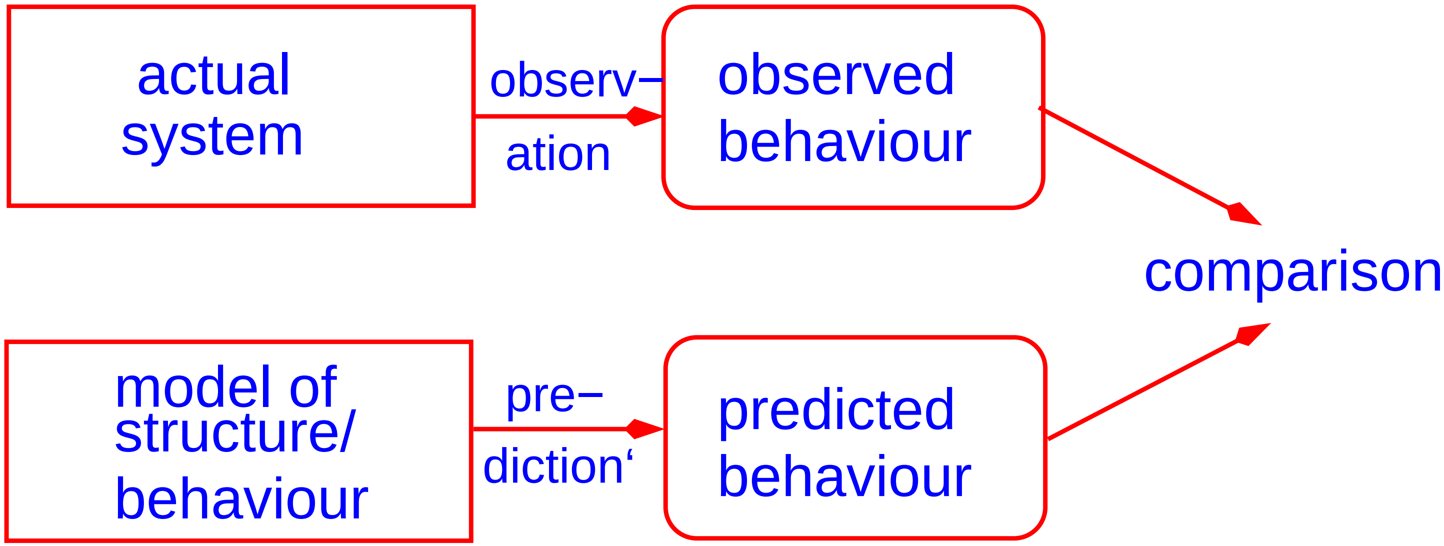
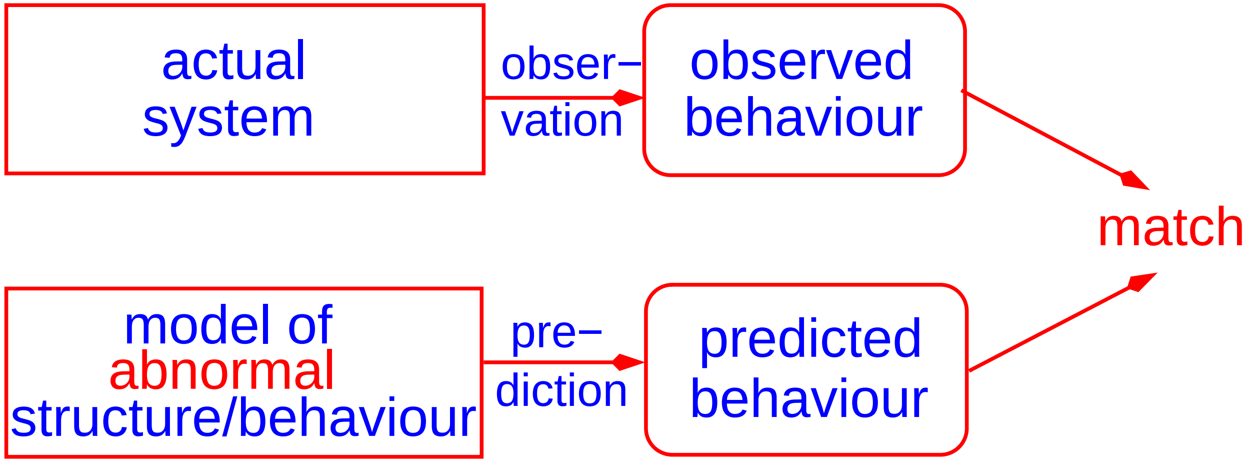
Model-based Reasoning – Abduction



- Model: representation of normal or abnormal behaviour and, possibly, internal structure
- Formalisation of model-based diagnosis:
  - *consistency-based diagnosis* (normal behaviour), and
  - *abductive diagnosis* (abnormal behaviour)

Abductive diagnosis

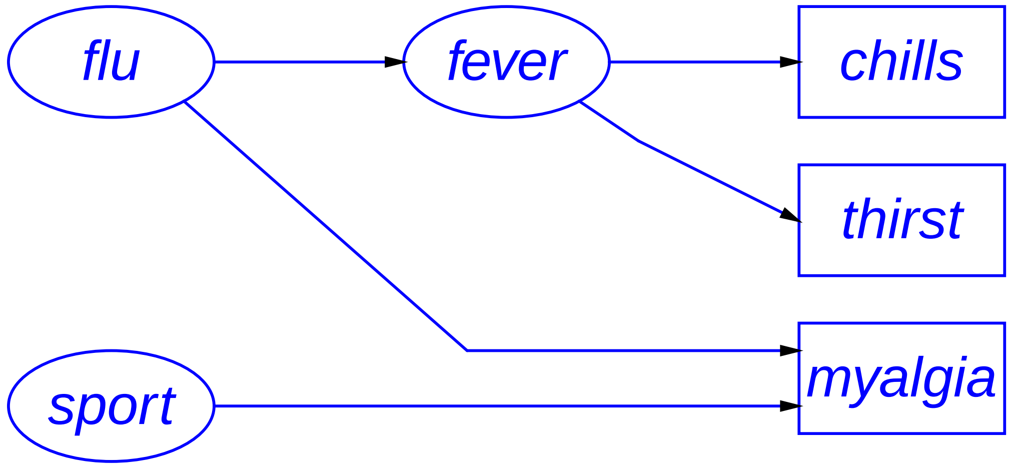


Correspondence between predicted *abnormal* behaviour and observed behaviour $\Rightarrow$ defect!

Originators:

- L. Console, D. Theseider Dupré and P. Torasso, "A theory of diagnosis for incomplete causal models", In: *IJCAI'89*, 1311–1317, 1989
- D. Poole, "Explanation and prediction: an architecture for default and abductive reasoning", *Computational Intelligence*, vol. 5, nr. 2, 97–110, 1989
- Y. Peng and J.A. Reggia, *Abductive Inference Models for Diagnostic Problem Solving*, New York: Springer-Verlag, 1990

Causal models



- Causality: combination of causes gives rise to effects

  flu causes fever
  fever causes chills
  fever causes thirst
  flu causes myalgia
  sport also causes myalgia

- Using logic: $(\text{Cause}_1 \wedge \cdots \wedge \text{Cause}_n) \rightarrow \text{Effect}$
- Example:

  fever $\longrightarrow$ *chills*

  fever $\longrightarrow$ *thirst*

  sport $\longrightarrow$ *myalgia* $\cdots$

Causality and implication

x causes y: Causes(x, y)

Axiomatisation:

- transitivity:
  $\forall x \forall y \forall z\,((\text{Causes}(x,z) \wedge \text{Causes}(z,y)) \rightarrow \text{Causes}(x,y))$
- antisymmetry: $\forall x \forall y (\text{Causes}(x,y) \rightarrow \neg\text{Causes}(y,x))$
- reflexivity: $\forall x$ Causes(x,x) (by definition this excludes antisymmetry)

With implication: x causes $y \equiv x \rightarrow y$

- transitivity ✓: $\{P \rightarrow Q, Q \rightarrow R\} \models P \rightarrow R$
- no antisymmetry but contraposition: $\{P \rightarrow Q, \neg Q\} \models \neg P$ ($P \rightarrow Q \equiv \neg Q \rightarrow \neg P$)
- reflexivity ✓: $\models P \rightarrow P$

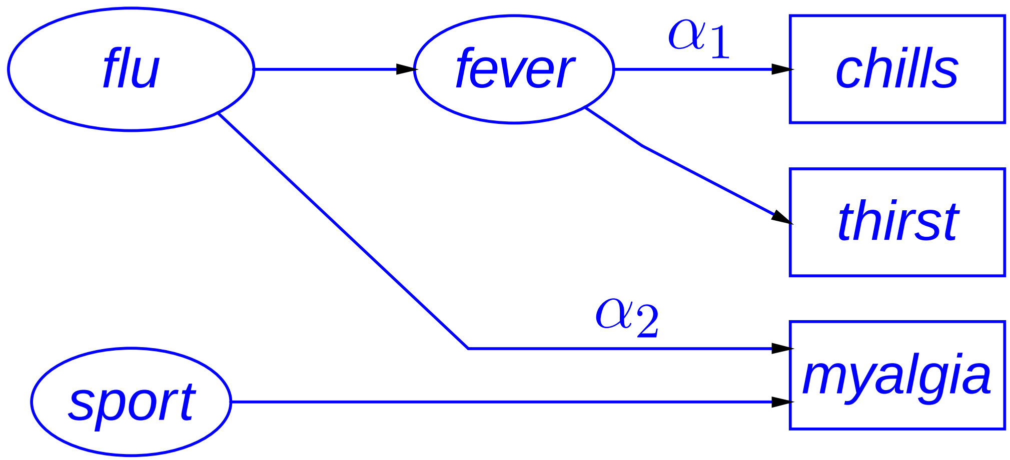
Weak and strong causality

- Strong causality: $C \rightarrow E$
  "If C present, then E *must* also be present"
- Weak causality: $(C \wedge \alpha) \rightarrow E$ or simplified $C \wedge \alpha \rightarrow E$
  "If C present, then E *may* be present" (α is incompleteness assumption)

$C \rightarrow E$

$(C \wedge \alpha) \rightarrow E$

Weak and strong causality



- Strong causality: $C \rightarrow E$
- Weak causality ("may cause"): $C \wedge \alpha \rightarrow E$ (α is incompleteness assumption)
- Example:

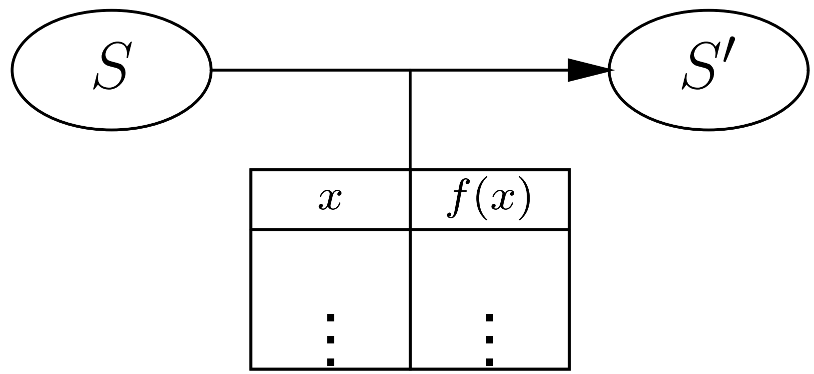
  fever $\wedge\, \alpha_1 \longrightarrow$ *chills*

  fever $\longrightarrow$ *thirst*

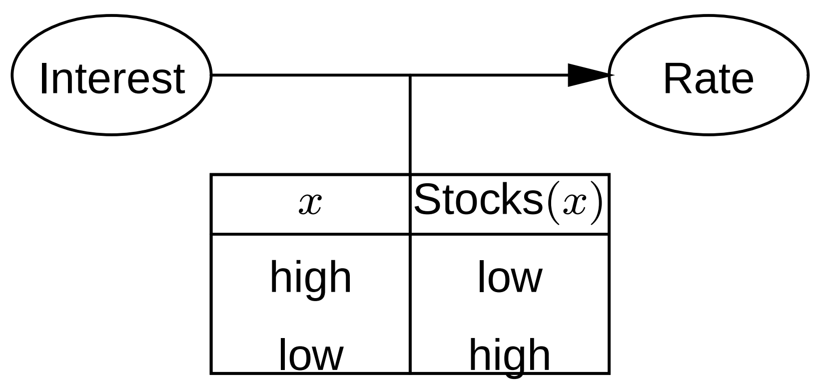
  sport $\longrightarrow$ *myalgia*

Generalisation

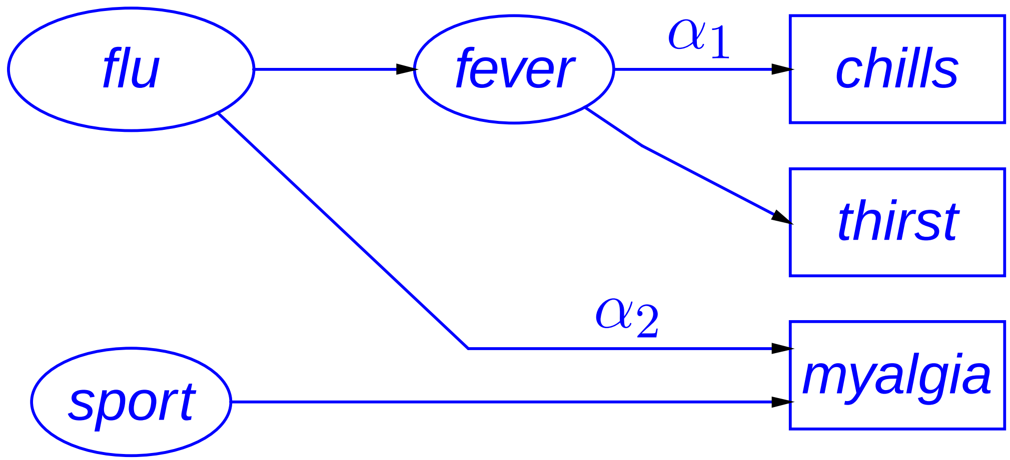
- Representation for classes of cases
- $\forall x(S(x) \rightarrow S'(f(x)))$, with S and S' states



- $\forall x(\text{Interest}(x) \rightarrow \text{Rate}(\text{Stocks}(x)))$

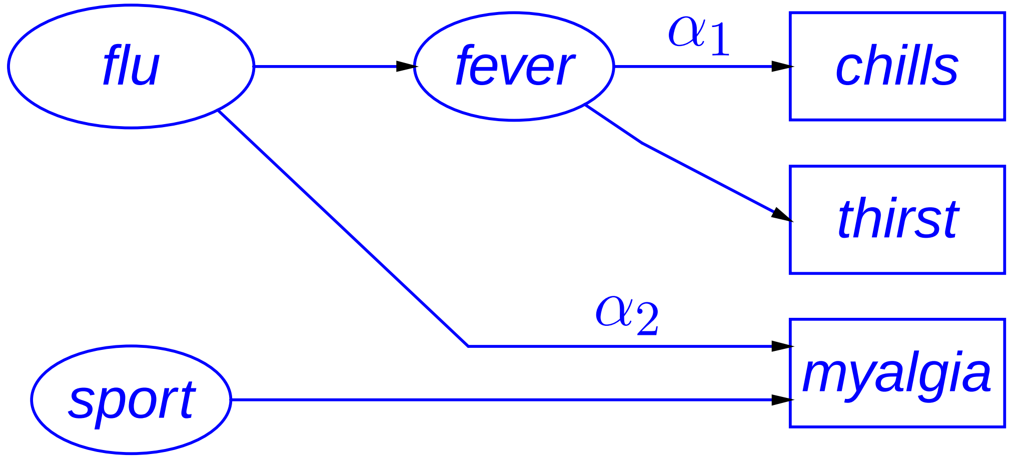


Prediction



- Causal specification: $\Sigma = (\Delta, \Phi, \mathcal{R})$, with:
  - Δ: potential causes and incompleteness assumptions
  - Φ: facts that can be observed
  - $\mathcal{R}$: causal model
- Prediction $V \subseteq \Delta$: $\boxed{\mathcal{R} \cup V \models E}$,
  with $E \subseteq \Phi$ (E can be observed)

Example prediction



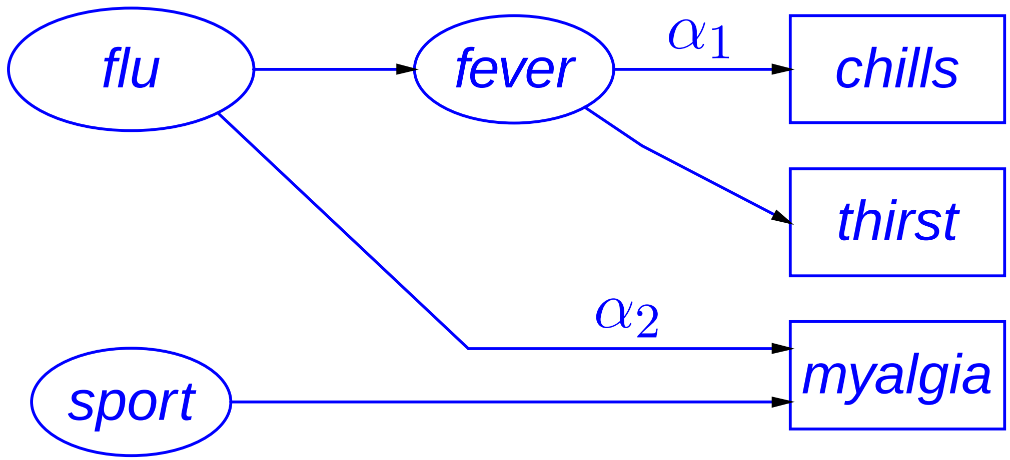
Causal specification: $\Sigma = (\Delta, \Phi, \mathcal{R})$

- Example 1: $\mathcal{R} \cup \{\textit{flu}, \alpha_1\} \models \{\textit{chills}, \textit{thirst}\}$
- Example 2:

  $$\mathcal{R} \cup \{\textit{flu}, \alpha_1, \alpha_2\} \models \{\textit{chills}, \textit{thirst}, \textit{myalgia}\}$$

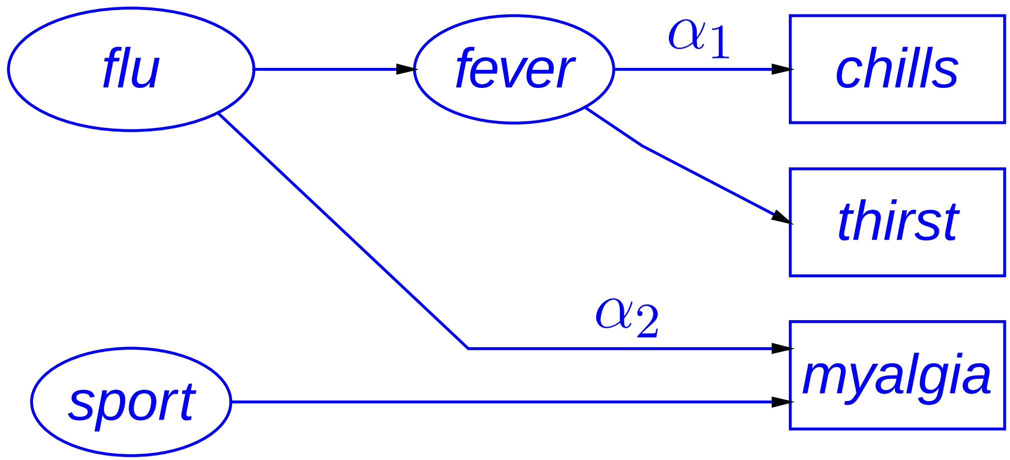
- Example 3: $\mathcal{R} \cup \{\textit{sport}\} \models \textit{myalgia}$

Diagnostic problem



- Causal specification: $\Sigma = (\Delta, \Phi, \mathcal{R})$
- Observed facts: $F = \{\textit{myalgia}, \textit{thirst}\}$
- Diagnosis D?
  - (1) Prediction that explains F, formal: $\mathcal{R} \cup D \models F$
  - (2) $\cdots$ but that does not explain too much
- Example diagnoses: $D = \{\textit{flu}, \alpha_2\}$, $D' = \{\textit{sport}, \textit{flu}\}$ and $D'' = \{\textit{flu}, \alpha_1\}$?

Don't explain too much!

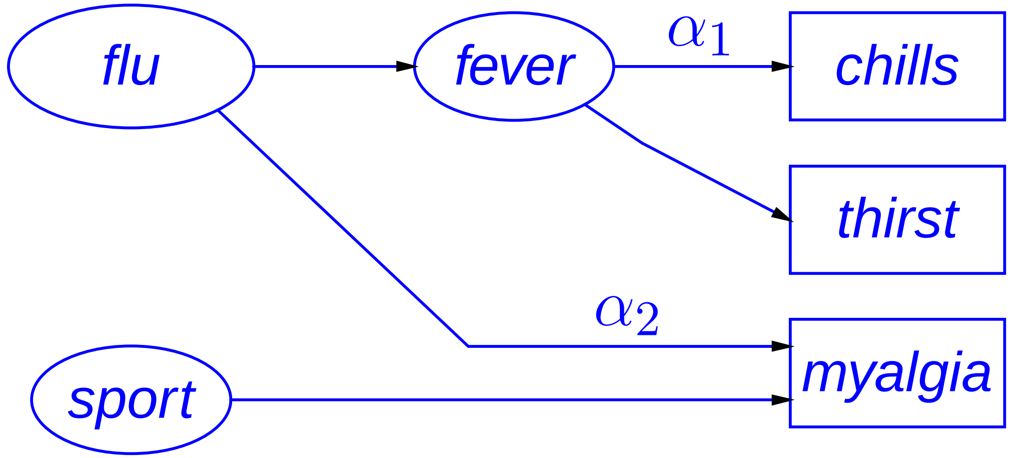


- Causal specification: $\Sigma = (\Delta, \Phi, \mathcal{R})$
- Observed facts: $F = \{\textit{myalgia}, \textit{thirst}\}$
- Facts that should *not* be explained: $C = \{\neg \textit{chills}\}$
- Formal: $D \subseteq \Delta$ is a *diagnosis* if:

(1) $\mathcal{R} \cup D \models F$ (covering condition)

(2) $\mathcal{R} \cup D \cup C \not\models \perp$ (consistency condition)

Consistency condition



- Causal specification: $\Sigma = (\Delta, \Phi, \mathcal{R})$
- Observed facts: $F = \{\textit{myalgia}, \textit{thirst}\}$
- Facts that should *not* be explained:

$$C = \{\neg \textit{chills}\}$$

$$\mathcal{R} \cup \{\textit{flu}, \alpha_1, \alpha_2\} \cup \{\neg \textit{chills}\} \models \bot$$

$$\Rightarrow D = \{\textit{flu}, \alpha_1, \alpha_2\}$$ *no* diagnosis

Abduction = anticausal reasoning

Abduction:

$$\frac{\text{Effect},\ \text{Cause} \longrightarrow \text{Effect}}{\text{Cause}}$$

Idea: *reversal of the causal relation*

Example:

$\textit{fever} \longrightarrow \textit{thirst}$ results in $\textit{thirst} \longrightarrow \textit{fever}$

Thus:

$$\{\textit{thirst} \longrightarrow \textit{fever}, \textit{thirst}\} \models \textit{fever}$$

Conclusion:

Abduction = deduction with implication reversal

Abduction and deduction

Reversal of the causal relations in $\mathcal{R}$ and addition to $\mathcal{R}$ is called the completion of $\mathcal{R}$

Basic idea:

$$d_1 \rightarrow f$$
$$d_2 \rightarrow f$$

indicates that d_1 and d_2 are possible explanations for f;

$$f \rightarrow (d_1 \vee d_2)$$

makes this explicit

Together:

$$\{d_1 \rightarrow f, d_2 \rightarrow f, f \rightarrow (d_1 \vee d_2)\} \equiv \{f \leftrightarrow (d_1 \vee d_2)\}$$

Abducibles

- Δ: defects, some of them derivable from other defects, some not:
  - abducible: defects d not derivable
  - non-abducible: each defect that can be derived
- Φ: findings, also non-abducible

A are abducibles; N are non-abducibles

Example:

$$d_1 \rightarrow d_2$$
$$d_3 \rightarrow f_1$$
$$d_2 \rightarrow f_2$$

$A = \{d_1, d_3\}$; $N = \{d_2, f_1, f_2\}$

Completion

$$\mathcal{R} = \{\varphi_{1,1} \rightarrow n_1, \ldots, \varphi_{1,n_1} \rightarrow n_1, \\ \vdots \\ \varphi_{m,1} \rightarrow n_m, \ldots, \varphi_{m,n_m} \rightarrow n_m\}$$

- $N = \{n_i \mid 1 \leq i \leq m\}$ is the set of non-abducible literals, and
- each $\varphi_{i,j}$ denotes a conjunction of defect literals, possibly including an assumption literal

Predicate completion of $\mathcal{R}$ with respect to N:

$$\mathrm{COMP}[\mathcal{R}; N] = \mathcal{R} \cup \{n_1 \rightarrow \varphi_{1,1} \vee \cdots \vee \varphi_{1,n_1}, \\ \vdots \\ n_m \rightarrow \varphi_{m,1} \vee \cdots \vee \varphi_{m,n_m}\}$$

Example

Example:

$$\mathcal{R} = \{P \wedge Q \rightarrow V, \quad T \rightarrow V, \quad T \rightarrow U\}$$

with $N = \{V, U\}$ results in

$\mathrm{COMP}[\mathcal{R}; N] = \{V \leftrightarrow ((P \wedge Q) \vee T), U \leftrightarrow T\}$

Let V be observed: $\mathrm{COMP}[\mathcal{R}; N] \cup \{V\} \models ((P \wedge Q) \vee T)$

i.e. two alternative diagnoses: $(P \wedge Q)$ and T

Conclusion: Abduction = deduction in a *completed theory*

Deduction of the solutions

- $\mathcal{P} = (\Sigma, F)$ is an abductive diagnostic problem
- $\mathrm{COMP}[\mathcal{R}; N]$ is the predicate completion of $\mathcal{R}$ with respect to N, the set of non-abducible literals in $\mathcal{P}$

A solution formula S for $\mathcal{P}$ is defined as the most specific formula consisting only of abducible literals, such that

$$\mathrm{COMP}[\mathcal{R}; N] \cup F \cup C \vDash S$$

where C is defined as:

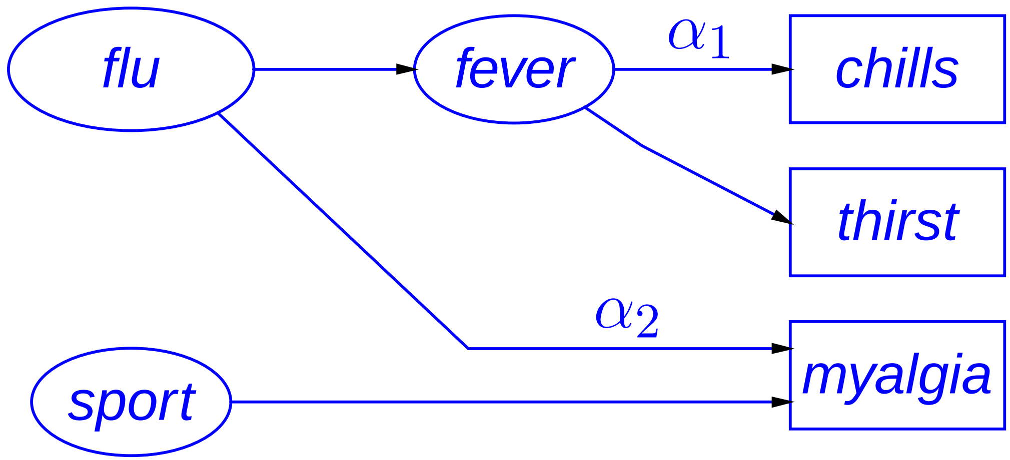
$$C = \{\neg f \in \Phi \mid f \in \Phi, f \notin F,\ f \text{ is a positive literal}\}$$

Solution formula

Theorem. Let $\mathcal{P} = (\Sigma, F)$ be an abductive diagnostic problem. Let C be constraints and S be a solution formula for $\mathcal{P}$. Let $H \subseteq \Delta$ be a set of abducible literals, and let I be an interpretation of $\mathcal{P}$, such that for each abducible literal $a \in A$: $\models_I a$ iff $a \in H$. Then, H is a solution to $\mathcal{P}$ iff $\models_I S$.

Proof.
Conjuncts in S are equivalent to observed findings $f \in F$, that are logically entailed by $\mathcal{R} \cup H$, or to non-observed findings $\neg f \in C$ that are consistent with $\mathcal{R} \cup H$. Hence, an interpretation I for which $\models_I H$, that falsifies each abducible in $\Delta \backslash H$, satisfying every $f \in F$ and each $\neg f \in C$ that has been rewritten, must satisfy this collection of conjuncts, i.e. S.

Example



$\mathcal{R}$:

fever $\wedge\ \alpha_1 \longrightarrow$ *chills*

flu $\longrightarrow$ *fever*

fever $\longrightarrow$ *thirst*

flu $\wedge\ \alpha_2 \longrightarrow$ *myalgia*

sport $\longrightarrow$ *myalgia*

Example

$$\mathrm{COMP}[\mathcal{R}; \{\textit{chills}, \textit{thirst}, \textit{myalgia}, \textit{fever}\}]$$

$$\begin{aligned} = \mathcal{R} \cup \{ & \textit{chills} \longrightarrow \textit{fever} \wedge \alpha_1, \\ & \textit{fever} \longrightarrow \textit{flu}, \textit{thirst} \longrightarrow \textit{fever}, \\ & \textit{myalgia} \longrightarrow (\textit{flu} \wedge \alpha_2) \vee \textit{sport}\} \end{aligned}$$

$$\begin{aligned} = \{ & \textit{chills} \leftrightarrow \textit{fever} \wedge \alpha_1, \\ & \textit{fever} \leftrightarrow \textit{flu}, \\ & \textit{thirst} \leftrightarrow \textit{fever}, \\ & \textit{myalgia} \leftrightarrow (\textit{flu} \wedge \alpha_2) \vee \textit{sport}\} \end{aligned}$$

Note that

$$\mathrm{COMP}[\mathcal{R}; \{\textit{chills}, \textit{thirst}, \textit{myalgia}, \textit{fever}\}] \cup F \cup C \models S \equiv (\textit{flu} \wedge \alpha_2) \vee (\textit{flu} \wedge \textit{sport})$$

given that $F = \{\textit{thirst}, \textit{myalgia}\}$ and $C = \{\neg \textit{chills}\}$

Example

$$\mathrm{COMP}[\mathcal{R}; \{\textit{chills}, \textit{thirst}, \textit{myalgia}, \textit{fever}\}]$$

$$\begin{aligned} = \mathcal{R} \cup \{&\textit{chills} \rightarrow \textit{fever} \wedge \alpha_1, \\ &\textit{fever} \rightarrow \textit{flu}, \textit{thirst} \rightarrow \textit{fever}, \\ &\textit{myalgia} \rightarrow (\textit{flu} \wedge \alpha_2) \vee \textit{sport}\} \end{aligned}$$

$$\begin{aligned} = \{&\textit{chills} \leftrightarrow \textit{fever} \wedge \alpha_1, \\ &\textit{fever} \leftrightarrow \textit{flu}, \\ &\textit{thirst} \leftrightarrow \textit{fever}, \\ &\textit{myalgia} \leftrightarrow (\textit{flu} \wedge \alpha_2) \vee \textit{sport}\} \end{aligned}$$

$$\mathrm{COMP}[\mathcal{R}; \{\textit{chills}, \textit{thirst}, \textit{myalgia}, \textit{fever}\}] \cup F \cup C \vDash \neg(\textit{fever} \wedge \alpha_1)$$

because $\{\neg\textit{chills}, \textit{chills} \leftrightarrow (\textit{fever} \wedge \alpha_1)\} \vDash \neg(\textit{fever} \wedge \alpha_1)$;
$\neg(\textit{fever} \wedge \alpha_1)$ is not part of S, because *fever* is non-abducible

Set covering diagnosis

- $\mathcal{N} = (\Delta, \Phi, C)$ is called a causal net, where:
  - Δ is a set of possible defects,
  - Φ is a set of elements called observable findings, and
  - C is a binary relation

$$C \subseteq \Delta \times \Phi$$

    called the causation relation

- A diagnostic problem in the set-covering theory of diagnosis: $\mathcal{D} = (\mathcal{N}, F)$, where $F \subseteq \Phi$ is a set of observed findings

Further notions

From defects to causes and vice versa:

- effects function $e : \wp(\Delta) \rightarrow \wp(\Phi)$ is defined as follows: for each $D \subseteq \Delta$:

  $$e(D) = \bigcup_{d \in D} e(\{d\})$$

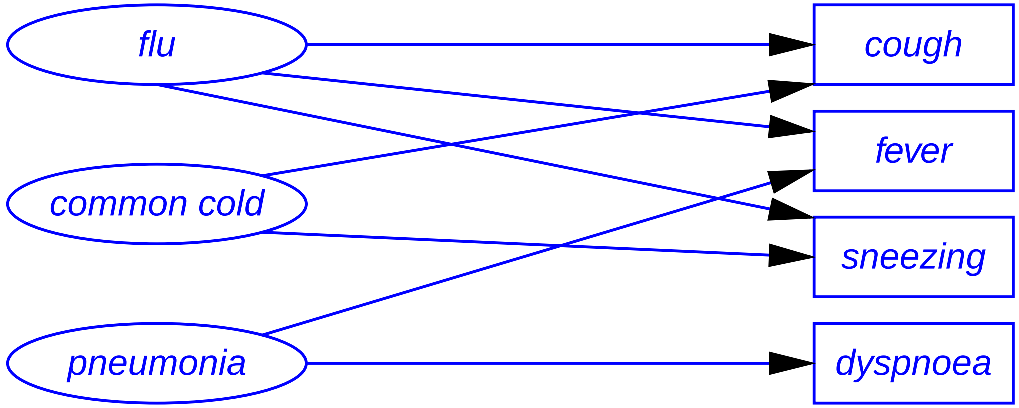
  where $e(\{d\}) = \{f \mid (d, f) \in C\}$

- causes function $c : \wp(\Phi) \rightarrow \wp(\Delta)$ is defined as follows: for each $E \subseteq \Phi$:

  $$c(E) = \bigcup_{f \in E} c(\{f\})$$

  where $c(\{f\}) = \{d \mid (d, f) \in C\}$

Example



$$e(D) = \bigcup_{d \in D} e(\{d\})$$

where

$$e(\{d\}) = \begin{cases} \{\textit{cough}, \textit{fever}, \textit{sneezing}\} & \text{if } d = \textit{flu} \\ \{\textit{cough}, \textit{sneezing}\} & \text{if } d = \textit{common cold} \\ \{\textit{fever}, \textit{dyspnoea}\} & \text{if } d = \textit{pneumonia} \end{cases}$$

Set-covering diagnosis

Let $\mathcal{D} = (\mathcal{N}, F)$ be a diagnostic problem, where F denotes a set of observed findings. Then, a set-covering diagnosis of $\mathcal{D}$ is a set of defects $D \subseteq \Delta$, such that:

$$e(D) \supseteq F$$

Let $F = \{\textit{cough}, \textit{fever}\}$ then

$$D_1 = \{\textit{flu}\}$$

is a diagnosis, but

$D_2 = \{\textit{flu}, \textit{common cold}\}$
$D_3 = \{\textit{common cold}, \textit{pneumonia}\}$

and $D_4 = \{\textit{flu}, \textit{common cold}, \textit{pneumonia}\}$ are also diagnoses for F

Mapping to abductive diagnosis

Define

$$e(\{d\}) = \{f_1, \ldots, f_n\}$$

and construct for each $d \in \Delta$:

$$\begin{array}{rcl} d \wedge \alpha_{f_1} & \rightarrow & f_1 \\ d \wedge \alpha_{f_2} & \rightarrow & f_2 \\ & \vdots & \\ d \wedge \alpha_{f_n} & \rightarrow & f_n \end{array}$$

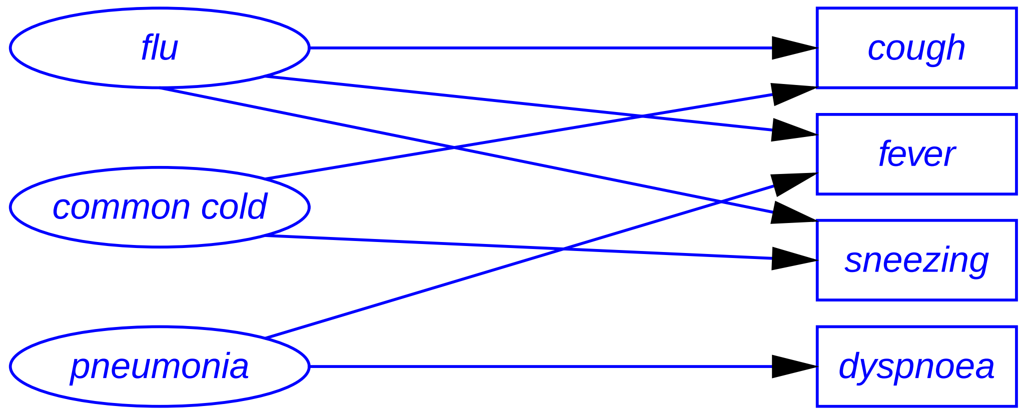
Note no interactions between defects!

$e(D) \supseteq F \Leftrightarrow \mathcal{R} \cup H \vDash F$ and $\mathcal{R} \cup H \nvDash \perp$, with D defects in H

Alternative diagnostic definitions

- Minimal cardinality: a diagnosis D of F is an explanation of $\mathcal{D}$ iff it contains the minimum number of elements among all diagnoses of F
- Irredundancy: a diagnosis D of F is an explanation of $\mathcal{D}$ iff no proper subset of D is a diagnosis of F
- Relevance: a diagnosis D of F is an explanation of $\mathcal{D}$ iff $D \subseteq c(F)$
- Most probable diagnosis: a diagnosis D of F is an explanation of $\mathcal{D}$ iff $P(D|F) \geq P(D'|F)$ for any diagnosis D' of F
- A diagnosis D is called a minimal-cost explanation of $\mathcal{D}$ iff $\sum_{d \in D} \textit{cost}(d) \leq \sum_{d \in D'} \textit{cost}(d)$

Example



$D_1 = \{$*flu*$\}$

$D_2 = \{$*flu*, *common cold*$\}$

$D_3 = \{$*common cold*, *pneumonia*$\}$

$D_4 = \{$*flu*, *common cold*, *pneumonia*$\}$

Observed $F = \{$*cough*, *fever*$\}$

- Diagnoses D_i, $i = 1, \ldots 4$, are relevant diagnoses, because $c(\{$*cough*, *fever*$\}) \supseteq D_i$
- Irredundant diagnoses of F are D_1 and D_3
- There is only one minimal cardinality diagnosis: D_1

Software: AILog

Diagnostic problem in AILog (developed by David Poole):

- facts, denoted by FACTS
- a set of hypotheses, denoted by HYP, and
- a set of constraints, denoted by C

FACTS and constraints C are formulae in first-order logic; hypotheses act as abducibles $=$ assumables in AILog

A set $\mathrm{FACTS} \cup H$ is called an explanation of a closed formula g, where H is a set of ground instances of hypothesis elements in HYP, iff:

(1) $\mathrm{FACTS} \cup H \vDash g$, and

(2) $\mathrm{FACTS} \cup H \cup C \nvDash \bot$.

Example

```
assumable a1.
assumable a2.
assumable fever.
assumable flu.
assumable sport.

chills   <- fever & a1.
fever    <- flu.
thirst   <- fever.
myalgia <- flu & a2.
myalgia <- sport.

false <- chills.  % constraint
```

Calling AILog

```
ailog: create_nogoods.  % enforce consistency
ailog: ask thirst & myalgia.
```

Yields the following results:

```
Answer: thirst & myalgia.
Assuming: [a2, fever, flu].
  [more,ok,how,help]: more.
Answer: thirst & myalgia.
Assuming: [fever, sport].
  [more,ok,how,help]: more.
Answer: thirst & myalgia.
Assuming: [a2, flu].
  [more,ok,how,help]: more.
Answer: thirst & myalgia.
Assuming: [flu, sport].
  [more,ok,how,help]: more.
No more answers.
```

Conclusions

- Model-based (consistency-based and abductive) reasoning is suitable if there are domain models available
- Consistency-gebased diagnosis: no or only scarce knowledge of problems in domain
- Abductive diagnosis: causal models of abnormal behaviour available
- Integration of consistency-based and abductive diagnosis is possible
- Relationship with Bayesian networks