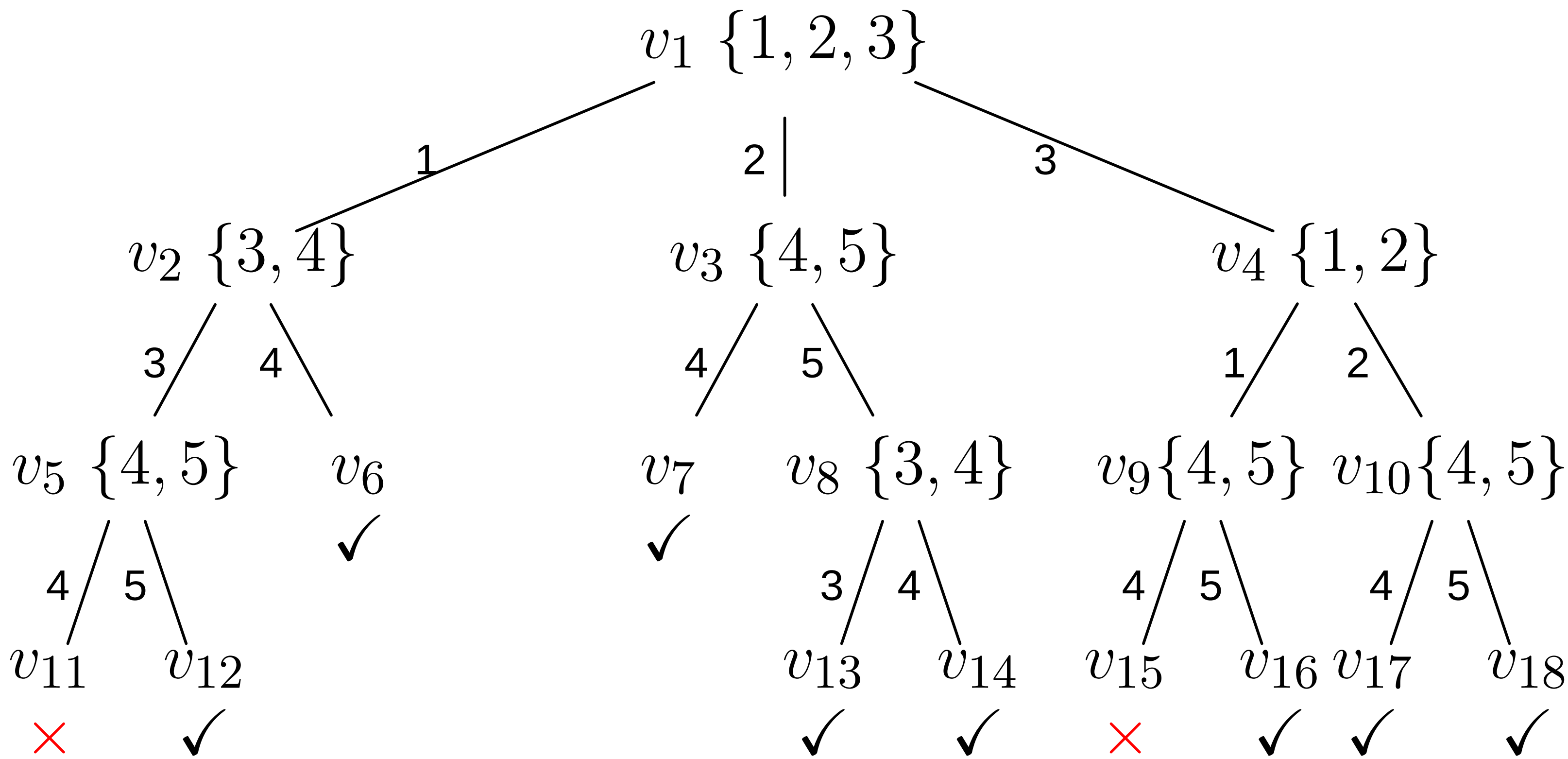
Example

$$F = \{\{1,2\},\{4,5\},\{1,2,3\},\{3,4\}\}$$



Example

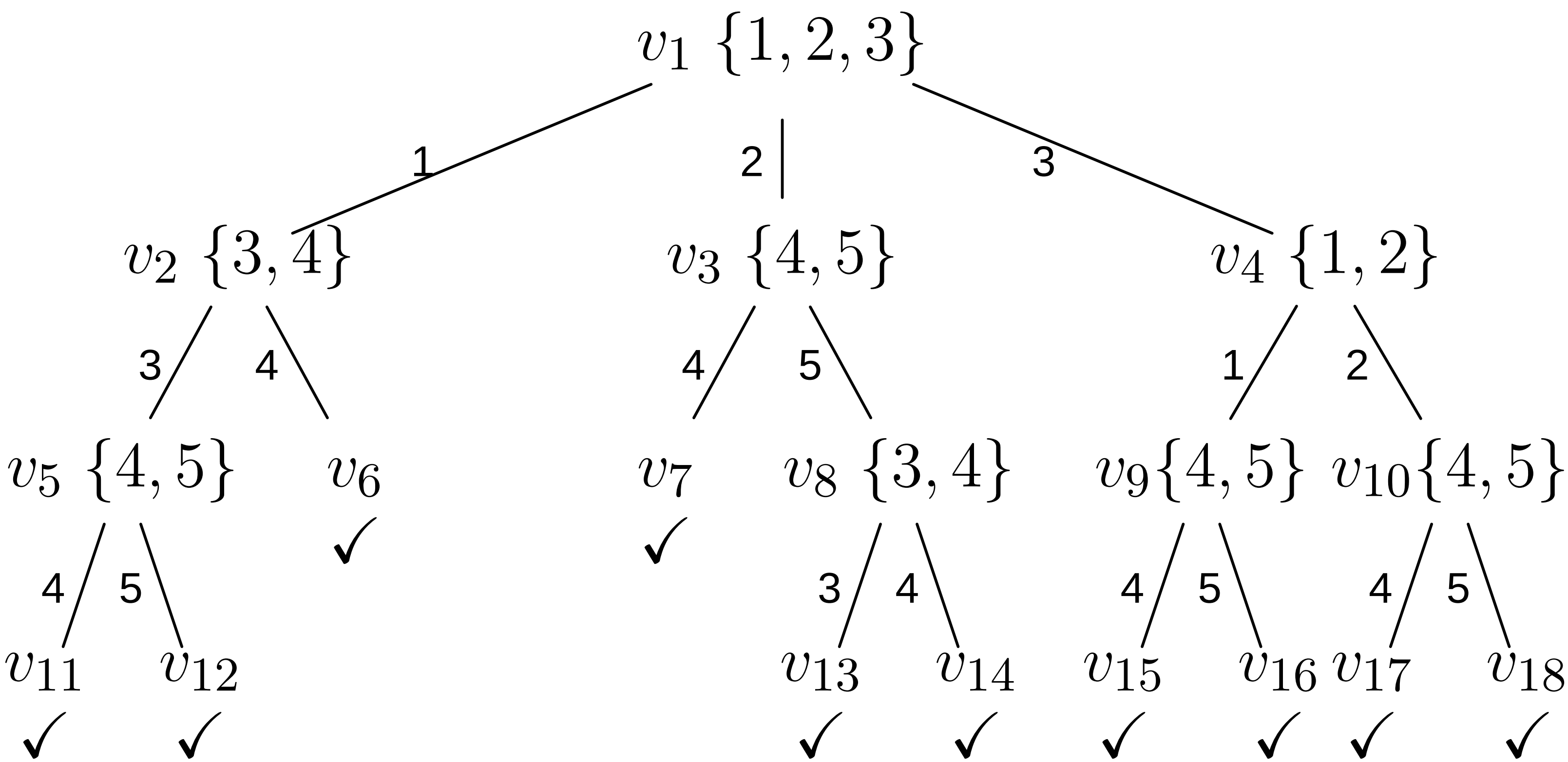
$$F = \{\{1,2\},\{4,5\},\{1,2,3\},\{3,4\}\}$$



Rule (1) and (2)

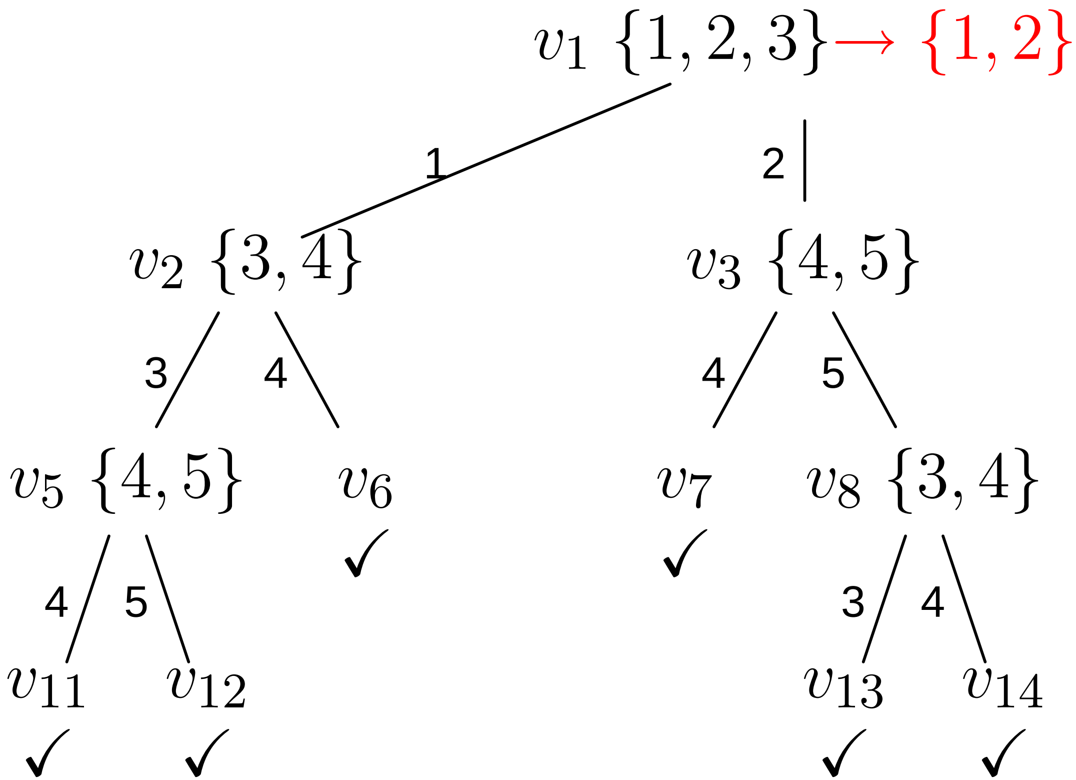
Example

$$F = \{\{1,2\},\{4,5\},\{1,2,3\},\{3,4\}\}$$



Example

$$F = \{\{1,2\},\{4,5\},\{1,2,3\},\{3,4\}\}$$



Rule (3)

Non-monotonic reasoning

- Knowledge base KB
- Add knowledge to KB and obtain new knowledge base KB'
- If $\mathrm{KB} \vdash$ Results and $\mathrm{KB}' \vdash$ Results$'$ then Results $\subseteq$ Results$'$ does not hold in general
  $\Rightarrow$ more knowledge does not always yield more results
- Consistency-based reasoning is an example of non-monotonic reasoning. Why?

$$\mathrm{SD} \cup \mathrm{OBS} \cup \{\neg \mathrm{Ab}(c) \mid c \in \mathrm{COMPS} - D\} \nvDash \bot$$

e.g., larger OBS or SD may make D smaller or different

Default logic

- $\mathrm{DT} = (W, R)$ is a default theory, where $W = \{\mathrm{Elephant(john)}\}$, i.e., John is an elephant, and the following default R:

$$\frac{\mathrm{Elephant}(x) : \mathrm{Grey}(x)}{\mathrm{Grey}(x)}$$

If being grey is consistent with our knowledge, conclude 'grey', so conclude $\mathrm{Grey(john)}$

- general form default

$$\frac{\textit{prerequisite} : \textit{justifications}}{\textit{consequent}}$$

Reasoning in default logic

Let $\mathrm{DT} = (W, R)$ be a default theory (W a set of logical formulas and a set of defaults R):

- $E = \mathrm{Th}(E)$ (so-called fixed point)
- $W \subseteq E$
- E includes the maximal set of conclusions obtained by applying defaults in R
- If $\frac{A:B_1,\ldots,B_n}{C} \in R$, $A \in E$ and $\neg B_1, \ldots, \neg B_n \notin W$, then $C \in E$

E is called an extension and Th is the derivation operator (deduction + default rule application)

Example

$\mathrm{DT} = (W, R)$, where

$$W = \{\mathrm{Elephant(clyde)}, \neg\mathrm{Grey(john)}\}$$

i.e., Clyde is an elephant and John is not grey, and the following default R:

$$\frac{\mathrm{Elephant}(x) : \mathrm{Grey}(x)}{\mathrm{Grey}(x)}$$

‘elephants are normally grey’

Extension: $E = \{\mathrm{Elephant(clyde)}, \mathrm{Grey(clyde)}, \neg\mathrm{Grey(john)}\}$

Diagnosis as non-monotonic reasoning

- Map a diagnostic problem to a default theory DT
- A diagnosis D predicts a formula φ iff

$$\mathrm{SD} \cup \mathrm{OBS} \cup \{\neg\mathrm{Ab}(c) \mid c \in \mathrm{COMPS} - D\} \cup \{\mathrm{Ab}(c) \mid c \in D\} \models \varphi$$

- Lemma: $\mathrm{DT} = (W, R)$ is a default theory with extension E iff

$$E = \mathrm{Th}\left(W \cup \{L \mid {:}L/L \in \Delta\}\right)$$

with subset-maximal set of defaults $\Delta \subseteq R$ such that

$$W \cup \{L \mid {:}L/L \in \Delta\} \not\models \bot$$

- Remark: $:\psi/\psi$ is a so-called normal default (default without prerequisite and justification that is the same as the conclusion)

Logical characterisation of diagnosis

Theorem. Let $\mathrm{DP} = (\mathrm{SYS}, \mathrm{OBS})$ be a diagnostic problem. Let

$$\mathrm{DT} = \left(\mathrm{SD} \cup \mathrm{OBS}, \left\{ \frac{: \neg \mathrm{Ab}(c)}{\neg \mathrm{Ab}(c)} \mid c \in \mathrm{COMPS} \right\}\right)$$

be a default theory with extension E, then D is a diagnosis for DP iff $E = \{\varphi \mid D$ predicts $\varphi\}$

Proof: E is an extension of DT, thus (Lemma):

$$\mathrm{SD} \cup \mathrm{OBS} \cup \left\{ \neg \mathrm{Ab}(c) \mid \frac{: \neg \mathrm{Ab}(c)}{\neg \mathrm{Ab}(c)} \in \Delta \right\} \nvDash \bot$$

with $\Delta \subseteq R$ such that Δ subset-maximal. Suppose that $D = \{c \mid c \in \mathrm{COMPS}, \frac{:\neg \mathrm{Ab}(c)}{\neg \mathrm{Ab}(c)} \notin \Delta\}$, then $\ldots$

Logical characterisation of diagnosis

Theorem. Let $\mathrm{DP} = (\mathrm{SYS}, \mathrm{OBS})$ be a diagnostic problem. Let

$$\mathrm{DT} = \left(\mathrm{SD} \cup \mathrm{OBS}, \left\{ \frac{: \neg\mathrm{Ab}(c)}{\neg\mathrm{Ab}(c)} \mid c \in \mathrm{COMPS} \right\}\right)$$

be a default theory with extension E, then D is a diagnosis for DP iff $E = \{\varphi \mid D \text{ predicts } \varphi\}$

Proof (continued):

$$\left\{ \neg\mathrm{Ab}(c) \mid \frac{: \neg\mathrm{Ab}(c)}{\neg\mathrm{Ab}(c)} \in \Delta \right\} = \{\neg\mathrm{Ab}(c) \mid c \in \mathrm{COMPS} - D\}$$

Thus, $E = \mathrm{Th}(\mathrm{SD} \cup \mathrm{OBS} \cup \{\neg\mathrm{Ab}(c) \mid c \in \mathrm{COMPS} - D\})$, and $E = \{\varphi \mid D \text{ predicts } \varphi\}$

Example

$\mathrm{DP} = (\mathrm{SYS}, \mathrm{OBS})$, with

$$\begin{aligned}\mathrm{SD} = \{&\forall x((\mathrm{ANDG}(x) \wedge \neg\mathrm{Ab}(x)) \rightarrow (\mathrm{out}(x) = \mathrm{and}(\mathrm{in}_1(x), \mathrm{in}_2(x)))),\\ &\forall x((\mathrm{XORG}(x) \wedge \neg\mathrm{Ab}(x)) \rightarrow (\mathrm{out}(x) = \mathrm{xorg}(\mathrm{in}_1(x), \mathrm{in}_2(x)))),\\ &\mathrm{ANDG}(A), \mathrm{XORG}(X),\\ &\mathrm{out}(A) = \mathrm{in}_1(X)\}\end{aligned}$$

$\mathrm{COMPS} = \{A, X\}$ and $\mathrm{OBS} = \{\mathrm{in}_1(A) = 1, \mathrm{in}_2(A) = 1, \mathrm{in}_2(X) = 0, \mathrm{out}(A) = 0, \mathrm{out}(X) = 1\}$

Default theory $\mathrm{DT} = (\mathrm{SD} \cup \mathrm{OBS}, R)$, with

$$R = \left\{ \frac{: \neg\mathrm{Ab}(A)}{\neg\mathrm{Ab}(A)}, \frac{: \neg\mathrm{Ab}(X)}{\neg\mathrm{Ab}(X)} \right\}$$

$E = \mathrm{Th}(\mathrm{SD} \cup \mathrm{OBS} \cup \{\neg\mathrm{Ab}(X)\})$, e.g., $\mathrm{Ab}(A) \in E$ (we can predict that A is abnormal)

Extra measurements

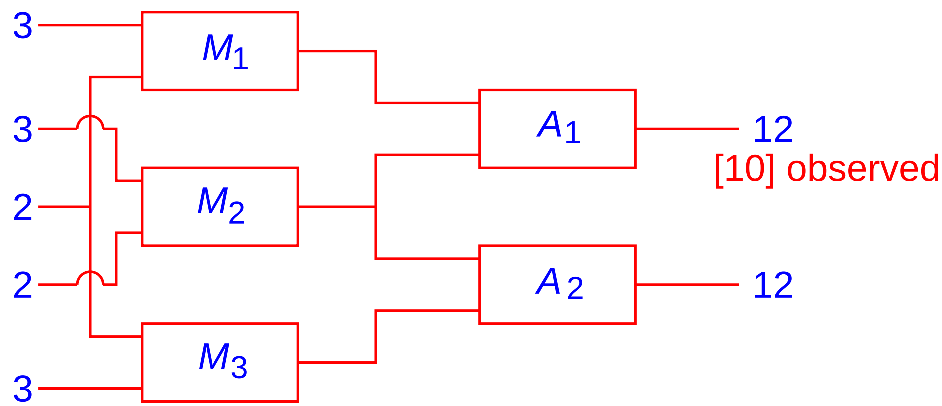
Recall: a diagnosis D predicts a formula φ iff

$$\mathrm{SD} \cup \mathrm{OBS} \cup \{\neg \mathrm{Ab}(c) \mid c \in \mathrm{COMPS} - D\} \vDash \varphi$$

if φ is a set of extra observations (measurements), then:

1. Every diagnosis D for $\mathrm{DP} = (\mathrm{SYS}, \mathrm{OBS})$ that predicts φ is also a diagnosis for $\mathrm{DP} = (\mathrm{SYS}, \mathrm{OBS} \cup \{\varphi\})$, i.e., the measurement φ confirms D
2. No diagnosis for $\mathrm{DP} = (\mathrm{SYS}, \mathrm{OBS})$ that predicts $\neg\varphi$ is also a diagnosis for $\mathrm{DP} = (\mathrm{SYS}, \mathrm{OBS} \cup \{\varphi\})$, i.e., the measurment φ disconfirms D
3. Any diagnosis D for $\mathrm{DP} = (\mathrm{SYS}, \mathrm{OBS} \cup \{\varphi\})$ which is not a diagnosis for $\mathrm{DP}' = (\mathrm{SYS}, \mathrm{OBS})$ is a strict superset of a diagnosis of DP' which predicts $\neg\varphi$

Example

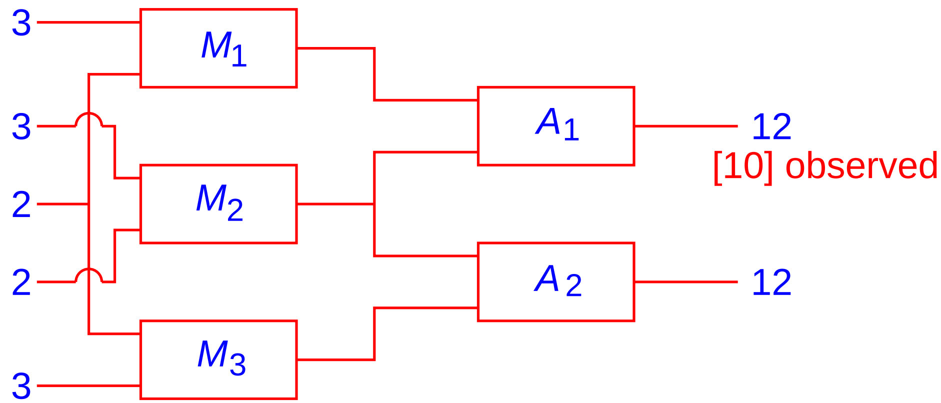


- Diagnostic problem $\mathrm{DP} = (\mathrm{SYS}, \mathrm{OBS})$, with
- Set of observations

$$\begin{aligned}\mathrm{OBS} = \{&\mathrm{in}_1(M_1) = 3, \mathrm{in}_2(M_1) = 2, \mathrm{in}_1(M_2) = 3, \mathrm{in}_2(M_2) = 2,\\ &\mathrm{in}_1(M_3) = 2, \mathrm{in}_2(M_3) = 3, \mathrm{out}(A_1) = 10, \mathrm{out}(A_2) = 12\}\end{aligned}$$

- Predictions w.r.t. $\mathrm{out}(M_2)$: diagnosis $D_1 = \{M_1\}$ predicts $\mathrm{out}(M_2) = 6$, $D_2 = \{A_1\}$ predicts $\mathrm{out}(M_2) = 6$, $D_3 = \{M_2, M_3\}$ predicts $\mathrm{out}(M_2) = 4$, $D_4 = \{M_2, A_2\}$ predicts $\mathrm{out}(M_2) = 4$

Example (continued)

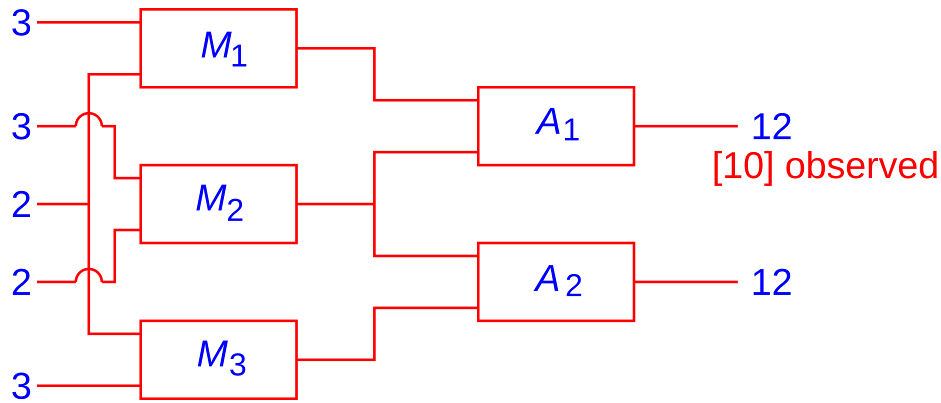


- Diagnostic problem $\mathrm{DP} = (\mathrm{SYS}, \mathrm{OBS})$, with
- Set of observations (with new one on M_2):

$$\mathrm{OBS} = \{\mathrm{in}_1(M_1) = 3, \mathrm{in}_2(M_1) = 2, \mathrm{in}_1(M_2) = 3, \mathrm{in}_2(M_2) = 2, \mathrm{in}_1(M_3) = 2, \mathrm{in}_2(M_3) = 3, \mathrm{out}(A_1) = 10, \mathrm{out}(A_2) = 12, \mathrm{out}(M_2) = 5\}$$

- The new observation $\mathrm{out}(M_2) = 5$ disconfirms all previous diagnoses

Example (continued)



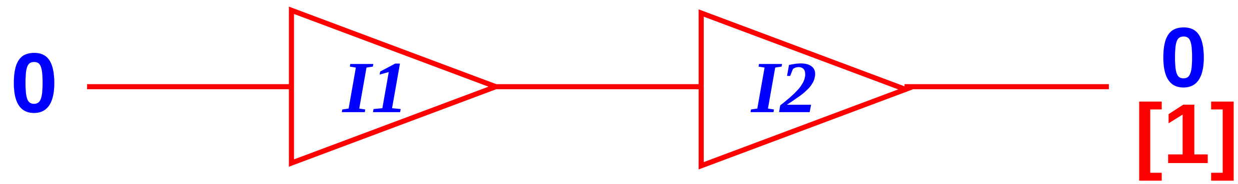
- Diagnostic problem $\mathrm{DP} = (\mathrm{SYS}, \mathrm{OBS})$, with
- Set of observations (with new one on M_2):

$$\mathrm{OBS} = \{\mathrm{in}_1(M_1) = 3, \mathrm{in}_2(M_1) = 2, \mathrm{in}_1(M_2) = 3, \mathrm{in}_2(M_2) = 2, \mathrm{in}_1(M_3) = 2, \mathrm{in}_2(M_3) = 3, \mathrm{out}(A_1) = 10, \mathrm{out}(A_2) = 12, \mathrm{out}(M_2) = 5\}$$

- New diagnoses: $D'_1 = \{M_1, M_2, M_3\}$, $D'_2 = \{M_1, M_2, A_2\}$, $D'_3 = \{M_2, M_3, A_1\}$, $D'_4 = \{M_2, A_1, A_2\}$; note supersets of the old diagnoses, e.g., $D_1, D_3 \subseteq D'_1$ (case 3)

Definition of diagnosis revisited

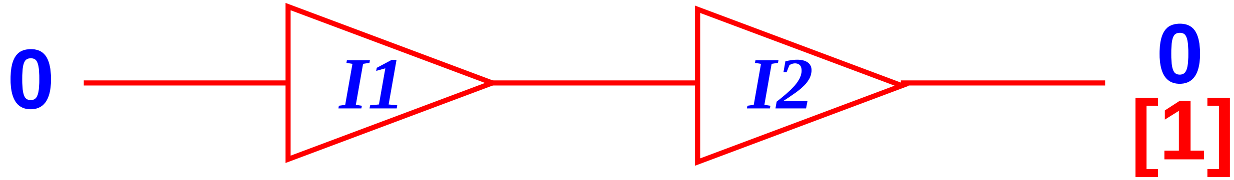
Example two inverters I_k, $k = 1, 2$:



$$\begin{aligned} \mathrm{SD} = \{ & \forall x((\mathrm{INV}(x) \wedge \neg \mathrm{Ab}(x)) \rightarrow \neg(\mathrm{out}(x) = \mathrm{in}(x))), \\ & \mathrm{INV}(I_1), \\ & \mathrm{INV}(I_2)\} \end{aligned}$$

- $\mathrm{OBS} = \{\mathrm{in}(I_1) = 0, \mathrm{out}(I) = 1\}$
- Diagnoses: $\{I_1\}$ and $\{I_2\}$
- However, $\{I_1, I_2\}$ might also be a diagnosis (only excluded because of definition using subset-minimality condition)

Faul models



$$
\begin{aligned}
\mathrm{SD} = \{&\forall x((\mathrm{INV}(x) \land \neg \mathrm{Ab}(x)) \rightarrow \neg(\mathrm{out}(x) = \mathrm{in}(x))),\\
&\forall x((\mathrm{INV}(x) \land \mathrm{Ab}(x)) \rightarrow (\mathrm{Stuck\_at0}(x) \lor (\mathrm{out}(x) = \mathrm{in}(x)))),\\
&\forall x(\mathrm{Stuck\_at0}(x) \rightarrow \mathrm{out}(x) = 0),\\
&\mathrm{INV}(I_1),\\
&\mathrm{INV}(I_2)\}
\end{aligned}
$$

Knowledge about behaviour for $\mathrm{Ab}(c)$ is called fault model

- $\{I_1\}$ and $\{I_2\}$ still diagnoses,
- $\{I_1, I_2\}$ no longer a diagnosis, because

$$\mathrm{SD} \cup \mathrm{OBS} \cup \{\mathrm{Ab}(I_1), \mathrm{Ab}(I_2)\} \vDash \bot$$

Conclusions

- Consistency-based diagnosis popular for trouble shooting of equipment and devices
- Extensions: temporal behaviour and continuous behaviour
- Diagnoses may be ranked using probability theory and entropy (General Diagnostic Engine, GDE)
- Software:
  - GDE/ATMS (Palo Alto Research Center): http://www2.parc.com/spl/members/dekleer/
  - Leancop (University Darmstad/Potsdam): http://www.leancop.de/
  - AILog!