Chapter 5 Bayesian Networks and Influence Diagrams

"The formulation of the problem is often more essential than its solution, which may be merely a matter of mathematical or experimental skill."

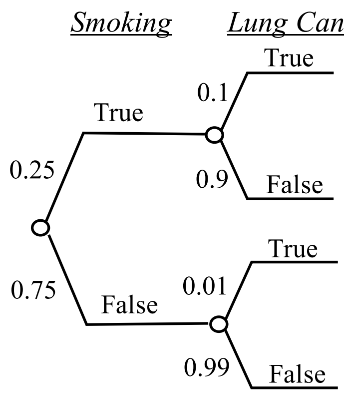
Albert Einstein

5.1 Bayesian Networks

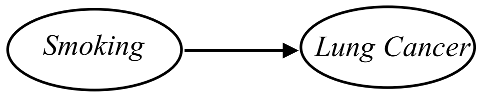
5.1.1 Probabilistic Modeling Using Bayesian Networks

Modeling Relationship between Two Variables

- Suppose you would like to model the relationship between "Smoking" and "Lung Cancer".
- We know from our knowledge that smoking may cause lung cancer, but the relation is probabilistic.
- In Chapter 2, we modeled this relation with a probability tree:



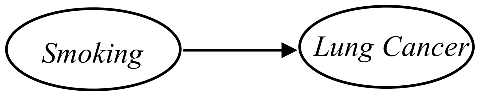
- Another way to represent the relation between "Smoking" and "Lung Cancer" is to use a directed graph or network as follows:



- The above network shows only the "causal" or relevance relation between the two events. The numerical probabilistic information is stored in each node as follows:
- At the node "smoking", we store the probability distribution:
  - p(Smoking = true | ξ) = 0.25
  - p(Smoking = false | ξ) = 0.75
- The probability of having lung cancer is dependent on whether smoking is true or false. Hence at the node "Lung cancer" we store the conditional probability table:
  - p(Lung cancer = true | Smoking = true, ξ) = 0.10
  - p(Lung cancer = false | Smoking = true, ξ) = 0.90

  - p(Lung cancer = true | Smoking = false, ξ) = 0.01
  - p(Lung cancer = false | Smoking = false, ξ) = 0.99

- The complete network is as follows:



Smoking	$p(S \mid \xi)$
True	0.25
False	0.75

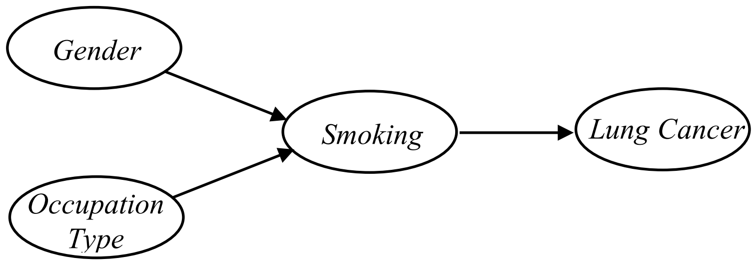
	$p(L \mid S, \xi)$	
	Lung Cancer	
Smoking	True	False
True	0.10	0.90
False	0.01	0.99

Expanding the Network

- In the above example, we have assumed that the prevalence rate of smokers is 25% and that of non-smokers is 75%.

- Suppose that we do not know these figures, but we can indirectly determine them from the characteristics of the population under study.

- Assuming that the two factors that are relevant to smoking are "Gender" and "Occupation Type", we model the information as follows:



- We assume that the possible states or outcomes of the two new variables are:
  - Gender: male or female
  - Occupation type: office or non-office

- We now store the following conditional probabilities at the node "Smoking":

Gender	**Occupation type**	**p(Smoking | Gender, Occupation, ξ)**	
		True	False
Male	Office	0.10	0.90
Male	Non Office	0.20	0.80
Female	Office	0.05	0.95
Female	Non Office	0.10	0.90

- Interpretation of the CPT:

  - p(Smoking = true | Gender = male, Occupation type = office, ξ) = 0.10
  - p(Smoking = false | Gender = male, Occupation type = office, ξ) = 0.90

  - p(Smoking = true | Gender = female, Occupation type = office, ξ) = 0.05
  - p(Smoking = false | Gender = female, Occupation = office, ξ) = 0.95

  - p(Smoking = true | Gender = male, Occupation type = non-office, ξ) = 0.20
  - p(Smoking = false | Gender = male, Occupation type = non-office, ξ) = 0.80

  - p(Smoking = true | Gender = female, Occupation type = non-office, ξ) = 0.10
  - p(Smoking = false | Gender = female, Occupation type = non-office, ξ) = 0.90

- At the node "Gender", we store the following information:

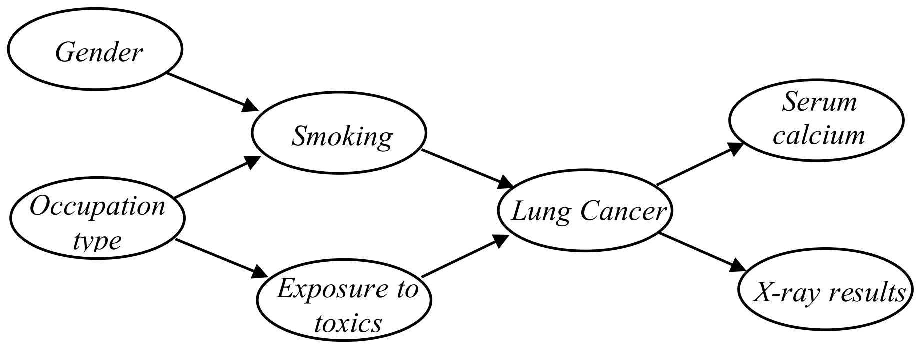
  - p(Gender = male | ξ) = 0.5
  - p(Gender = female | ξ) = 0.5

- At the node "Occupation type", we store the following information:

  - p(Occupation type = office | ξ) = 0.25
  - p(Occupation type = non-office | ξ) = 0.75

Completing the network with more variables:

- Another possible cause of lung cancer is "Exposure to toxics" which is also dependent on the person's "Occupation Type". Also, a person with lung cancer will have an abnormal level of "Serum calcium" and "X-ray results".



- The following conditional probability tables are now stored in each node:

  - p(Gender | ξ)
  - p(Occuptation_type | ξ)
  - p(Smoking | Gender, Occupation_type, ξ)
  - p(Exposure_to_toxics | Occupation_type, ξ)
  - p(Lung_Cancer | Smoking, Exposure_to_toxics, ξ)
  - p(Serum_calcium | Lung_Cancer, ξ)
  - p(X-ray_results | Lung_Cancer, ξ)

- The above network representation of probabilistic relations between uncertain variables is called a **Bayesian Network**.

Definition:

- A **Bayesian Network** is a **Directed Acyclic Graph**; the nodes in the graph indicate the variables of concern, while the arcs between the nodes indicate **possible** probabilistic relations among the nodes.

  In each node, we store a conditional probability distribution of the variable represented by that node, conditioned on the outcomes of all the uncertain variables that are parents of that node.

Qualitative and Quantitative Representation Layers

- A Bayesian network has two layers of representation of knowledge:

  1. At the **Qualitative** level, the graphical structure of the network represents the probabilistic dependence or relevance between the variables. This layer of representation is often referred to as the **graphical structure** of the network

  2. At the **Quantitative** level, the conditional probabilities at each node represent the local "strengths" of the dependence relationships.

Where do all the Numbers (or Probabilities) in a Bayesian Network come from?

- The marginal or conditional probabilities at each node can be obtained in the following manners:

  1. Direct assessment by domain experts (c/o Chapter 7)

  2. Learn from a sufficient amount of data using:
     - Statistical methods
     - Machine learning methods

  3. Outputs from other mathematical models
     - Simulation models (Monte Carlo and Discrete Events)
     - Stochastic models, e.g., Queueing, Markov chains, etc.
     - etc.

  4. Combinations of the above
     - Experts assess the graphical structure and learning algorithms or other models to fill in the numbers.
     - Learn both graphical structure and probabilities and let the experts fine-tune the results.

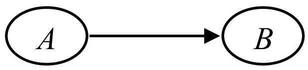
5.1.2 Properties of Bayesian Networks

Presence of an Arc indicates Possible Relevance while Absence of an Arc indicates Definite Non-Relevance

- The **Presence** of an arc between two nodes in a Bayesian network indicates possible relevance or probabilistic dependence between the two nodes. It may turn out later that the two nodes are not relevant.

Example

- If you do not know for sure that *A* and *B* are not relevant or not probabilistically dependent, put an arc between them.

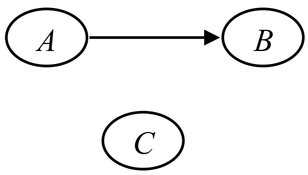


A might be relevant to *B*.

- But if you are sure that *A* and *B* are not relevant or not probabilistically dependent on each other, then omit the arc between them as shown below.

A is definitely not relevant to *B*

Example



A might be relevant to *B*.
C is definitely not relevant to *A*
C is definitely not relevant to *B*

Interpretation of the Arcs: Relevance or Causality?

- Bayesian networks can be used to represent causal relations between variables, and they are often drawn or constructed by experts based on this knowledge.

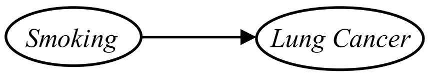
- However, we would like to adapt the interpretation that the arcs represent relevance, probabilistic or statistical correlation instead of only causality.

- This is to allow us to process the network which may include removal of nodes and reversal of arcs.

- Knowledge of causality may be used to construct the network, but once it is completed, we drop that notion and adopt the more general statistical interpretation.

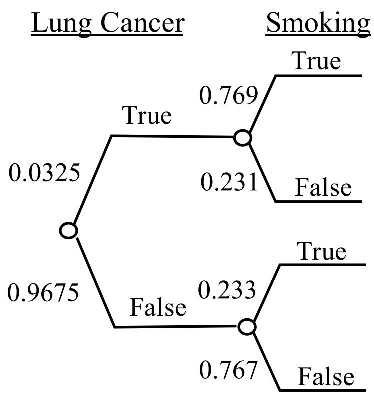
Example

- The following network expresses the risk of lung cancer from smoking.

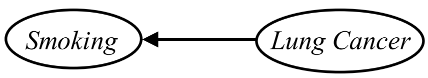


- Suppose we are interested to know if a specific person has lung cancer, what is the probability that he or she smokes?

- In Chapter 2, we did this by flipping the probability tree:



- In Bayesian networks, we may apply the arc-reversal operation and obtain a new network as follows:



- The operation will compute and replace the probabilities at the two nodes as follows:

  At the "Lung cancer" node:
  - p(Lung Cancer = true | ξ) = 0.0325
  - p(Lung Cancer = false | ξ) = 0.9675

  At the "Smoking" node:
  - p(Smoking= true | Lung Cancer = true, ξ) = 0.769
  - p(Smoking= false | Lung Cancer = true, ξ) = 0.231
  - p(Smoking= true | Lung Cancer = false, ξ) = 0.233
  - p(Smoking= false | Lung Cancer = false, ξ) = 0.767

Bayesian Network Represents a Joint Probability Distribution

- A Bayesian network represents the **Joint Probability Distribution** (JPD) of all the variables in the network.

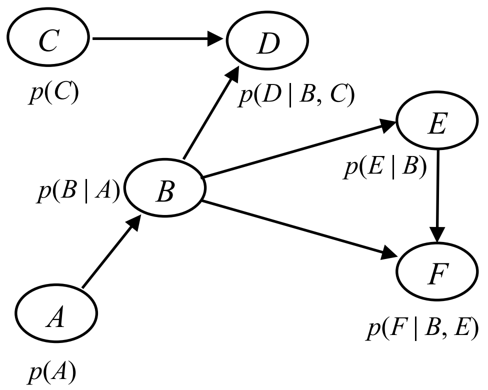
- The JPD of all the variables in the network is equal to the product of all the conditional probabilities stored at each node. That is for an n-node Bayesian network:

$$p(X_1, X_2, ..., X_n) = \prod_{j=1}^{n} p(X_j \mid \pi(X_j))$$

where $p(X_j \mid \pi(X_j))$ is the conditional probability distribution for X_j stored at node j, and $\pi(X_j)$ is the set of parent nodes of X_j in the network.

Example

- Consider a Bayesian network with six variables:



- The JPD for the six variables is equal to the product of all the conditional probabilities stored in each node:

$$p(A, B, C, D, E, F) = p(A)\, p(B \mid A)\, p(C)\, p(D \mid B, C)\, p(E \mid B)\, p(F \mid B, E)$$

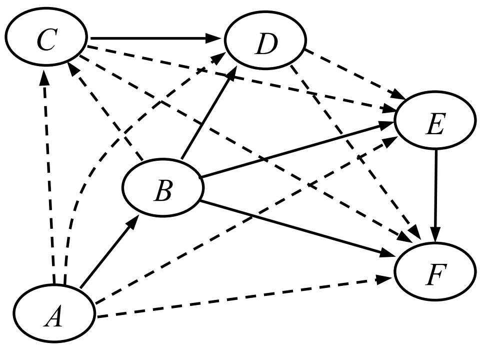
How does a Bayesian Network help in simplifying the JPD?

- A Bayesian network helps us simplify the JPD of a set of non-mutually independent variables.

- For example, without constructing a Bayesian network first, we can still derive the JPD of the six variables, but it would be highly complex based on the standard factorization rule:

$$p(A, B, C, D, E, F) = p(A)\, p(B \mid A)\, p(C \mid A, B)\, p(D \mid A, B, C)\, p(E \mid A, B, C, D)\, p(F \mid A, B, C, D, E)$$

- But the above corresponds to the case when the network is fully connected with 15 arcs, i.e., there is an arc between every pair of nodes.



A Bayesian network should have as little number of arcs as possible.

- We observe that the lesser the number of arcs there are in the network, the simpler will be the JPD.

- Hence the simplicity of the factorization depends on how sparse is the network.

- It is important to always construct the most compact network, i.e., insert only necessary arcs.

5.1.3 Conditional Independence

- Consider three variables A, B, and C. Let $p(A \mid B, C)$ be the conditional probability distribution of A given B and C.

- If $p(A \mid B, C)$ does not depend on the actual value of B, i.e., $p(A \mid B{=}b_1, C) = p(A \mid B{=}b_2, C)$, then we may write $p(A \mid B, C) = p(A \mid C)$. See the example below.

- We say that A is **Conditionally Independent** of B given C. We use the notation $A \perp B \mid C$.

- Note that $A \perp B \mid C \Leftrightarrow B \perp A \mid C$

Example

- Suppose the CPT for $p(A \mid B, C) =$

		a_1	a_2
b_1	c_1	0.2	0.8
b_2	c_1	0.2	0.8
b_1	c_2	0.7	0.3
b_2	c_2	0.7	0.3

- We observe that
  - When $C{=}c_1$, the conditional probability of A given $C{=}c_1$ does not depend on the value of B.
  - When $C{=}c_2$, the conditional probability of A given $C{=}c_2$ does not depend on the value of B.

- We can simply the CPT to $p(A \mid C) =$

	a_1	a_2
c_1	0.2	0.8
c_2	0.7	0.3

5.1.4 Graphical Representation of Conditional Independence in 3-Node BN

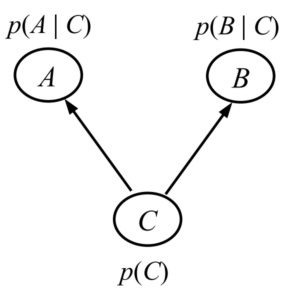
- In general, given any three variables, the joint distribution $p(A, B, C)$ can be factorized as

$$p(A, B, C) = p(A \mid B, C)\, p(B \mid C)\, p(C)$$

- Now, if $A \perp B \mid C$, then using the relation $p(A \mid B, C) = p(A \mid C)$, we have

$$p(A, B, C) = p(A \mid C)\, p(B \mid C)\, p(C)$$

- The BN representing $A \perp B \mid C$ can be drawn as follows:



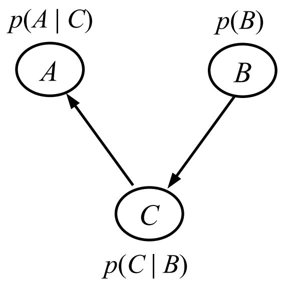
- Another way to factorize the joint distribution is

$$p(A, B, C) = p(A \mid B, C)\, p(C \mid B)\, p(B)$$

- Now, if $A \perp B \mid C$, then again using the relation $p(A \mid B, C) = p(A \mid C)$, we have

$$p(A, B, C) = p(A \mid C)\, p(C \mid B)\, p(B)$$

- A BN also representing $A \perp B \mid C$ is therefore



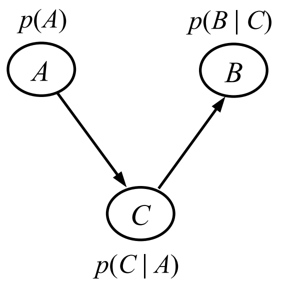
- Yet, another way to factorize the joint distribution is

$$p(A, B, C) = p(B \mid A, C)\, p(C \mid A)\, p(A)$$

- Now, if $A \perp B \mid C$, then $B \perp A \mid C$ and using $p(B \mid A, C) = p(B \mid C)$, we have

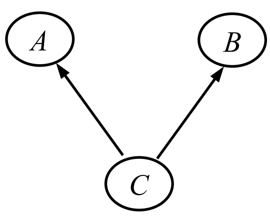
$$p(A, B, C) = p(B \mid C)\, p(C \mid A)\, p(A)$$

- Another BN also representing $A \perp B \mid C$ is therefore

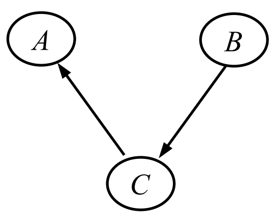


Summary of 3-Node BN Representing Conditional Independence

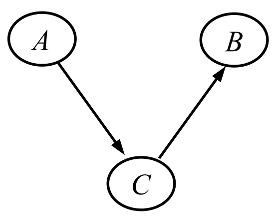
- The following three BNs indicate that A is conditional independent of B given C:



$A \perp B \mid C$



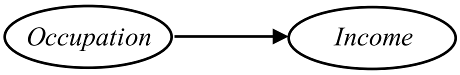
$A \perp B \mid C$



$A \perp B \mid C$

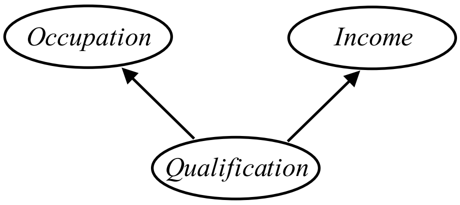
Example

- *Income* is generally dependent on *Occupation*. Hence the BN with only these two variables is as follows:



Income and Occupation are relevant to each other

- Now, if you add a third variable *Qualification* into the BN, and suppose you know the qualification of a person then his/her income and occupation become independent, the BN with the three nodes is:

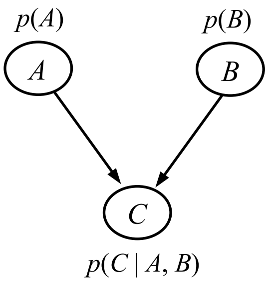


$$Income \perp Occupation \mid Qualification$$

Non-Conditional Independence

- Suppose A and B are generally independent of each other, then the 2-node BN is:

- Suppose C is dependent on both A and B, then the 3-node BN is

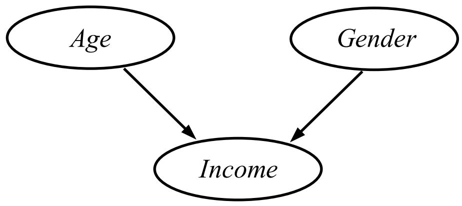


- Under the above conditions, there is no conditional independence among A, B, and C.

Example

- *Age* and *Gender* are independent of each other.

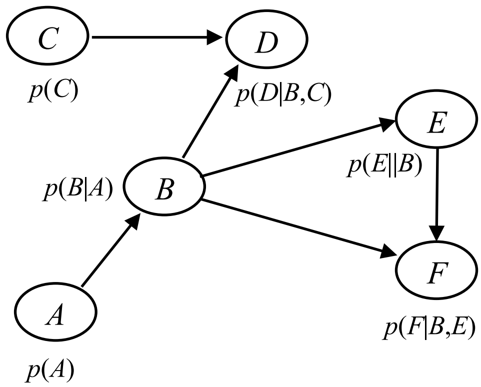
- *Income* depends on both *Age* and *Gender*:



- But there is no conditional independence relation among the three variables.

5.1.5 Identifying Conditional Independence in General Networks

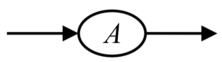
- Can you identify some of the conditional independence relations in the following network?



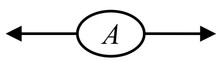
Conditional Independence in Directed Acyclic Graphs (DAG)

- In general, it is possible to identify conditional independence of any two distinct sets of nodes in the network given any third set of nodes using the notion of ***d*-separation**.
- For example, in the above network, we want to check if the relation ($A \perp G \mid E$) is true or false.

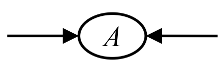
- We will use the following terminologies regarding the pattern of arcs in and out of a node:



The two arcs meet **Head-to-Tail** (H2T) at node *A*.



The two arcs meet **Tail-to-Tail** (T2T) at node *A*.



The two arcs meet **Head-to-Head** (H2H) at node *A*.

Blocking of a Path

- **Definition (Blocking of a Path)**

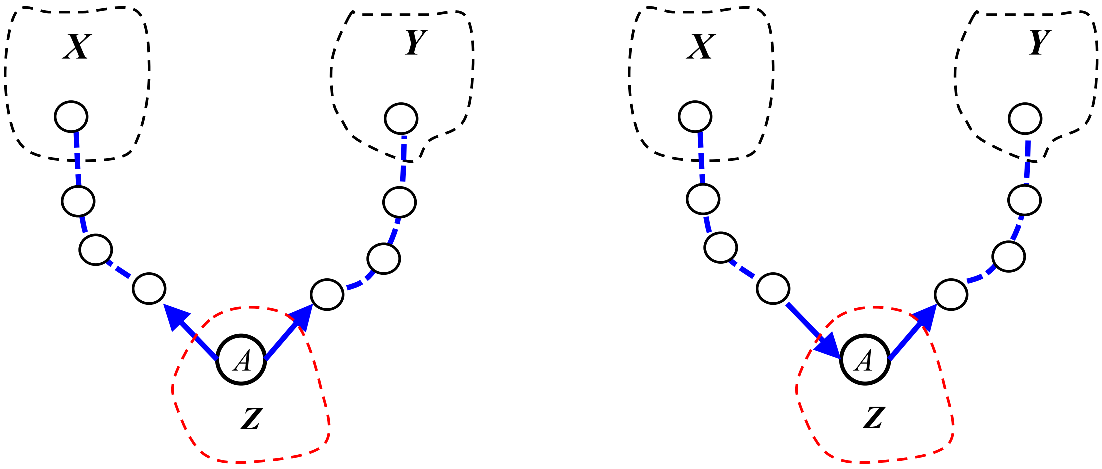
  Let ***G*** be a directed acyclic graph and let ***X***, ***Y*** and ***Z*** be three *disjoint* subsets of nodes in ***G***. Then a path (ignoring the arcs' direction) between a node in ***X*** and a node in ***Y*** is said to be **blocked** by ***Z*** if there exists an intermediate node A on the path such that **at least one** of the following conditions is true:

  1. A is either a head-to-tail **or** a tail-to-tail node in the path, **and** A is in ***Z.***

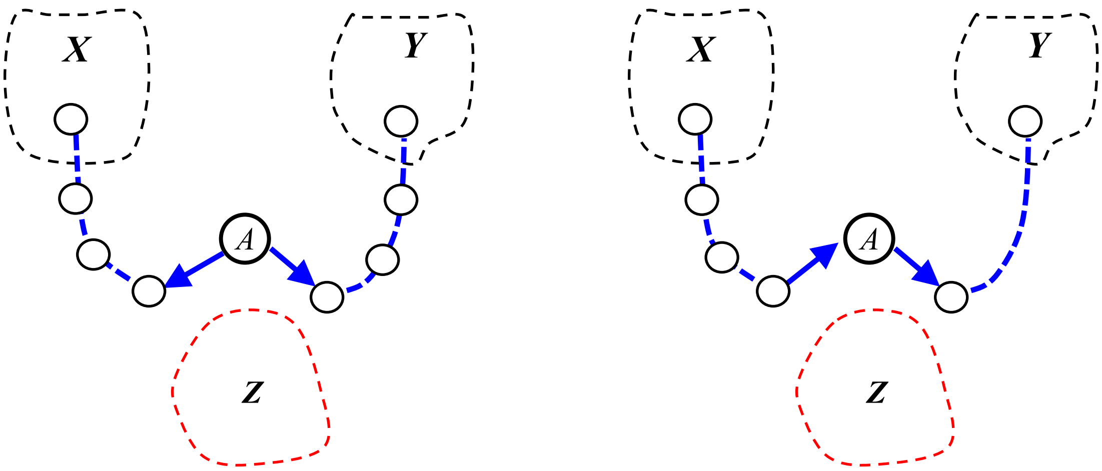
  2. A is a head-to-head node in the path, **and** neither A nor any of its descendants are in ***Z.***

Illustrations on Condition (1)

- **Condition (1) is True:** A is (T2T or H2T) and $A \in \boldsymbol{Z}$

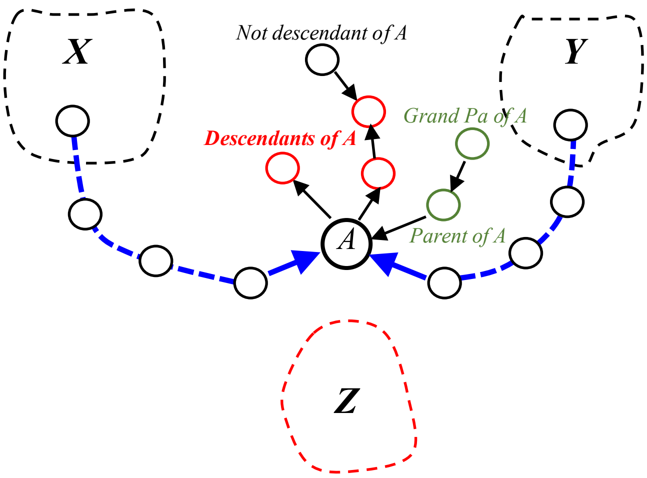


- **Condition (1) is False:** A is (T2T or H2T) but $A \notin \boldsymbol{Z}$

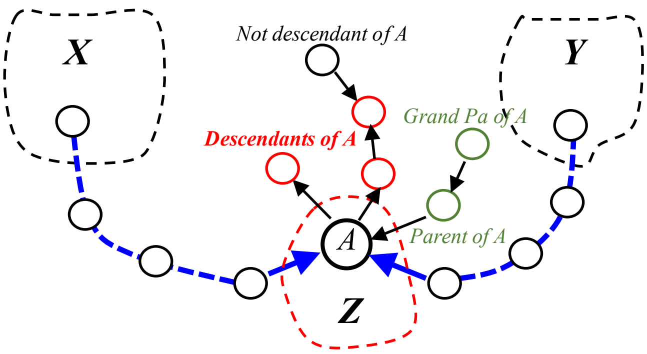


Illustrations on Condition (2)

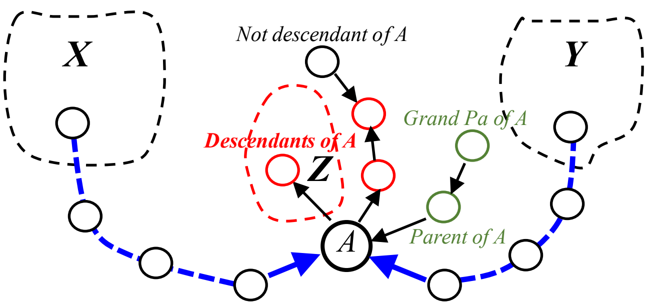
- **Condition (2) is True:** A is H2H, $A \notin \boldsymbol{Z}$ **and** None of A's descendants $\in \boldsymbol{Z}$.



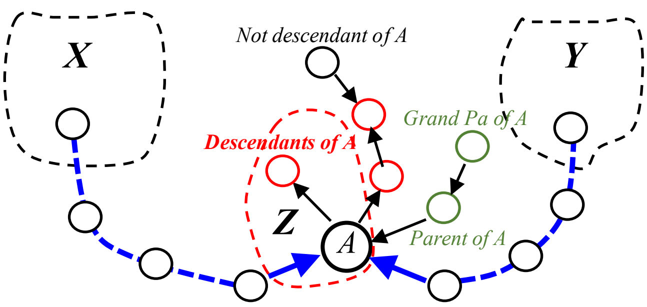
- **Condition (2) is False:** A is H2H but $A \in \boldsymbol{Z}$



- **Condition (2) is False:** A is H2H, $A \notin \boldsymbol{Z}$ but at least one of A*'s descendants* $\in \boldsymbol{Z}$

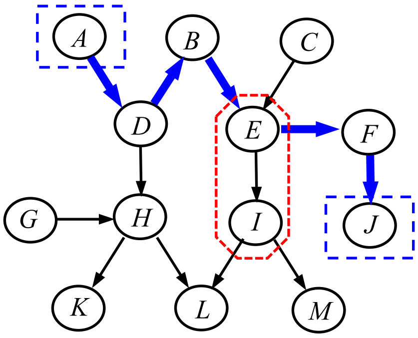


- **Condition (2) is False:** A is H2H, but $A \in \boldsymbol{Z}$ and at least one of its descendants $\in \boldsymbol{Z}$



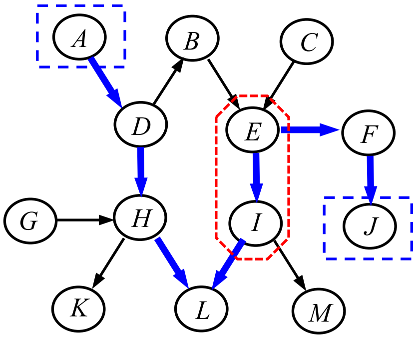
Examples on Blocking and Non-Blocking

Example (blocking)



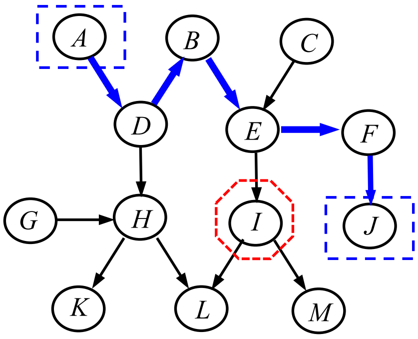
- Let $\boldsymbol{X} = \{A\}$, $\boldsymbol{Y} = \{J\}$ and $\boldsymbol{Z} = \{E, I\}$.
- The path A-D-B-E-F-J is blocked by $\boldsymbol{Z} = \{E, I\}$ because there exists node E that is H2T along the path and $E \in \boldsymbol{Z}$ (Condition 1).

Example (blocking)



- Let $\boldsymbol{X} = \{A\}$, $\boldsymbol{Y} = \{J\}$ and $\boldsymbol{Z} = \{E, I\}$.
- The path A-D-H-L-I-E-F-J is blocked by $\boldsymbol{Z} = \{E, I\}$ because there exists node L that is H2H along the path and $L \notin \boldsymbol{Z}$ (Condition 2)
- Note that we could also use node E or node I as the candidate blocking node:
  - node E is T2T along the path and $E \in \boldsymbol{Z}$ (Condition 1).
  - node I is H2T along the path and $E \in \boldsymbol{Z}$ (Condition 1).

Example (None Blocking)

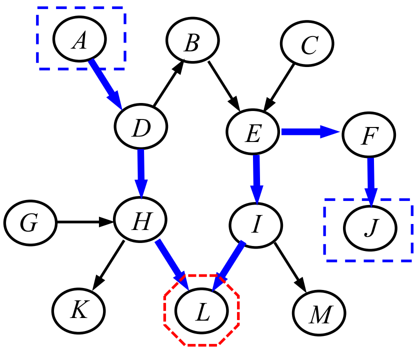


- Let $\boldsymbol{X} = \{A\}$, $\boldsymbol{Y} = \{J\}$ and $\boldsymbol{Z} = \{ I \}$.

- The path A-D-B-E-F-J is not blocked by $\boldsymbol{Z} = \{ I \}$ because none of the nodes along this path satisfy either Condition 1 or Condition 2:

  - D is H2T along the path, but $D \notin \boldsymbol{Z}$.
  - B is H2T along the path, but $B \notin \boldsymbol{Z}$.
  - E is H2T along the path, but $E \notin \boldsymbol{Z}$.
  - F is H2T along the path, but $F \notin \boldsymbol{Z}$.

Example (None Blocking)

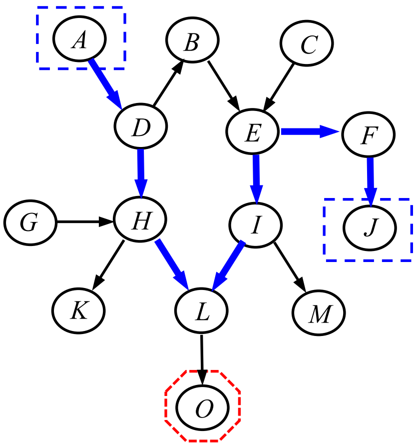


- Let $\boldsymbol{X} = \{A\}$, $\boldsymbol{Y} = \{J\}$ and $\boldsymbol{Z} = \{ L \}$.

- The path A-D-H-L-I-E-F-J is **not blocked** by $\boldsymbol{Z} = \{ L \}$ because none of the nodes along this path satisfy either Condition 1 or Condition 2:

  - D is H2T along the path, but $D \notin \boldsymbol{Z}$.
  - H is H2T along the path, but $H \notin \boldsymbol{Z}$.
  - L is H2H along the path and $L \in \boldsymbol{Z}$
  - I is H2T along the path, but $I \notin \boldsymbol{Z}$.
  - E is T2T along the path, but $E \notin \boldsymbol{Z}$.
  - F is H2T along the path, but $F \notin \boldsymbol{Z}$.

Example (None Blocking)



- Let $\boldsymbol{X} = \{A\}$, $\boldsymbol{Y} = \{J\}$ and $\boldsymbol{Z} = \{\ O\ \}$.

- The path *A-D-H-L-I-E-F-J* is **not blocked** by $\boldsymbol{Z} = \{\ O\ \}$ because none of the nodes along this path satisfy either Condition 1 or Condition 2:

  - D is H2T along the path, but $D \notin \boldsymbol{Z}$.
  - H is H2T along the path, but $H \notin \boldsymbol{Z}$.
  - **$\boldsymbol{L}$ is H2H along the path but $\boldsymbol{L}$ has a descendant $\boldsymbol{O} \in \mathbf{Z}$.**
  - I is H2T along the path, but $I \notin \boldsymbol{Z}$.
  - E is T2T along the path, but $E \notin \boldsymbol{Z}$.
  - F is H2T along the path, but $F \notin \boldsymbol{Z}$.

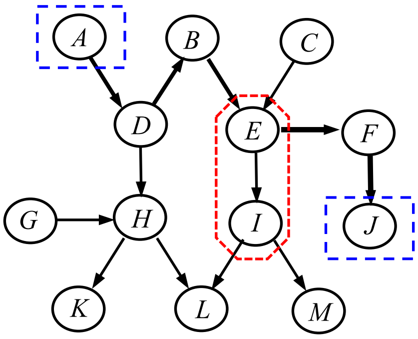
d-Separation

- **Definition (*d*-Separation)**
  Let ***G*** be a directed acyclic graph and let ***X***, ***Y*** and ***Z*** be three disjoint subsets of nodes in ***G***. Then ***X*** and ***Y*** are said to be ***d*-separated** by ***Z*** if every path from a node ***X*** to a node in ***Y*** is blocked by ***Z***.

- If ***X*** and ***Y*** are **not *d*-separated** by ***Z*** then we say that ***X*** and ***Y*** are ***d*-connected** by ***Z*.**

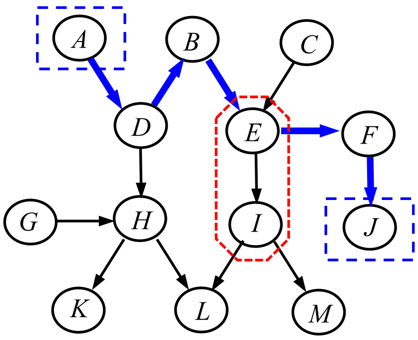
Example (*d*-Separation)

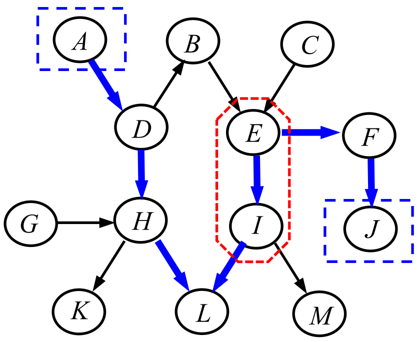
- Are the nodes in$\{ A \}$ and $\{ J \}$ *d*-separated by $\{ E, I \}$ in the BN below?



- Let $\boldsymbol{X} = \{ A \}$, $\boldsymbol{Y} = \{ J \}$ and $\boldsymbol{Z} = \{ E, I \}$.

- There are two paths from ***X*** to ***Y*** as shown below:



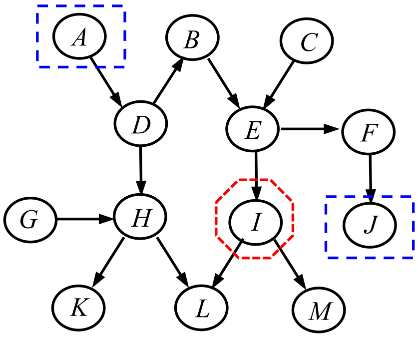


- We have previously shown that both of these paths are blocked by ***Z***.

- Since all paths from ***X*** to ***Y*** are blocked by ***Z***, it follows that ***X*** and ***Y*** are *d*-separated by ***Z***.

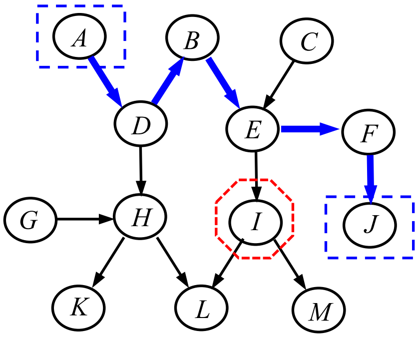
Example (None *d*-Separation)

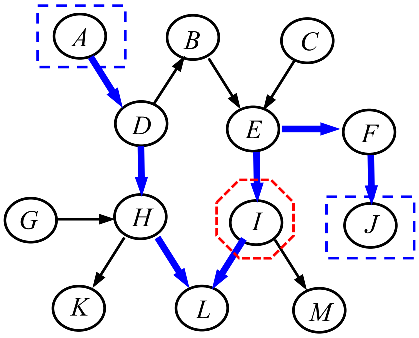
- Are the nodes in { A } and { J } d-separated by { I } in the BN below?



- Let $\boldsymbol{X}$ = { A }, $\boldsymbol{Y}$ = { J } and $\boldsymbol{Z}$ = { I }.

- There are two paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$ as shown below:





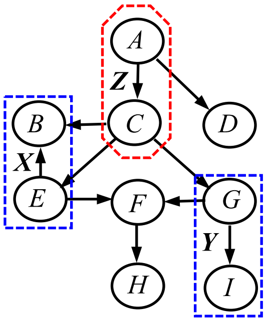
- We have previously shown that the first path A-D-B-E-F-G is not blocked by $\boldsymbol{Z}$.

- Since at least one of the paths is not blocked, it follows that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are not d-separated by $\boldsymbol{Z}$.

- Note that it is not necessary to show whether the second path A-D-H-L-I-E-F-J is blocked or not since we already have already shown that the first path is not blocked.

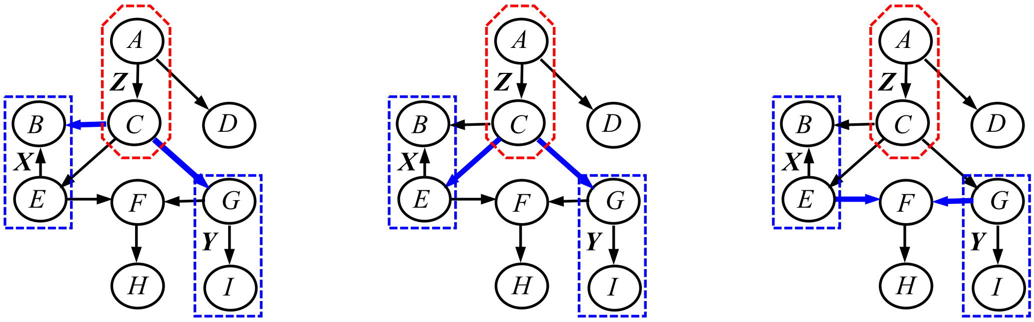
Example (*d*-Separation)

- Are the nodes in { *B, E* } and { *G, I* } *d*-separated by { *A, C* } in the BN below?



- Let $\boldsymbol{X}$ = { *B*, *E* }, $\boldsymbol{Y}$ = { *G, I* }, and $\boldsymbol{Z}$ = { *A, C* }.

- There are three paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$ as shown below:

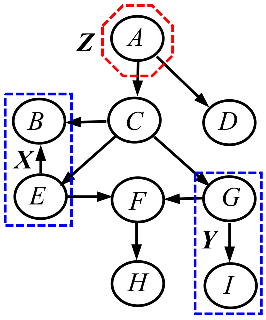


1. Path *B*-*C*-G is **blocked** since there is a node *C* that is T2T, and $C \in \boldsymbol{Z}$.
2. Path *E*-*C*-*G* is **blocked** since there is a node *C* that is T2T, and $C \in \boldsymbol{Z}$.
3. Path *E*-*F*-*G* is **blocked** since there is a node *F* that is H2H, $F \notin \boldsymbol{Z}$ and its descendant $H \notin \boldsymbol{Z}$.

- Since all paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$ are blocked by $\boldsymbol{Z}$, it follows that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are *d*-separated by $\boldsymbol{Z}$.

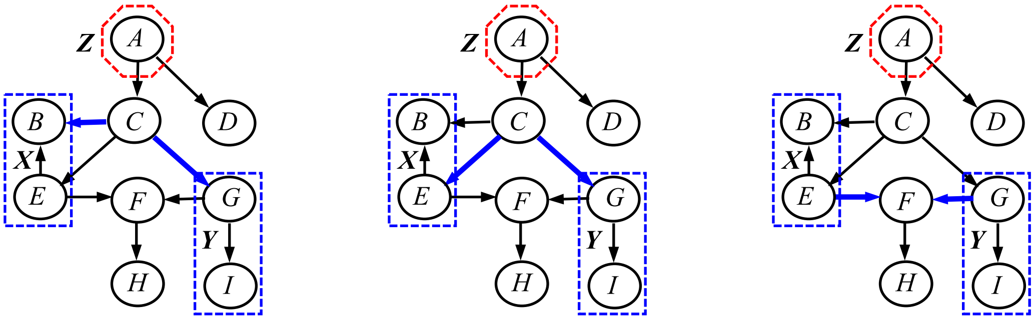
Example (None *d*-Separation)

- Are the nodes in{ B, E } and { G, I } *d*-separated by { A } in the BN below?



- Let $\boldsymbol{X}$ = { B, E }, $\boldsymbol{Y}$ = { G, I }, and $\boldsymbol{Z}$ = { A }.

- There are three paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$ as shown below:



1. Path B-C-G is **not blocked** as the only node C along the path is T2T but $C \notin \boldsymbol{Z}$.

2. Path E-C-G is **not blocked** as the only node C along the path is T2T but $C \notin \boldsymbol{Z}$.

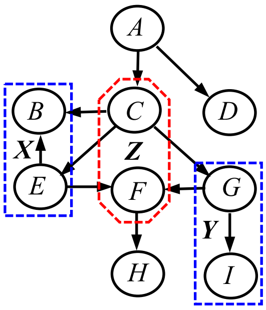
3. Path E-F-G is **blocked** as there exits node F which is H2H and its descendant $H \notin \boldsymbol{Z}$.

- Since at least one of the paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$ is not blocked by $\boldsymbol{Z}$, it follows that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are not *d*-separated by $\boldsymbol{Z}$.

- Note that was is not necessary to check the second and third paths after showing that the first path is not blocked.

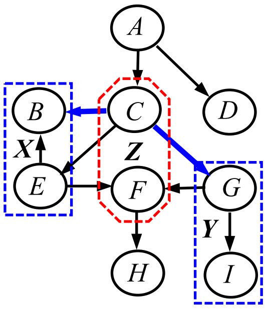
Example (None *d*-Separation)

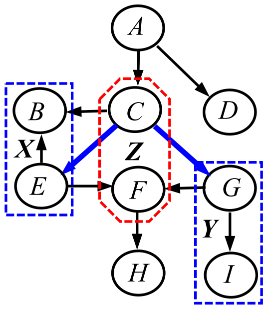
- Are the nodes in { *B, E* } and { *G, I* } *d*-separated by { *C, F* } in the BN below?

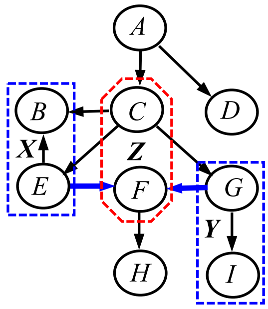


- Let $\boldsymbol{X} = \{B, E\}$, $\boldsymbol{Y} = \{G, I\}$, and $\boldsymbol{Z} = \{C, F\}$

- There are three paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$ as shown below:







1. Path *B*-*C*-*G* is **blocked** since there exists node *C* along the path that is T2T and $C \in \boldsymbol{Z}$.

2. Path *E*-*C*-*G* is **blocked** since there exists node *C* along the path that is T2T and $C \in \boldsymbol{Z}$.

3. Path *E*-*F*-*G* is **not blocked** since the only node *F* along the path is H2H but $E \in \boldsymbol{Z}$.

- Since at least one of the paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$ is not blocked by $\boldsymbol{Z}$, it follows that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are **not *d*-separated** by $\boldsymbol{Z}$.

- Note that if we had first shown that Path 3 is not blocked, we can straight away conclude none *d*-separation.

***d*-Separation ⇔ Conditional Independence**

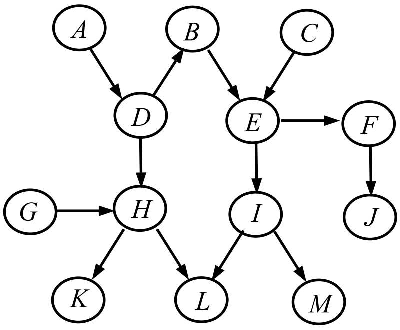
- **Theorem (Verma and Pearl, 1988)**
  Let ***X***, ***Y***, and ***Z*** be 3 disjoint subsets of nodes in a directed acyclic graph ***G***. *If* ***X*** and ***Y*** are *d-separated* by ***Z*** in ***G***, then ***X*** and ***Y*** are *Conditionally Independent* given ***Z*** in all probability distributions ***G*** represents.

- The converse of the above theorem is also true (Geiger and Pearl, 1988).

- Hence, *d*-separation in the Graphical Structure of the BN is necessary and sufficient for conditional probabilistic independence in the BN.

- **We can identify Conditional Independence relations in the probabilities among any three disjoint subsets of variables in a BN by just checking for *d*-separation in the graphical structure of the DAG.**

Example

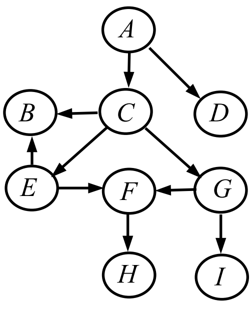


- For all (conditional) probability distributions that the above Bayesian network represents and by the *d*-separation relations shown earlier:

  1. The variables in { *A* } and { *J* } are **Conditional Independent** given the variables in { *E, I* }.

  2. The variables in { *A* } and { *J* } are **Not Conditional Independent** given the variables in { *I* }.

Example



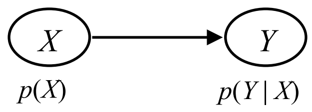
- For all (conditional) probability distributions that the above Bayesian network represents and by the *d*-separation relations shown earlier:

  1. The variables in $\{B, E\}$ and $\{G, I\}$ are **Conditional Independent** given variables in $\{A, C\}$.
  2. The variables in $\{B, E\}$ and $\{G, I\}$ are **Not Conditional Independent** given the variables in $\{A\}$.
  3. The variables in $\{B, E\}$ and $\{G, I\}$ are **Not Conditional Independent** given the variables in $\{C, F\}$.

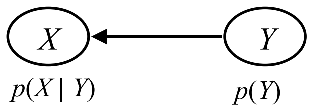
5.1.6 Arc Reversal Operations

Arc Reversal on a 2-Node Network

- Given a two-node network with variables X and Y:



- The arc between X and Y may be reversed, i.e., replaced by an arc from Y to X. This results in the following network:



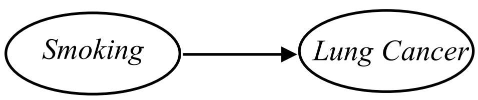
- The probability distributions $p(Y)$ and $p(X \mid Y)$ for the new network can be computed using Bayes' Theorem as follows:

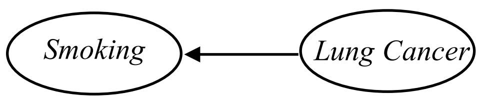
$$p(Y) = \sum_X p(X)\, p(Y|X)$$

$$p(X|Y) = \frac{p(X)\, p(Y|X)}{p(Y)}$$

- Note that the probabilities on the LHS are those required by the new network (after reversal) while those probabilities on the RHS are from the original network (before reversal).

Example (Lung Cancer and Smoking)





Smoking	$p(S \mid \xi)$
True	0.25
False	0.75

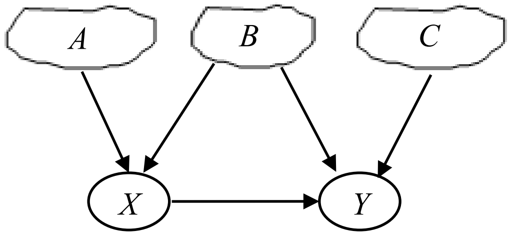
	$p(L \mid S, \xi)$	
	Lung Cancer	
Smoking	True	False
True	0.10	0.90
False	0.01	0.99

	$p(S \mid L, \xi)$	
	Smoking	
Lung Cancer	True	False
True	0.769	0.231
False	0.233	0.767

Lung Cancer	$p(S \mid \xi)$
True	0.0325
False	0.9675

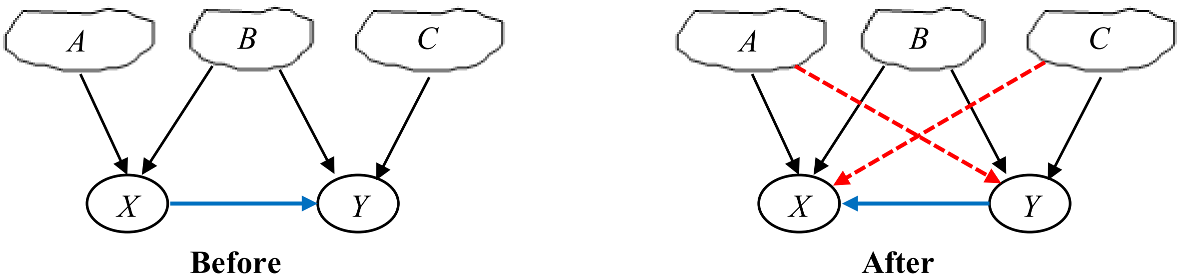
Arc Reversal in General Bayesian Networks

- We can generalize the arc reversal operation to any arc in a general Bayesian Network.
- Any arc in a general BN may be reversed in direction if no cycle is created after the reversal.
- Let X and Y be two directly connected nodes. The parent nodes of X and Y may be partitioned into 3 disjoint sets:



  - Set A contains nodes which are parents of X but not of Y
  - Set B contains nodes which are both parents of X and Y
  - Set C contains nodes which are parents of Y but not of X

- Suppose that the arc from X to Y can be reserved, i.e., it will not create a cycle after that.
- After the reversal operation, both nodes X and Y inherit each other's parents. That is we add arcs from all the nodes in set A to Y, and all the nodes in set C to X:



- The updated Conditional Probabilities for X and Y in the new network are computed as follows:

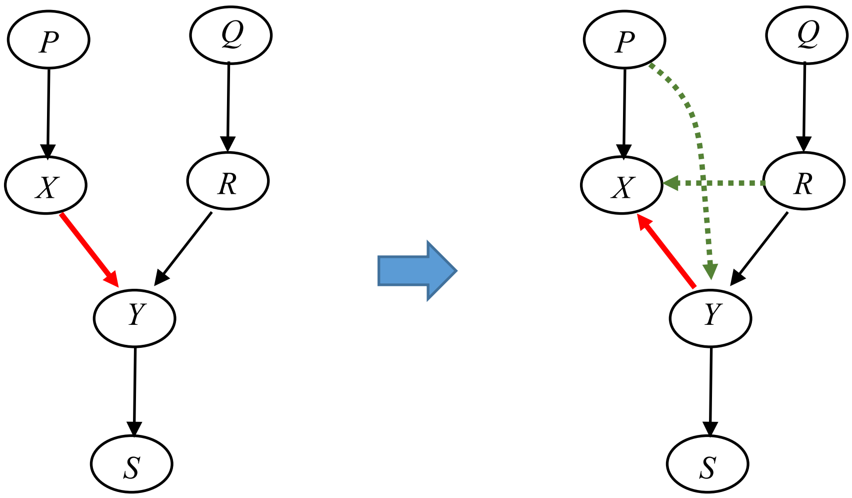
$$p(Y \mid A,B,C) = \sum_{X} p(X \mid A,B)\, p(Y \mid X,B,C)$$

$$p(X \mid Y,A,B,C) = \frac{p(X \mid A,B)\, p(Y \mid X,B,C)}{p(Y \mid A,B,C)}$$

- Notice that all the probabilities on the RHS are from the old network.

Example

- Reverse the arc from X to Y in the following BN:



<u>Before Reversal</u>

$p(Y \mid X, R)$

$p(X \mid P)$

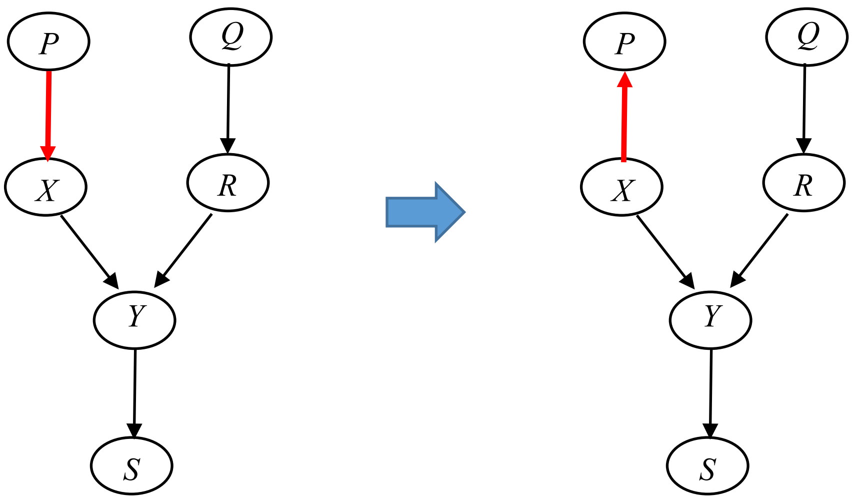
<u>After Reversal</u>

$p(Y \mid P, R)$

$p(X \mid P, Y, R)$

Example

- Reverse the arc from P to X:



<u>Before Reversal</u>

$p(P)$

$p(X \mid P)$

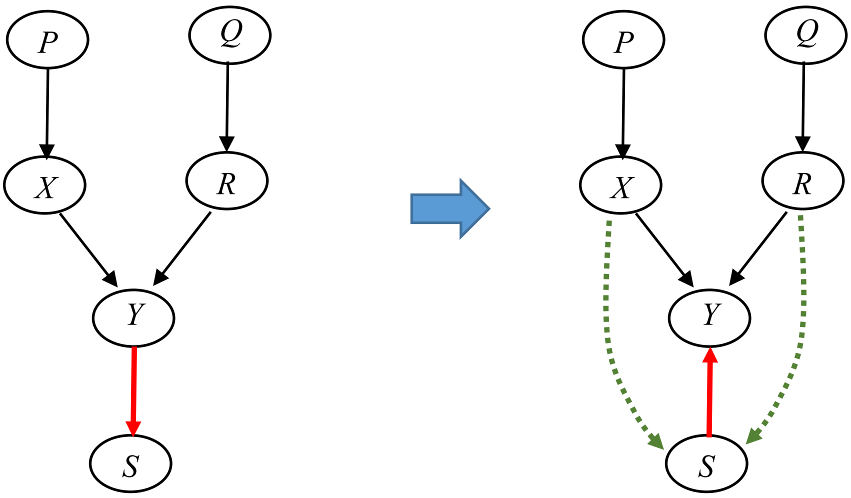
<u>After Reversal</u>

$p(P \mid X)$

$p(X)$

Example

- Reverse the arc from Y to S:



Before Reversal
$p(Y \mid X, R)$
$p(S \mid Y)$

After Reversal
$p(Y \mid S, X, R)$
$p(S \mid X, R)$

Why do we have to add arcs after the reversal operations?

- Reversing an arc between X and Y is equivalent to changing the order of conditioning between X and Y. But X and Y are also dependent on their parents.

- To compute the new conditional probabilities for X and Y, we need to apply Bayes' Theorem.

- It can be shown that the Bayes formula can only compute the values of $p(Y \mid A, B, C)$, and $p(X \mid Y, A, B, C)$, and nothing simpler using the probability distributions available from the original network.

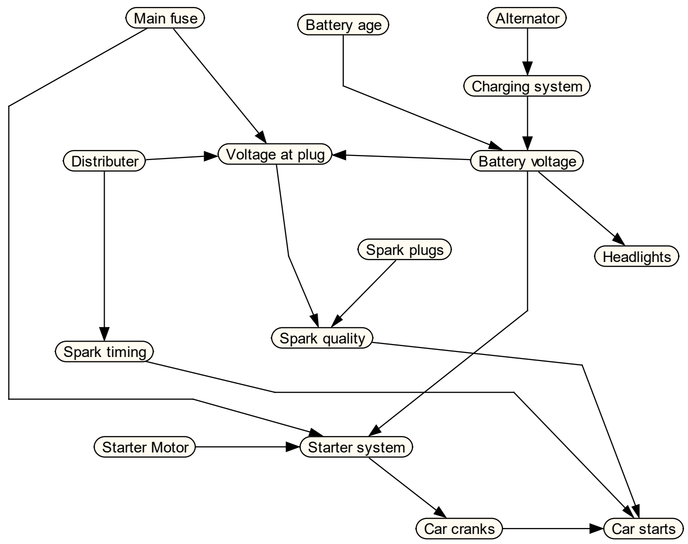
- Hence, we have no choice but to make all the nodes in A, B, and C parents of X and Y in the BN after the reversal operation.

Avoid arc reversal operations that introduce many additional arcs

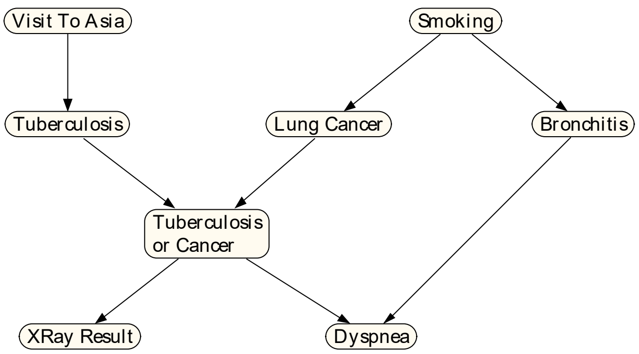
- Note that adding arcs after a reversal operation amounts to losing information about conditional independence. Hence if possible, we should avoid arc reversals that will introduce additional arcs.

5.1.7 Example Bayesian Networks

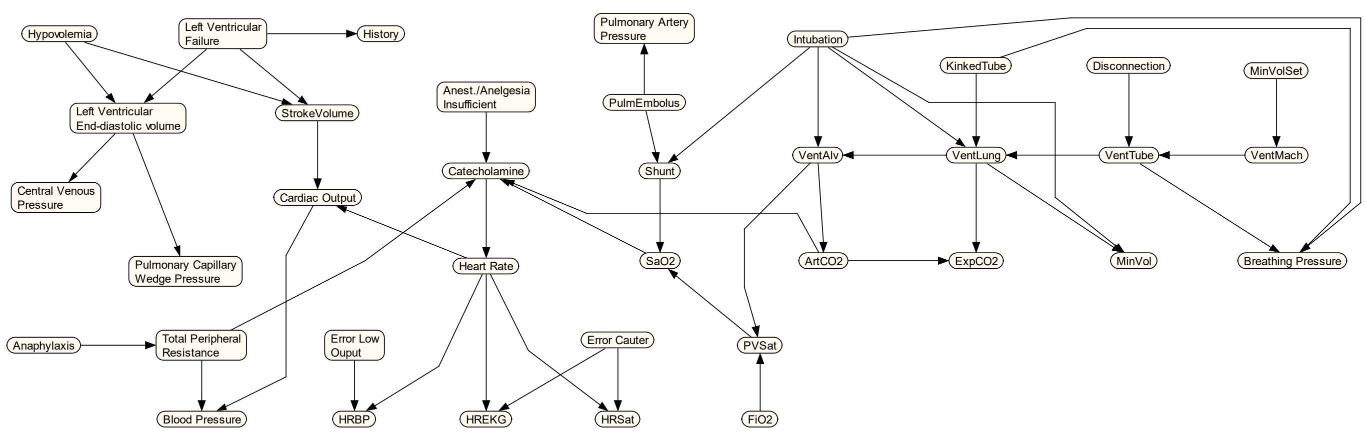
Car Starting System



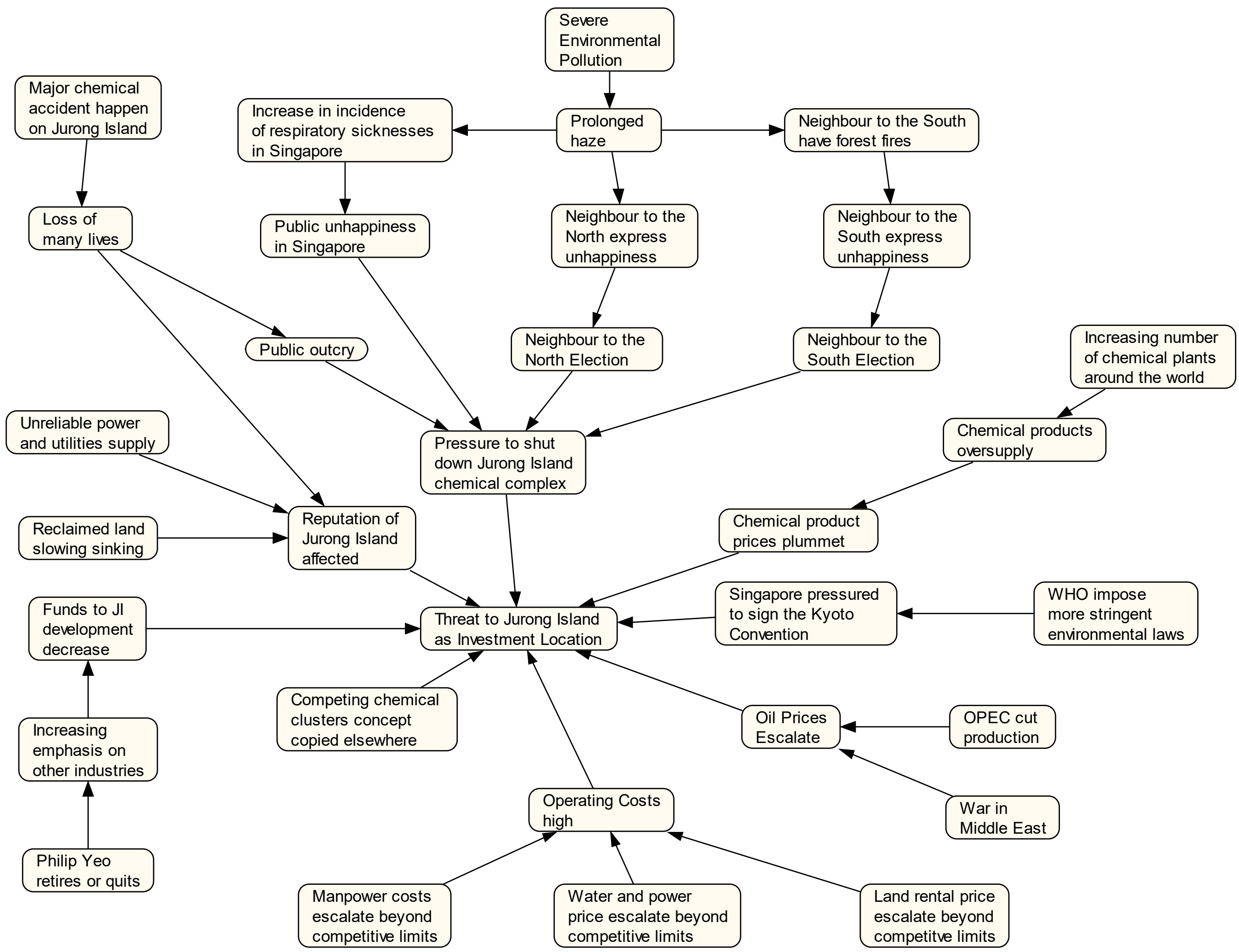
Cause of Dyspnea



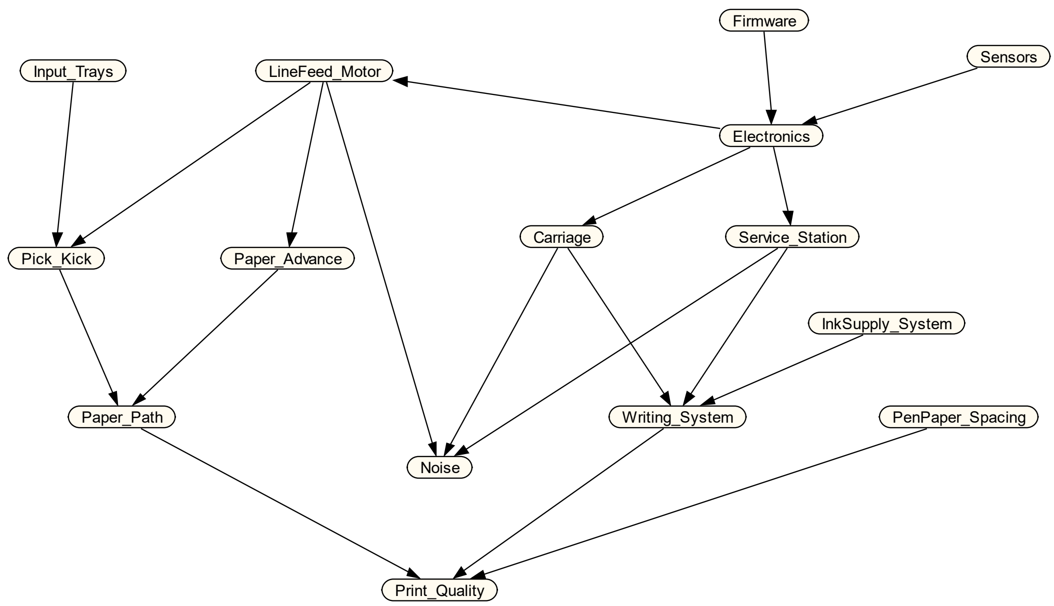
Patient Monitoring in an ICU (Stanford ALARM project)



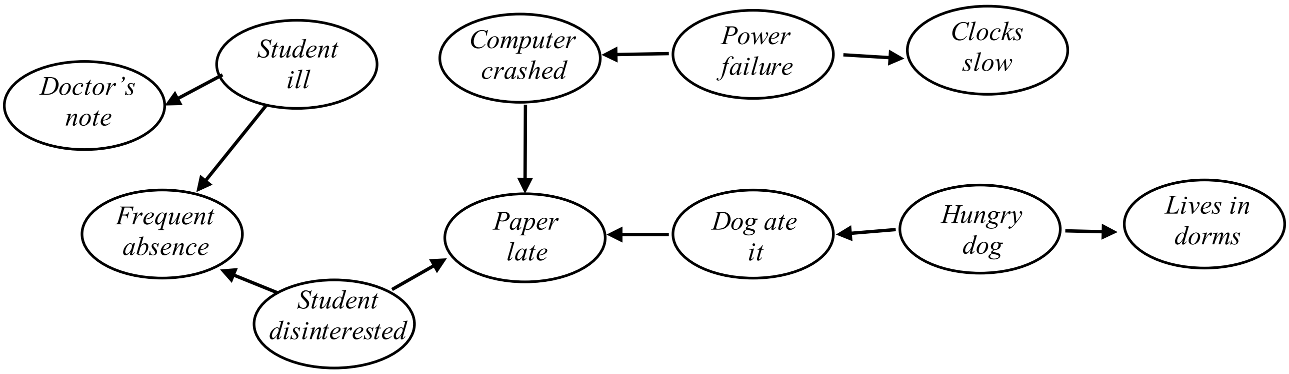
Threat to Jurong Island as an Investment Location



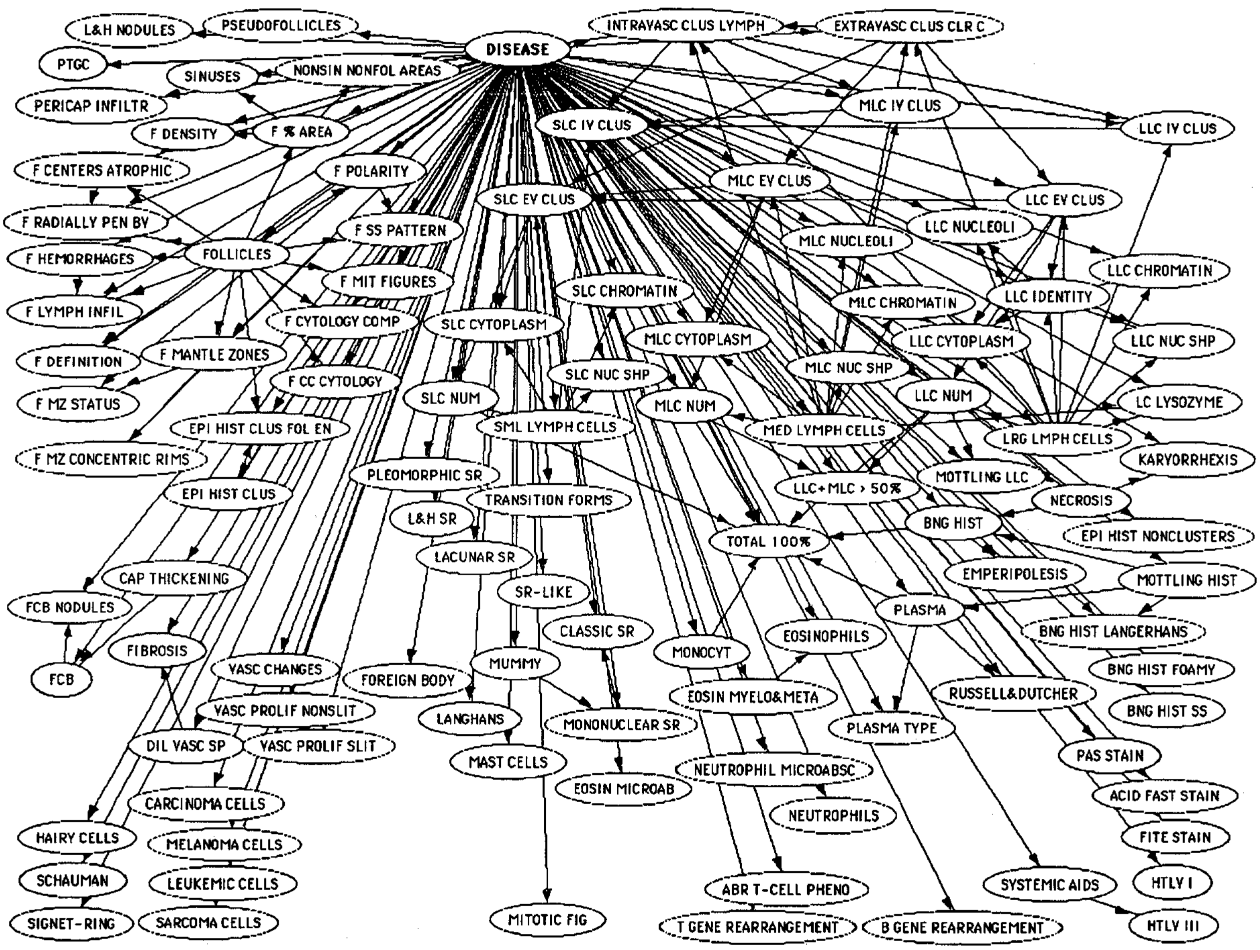
Ink-Jet Printer Trouble Shooting



Why did a student turn in an assignment late?



The Pathfinder Project (Stanford)



5.1.8 Probabilistic Inference using Bayesian Networks

The Probabilistic Inference Problem

- Let $\{X_1, X_2, \ldots, X_n\}$ be n uncertain variables that can be modeled with a Bayesian network.
- The *Probabilistic Inference Problem* is to compute the conditional probability $p(Y \mid \boldsymbol{E}=\boldsymbol{e})$

  where

  - $Y \in \{X_1, X_2, \ldots, X_n\}$, is any target variable of interests, and
  - $\boldsymbol{E} \subseteq \{X_1, X_2, \ldots, X_n\} \setminus \{ Y\}$, represents a set of variables or evidence nodes whose outcomes are already known or assumed to be equal to $\boldsymbol{e}$.

- It has been shown the Probabilistic Inference Problem is *NP*-hard when the Bayesian network is a general directed acyclic graph.
- Various algorithms have been developed to perform probabilistic inferences for different types of networks. See books on Bayesian networks for details,

Types of Probabilistic Inference Problem

- Probabilistic inference on Bayesian networks may be classified as follows:

  1. Predictive or Causal Inference
  2. Diagnostic Inference
  3. Inter-Causal Reasoning
  4. General or Mixed inference

- Each of the above methods of inference is illustrated using the lung cancer example.
- More detailed discussion of the methods will be covered in the laboratory session using Netica software.

5.2 Influence Diagrams

5.2.1 Decision Modeling using Influence Diagrams

- Bayesian networks represent probabilistic relationships among uncertain variables. They are useful for pure probabilistic reasoning and inferences.

- In this section, we will extend Bayesian networks to Influence diagrams to represent decision problems by adding decision nodes and value nodes.

- This is analogous to extending a probability tree to a decision tree by adding decision branches and adding values or utilities to the end points of the tree.

Definition of Influence Diagram (ID)

- An **Influence Diagram** (also known as a Decision Diagram) is a **Directed Acyclic Graph** representing a decision problem. The nodes and arcs of an influence diagram are described below:

Chance Node

- A *Chance Node* (denoted by an oval) in an influence diagram represents an uncertain variable.
- In each chance node, we assess a conditional probability distribution for the uncertain variable it represents, conditioned on its parent nodes.

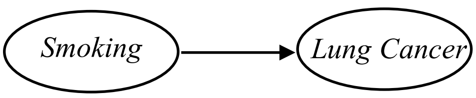
Relevance Arc

- An arc between two chance nodes in an influence diagram is known as a *Relevance Arc*.
- It denotes possible probabilistic dependence between the two variables.
- A relevance arc in an influence diagram may be reversed if doing so will not create any directed cycle in the diagram. Note that additional arcs may be added as a result of a relevant arc reversal.

Example

- The risk of having lung cancer is dependent on smoking habits.

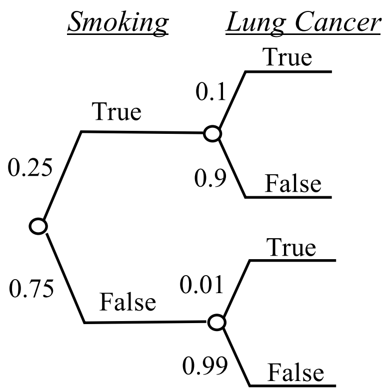
Influence Diagram



Smoking	$p(S \mid \xi)$
True	0.25
False	0.75

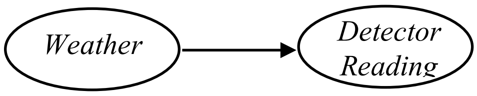
	$p(L \mid S, \xi)$	
	Lung Cancer	
Smoking	True	False
True	0.10	0.90
False	0.01	0.99

Equivalent Probability Tree



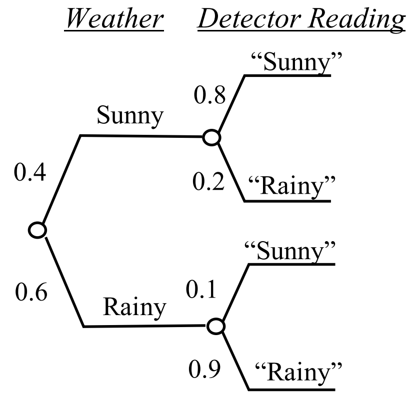
Example

- The detector's reading depends on the actual weather condition.

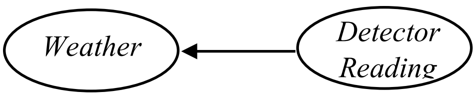


Weather	$p(W \mid \xi)$
Sunny	0.4
Rainy	0.6

	$p(R \mid W, \xi)$	
	Detector Reading	
Weather	"Sunny"	"Rainy"
Sunny	0.8	0.2
Rainy	0.1	0.9

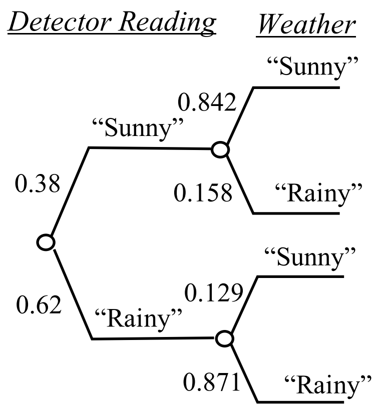


- The relevance arc between weather and detector reading may be reversed and the probabilities in the two chance nodes are updated.



	$p(W \mid R, \xi)$	
	Weather	
Detector Reading	"Sunny"	"Rainy"
"Sunny"	0.842	0.158
"Rainy"	0.129	0.871

Detector Reading	$p(R \mid \xi)$
"Sunny"	0.38
"Rainy"	0.62



Decision Node

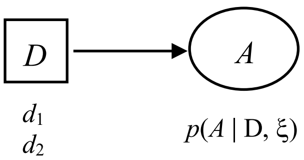
- A *Decision Node* (denoted by a rectangle) in an influence diagram represents a decision variable.
- In each decision node, we indicate a list of alternatives for the decision variable it represents.

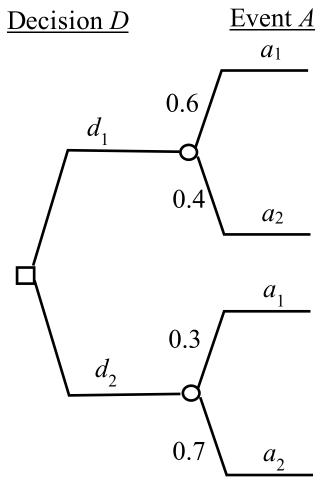
Influence Arc

- An arc from a decision node to a chance node in an influence diagram is known as an *Influence Arc*.
- It indicates that the probability of the uncertain variable is dependent on the alternatives of the decision node.
- An influence arc cannot be reversed.

Example

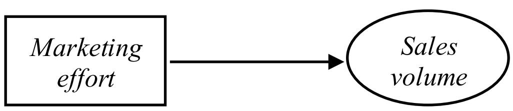
- The probability for the outcomes of A depends on the actual decision made at D. We therefore assess a conditional probability at A conditioned on each alternative at D.





Example

- The performance of a product in a market (sales volume) depends on how much marketing was done.



High
Low

p(Sales=good | Marketing = high, ξ) = 0.9
p(Sales=bad | Marketing = high, ξ) = 0.1

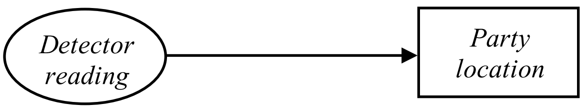
p(Sales=good | Marketing = low, ξ) = 0.3
p(Sales=bad | Marketing = low, ξ) = 0.7

Information Arc

- An arc from a chance node into a decision node in an influence diagram is known as an *Information Arc*.
- It indicates that the outcome of the chance node will be known to the decision maker when the decision is being carried out.
- An information arc cannot be reversed.

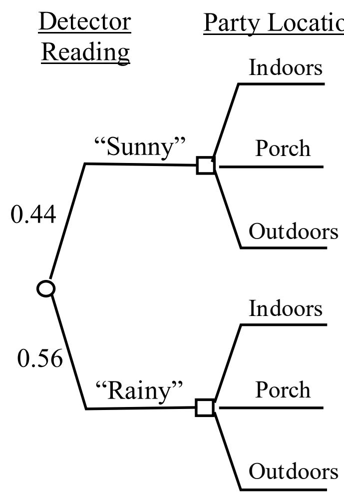
Example

- We are currently uncertain about the outcome of “Detector Reading”, but we do know its probabilities.
- However, the exact outcome of “Detector Reading” (i.e., “sunny” or “rainy”), will be known to the decision maker when he decides on the party location.



p(Reading=”Sunny”, ξ) = 0.44
p(Reading=”Rainy”, ξ) = 0.56

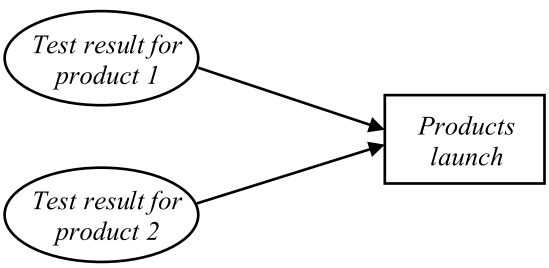
Indoors
Porch
Outdoors

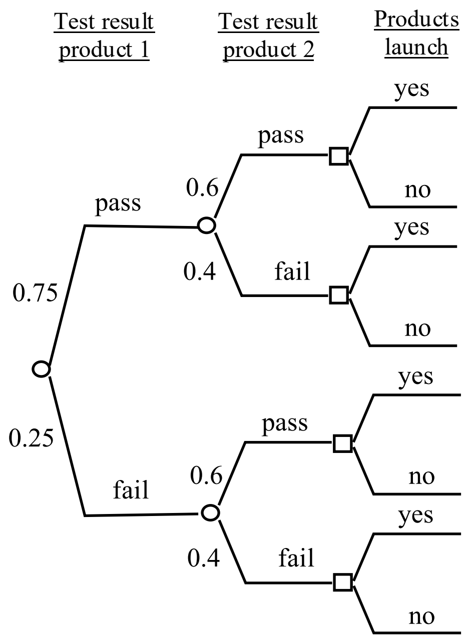


Multiple Information Arcs

- A decision node may have multiple information arcs.

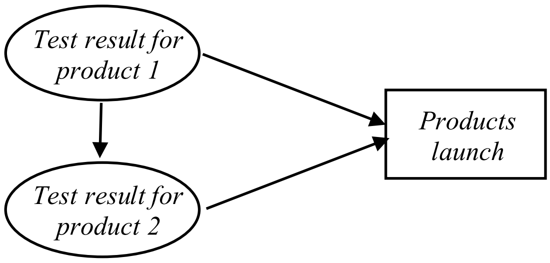
Example

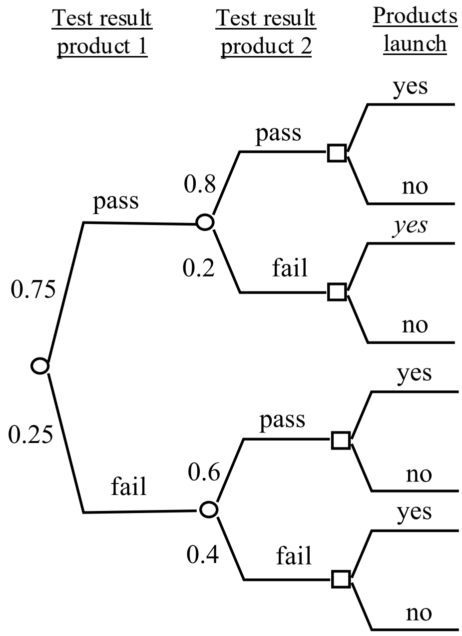




- The results of the two tests are currently uncertain. However, at the time when we carry out the decision to launch products, we will know the outcomes of “Test result for product 1” and “Test result for product 2.” Moreover, the two test results are mutually independent.

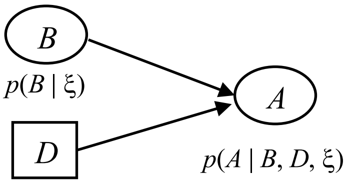
- What about the following influence diagram?

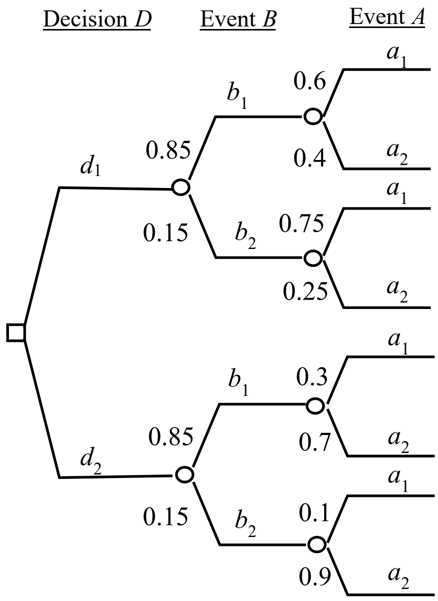




- This is the same as the previous example, except that two test results may not be independent.

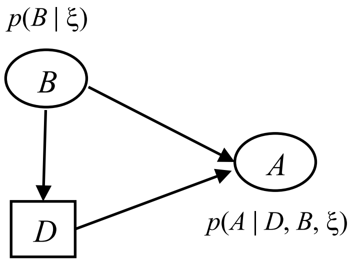
Example (Combining influence and relevance arcs)

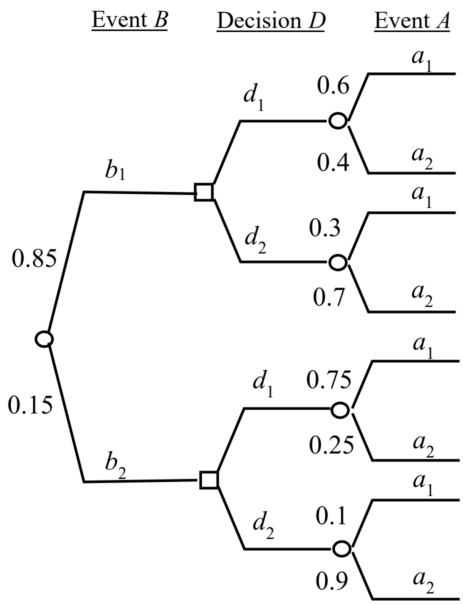




- The outcomes of A and B are uncertain. The probability of A depends on the outcome of B and the decision made at D. At the time decision D is carried out, we do not know the outcomes of B and A.
- Note that the decision tree with the nodes ordering B, D, A is wrong. Why?

Example (Combining influence, relevance and information arcs)



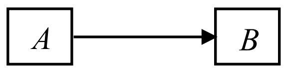


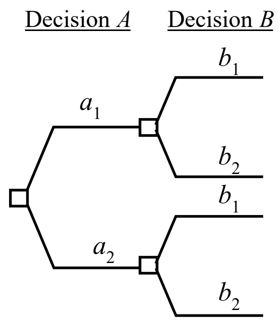
- The outcomes of A and B are uncertain. The probability of A depends on the outcome of B and the decision made at D. At the time decision D is carried out, we know the outcome of B, but not of A.

Chronological Arc

- An arc from a decision node to another decision node in an influence diagram is known as a *Chronological Arc.*
- It indicates the *chronological order* in which the decisions are being carried out.
- A chronological arc cannot be reversed.

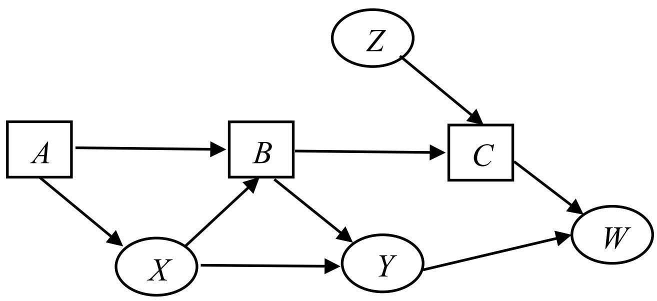
Example





- Decision B is made after Decision A.

Example (Relevance, Influence, Information and Chorological arcs)



- Decision A is carried out before decision B which is carried out before decision C.
- Uncertain variable X is influenced by decision A.
- At the time decision B is carried out, the outcome of X and the choice made at A are known.
- Uncertain variable Y is influenced by decision B and the outcome of X.
- At the time decision C is carried out, the outcome of Z and the choice made at B are known.
- Uncertain variable W is influenced by decision C and the outcome of Y.

- Can you draw (or specify the sequence of nodes for) the equivalent decision tree?

Value Node

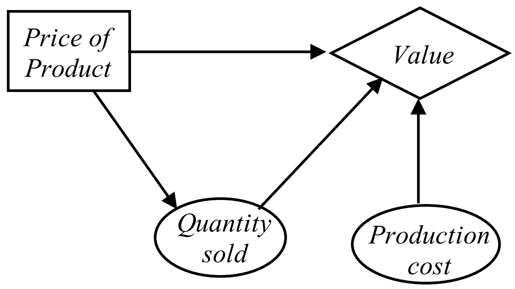
- A *Value Node* (denoted by a diamond) in an influence diagram represents the quantity whose expected value is to be optimized.
- A value node is used to represent the utility or value function of the decision maker.
- A value node must be a sink node, i.e., it has only incoming arcs.
- Only one value node is allowed in a standard influence diagram.

Value Arc

- An arc from any node to the value node is known as a *Value Arc*.
- Value arcs indicate the variables whose outcomes the decision maker cares about. i.e., have direct impact on his utility.
- A value arc cannot be reversed.

Example

- The decision maker is concerned about the profit which is affected by the product price, quantity sold and production cost.



Profit = Price * Quantity Sold – Total Production Cost

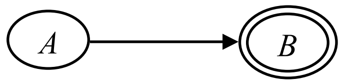
- Can you draw (or specify the sequence of nodes for) the equivalent decision tree?

Deterministic Node

- A *Deterministic Node* (denoted by a double oval) in an influence diagram is a special type chance node.
- It represents a variable whose outcome is known with certainty (i.e., has probability = 1), once the outcomes of other conditioning nodes are known.
- A deterministic node can be used to represent a function.

Example

- Variable B is deterministically dependent on A, i.e., if the outcome of A is known, then the outcome of B is also known exactly.



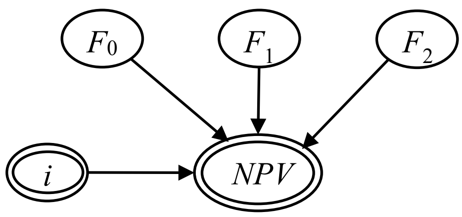
A	$p(A)$
a_1	0.2
a_2	0.7
a_3	0.1

	$p(B \mid A)$	
	b_1	b_2
a_1	1	0
a_2	0	1
a_3	1	0

- Also, B is function A such that $f(a_1) = b_1, f(a_2) = b_2$, and $f(a_3) = b_1$.

Example

- F_o, F_1, and F_2 are mutually independent uncertain cash flows at times 0, 1, and 2, respectively. The discounting (interest) rate i is certain.



$$NPV = F_0 + \frac{F_1}{(1+i)} + \frac{F_2}{(1+i)^2}$$

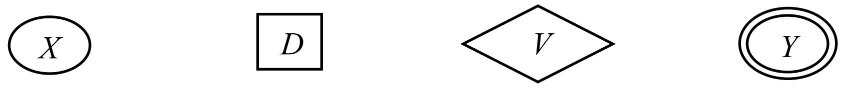
- Note that although NPV is a deterministic node, it is actually a random variable as it is a function of three random variables.

- If the cash flows F_o, F_1, and F_2 are not mutually independent, add relevance arcs between them.

Summary on Influence Diagram

- An **Influence Diagram** is a *Directed Acyclic Graph* (DAG) representing a decision model.

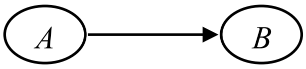
Types of Node:



Chance node Decision node Value node Deterministic node

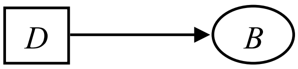
Types of Arc:

Relevance Arc



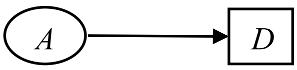
The probability for *B* may be dependent on the outcome of *A*.

Influence Arc



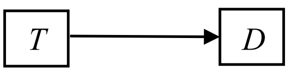
The probability for *B* may be dependent on alternative chosen at *D*.

Information Arc



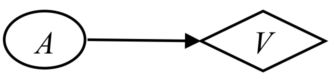
The decision maker knows the outcome of *A* when carrying out decision *D*.

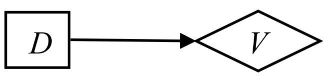
Chronological Arc



Decision *T* is made before decision *D*.

Value Arc





Outcome of *A* has direct impact on Value.

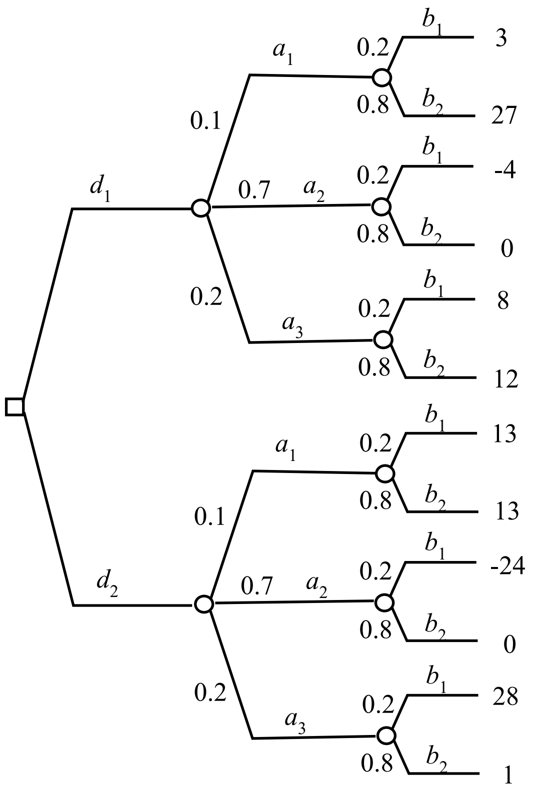
Decision made a *D* has direct impact on Value

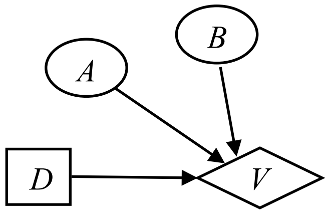
5.2.2 Comparing Influence Diagrams and Decision Trees

	Influence Diagram	Decision Tree
1	**Compact** The size of an influence diagram is equal to the total number of variables.	**Combinatory** The size of a decision tree grows exponentially with the total number of variables. A binary tree with n nodes has 2^n leaf nodes.
2	**Graphical Representation of Independence** Conditional independence relations among the variables are represented by the graphical structure of network. No numerical computations needed to determine conditional independence relations.	**Numerical Representation of Independence** Conditional independence relations among the variables can only be determined through numerical computation using the probabilities.
3	**Non-Directional** The nodes and arcs of an influence diagram may be added or deleted in any order. This makes the modeling process flexible.	**Unidirectional** A decision tree can only be built in the direction from the root to the leaf nodes. The exact sequence of the nodes or events must be known in advance.
4	**Symmetric Model only** The outcomes of all nodes must be conditioned on all outcomes of its parents. This implies that the equivalent tree must be symmetrical.	**Asymmetric Model possible** The outcomes of some nodes may be omitted for certain outcomes of its parent leading to an asymmetrical tree.

Examples

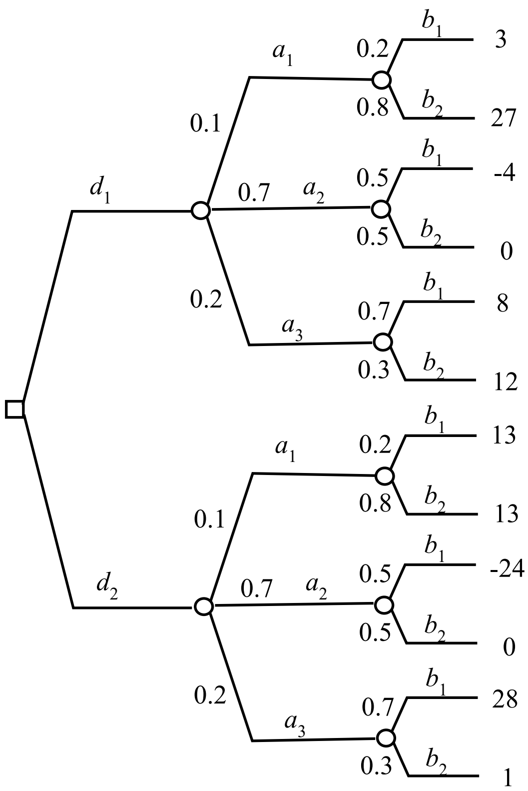
Case 1

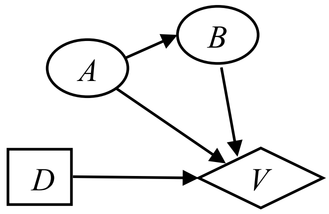




- In the decision tree, we can only infer that $A \perp B$, $A \perp D$, and $B \perp D$ through the numerical probability values.
- In the influence diagram, these independent conditions are explicitly expressed by the graphical structure.

Case 2

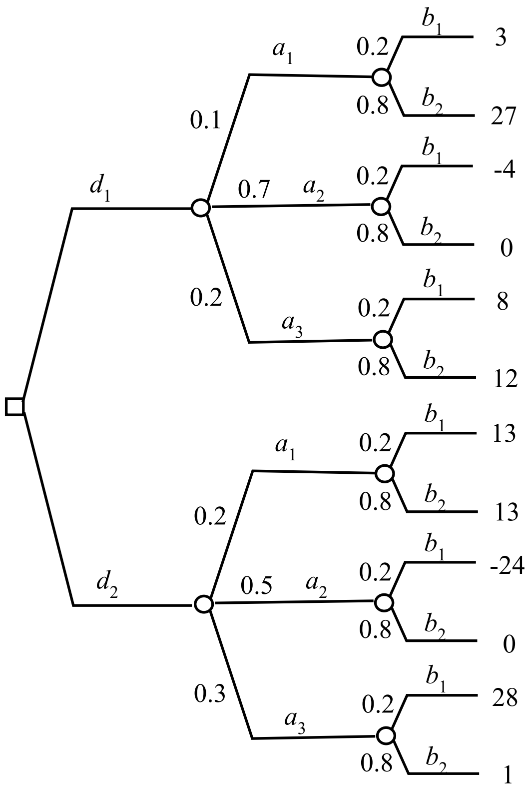


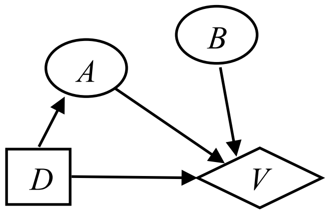


- In the decision tree, we can only infer that $A \perp D$, $B \perp D$, but B is dependent on A through the numerical probability values.

- In the influence diagram, these independent conditions are explicitly expressed by the graphical structure.

Case 3



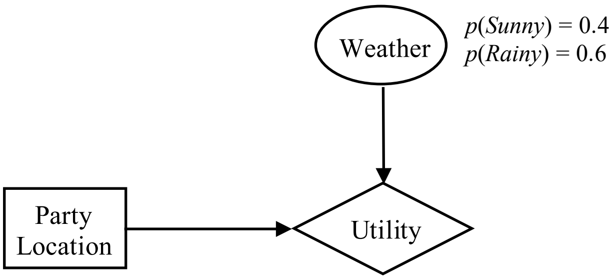


- In the decision tree, we can only infer that $A \perp B$, $B \perp D$, but A is dependent on D through the numerical probability values.

- In the influence diagram, these independent conditions are explicitly expressed by the graphical structure.

5.3 Modeling the Party Problem using Influence Diagram

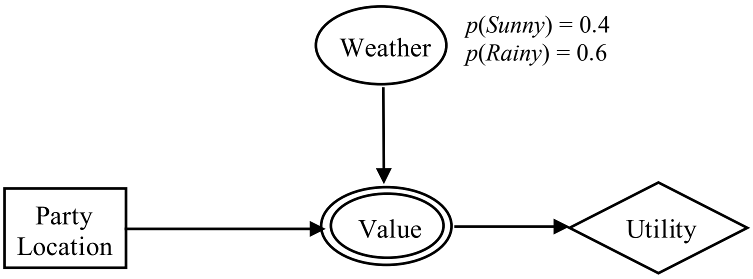
Base model for Kim's Party Problem with direct assessment of utilities



Indoors
Porch
Outdoors

	Sunny	*Rainy*
Indoors	0.57	0.67
Porch	0.95	0.32
Outdoors	1	0

Kim's Party Problem with equivalent dollar values and utility function

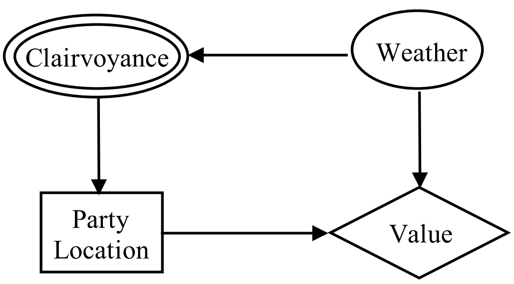


Indoors
Porch
Outdoors

	Sunny	*Rainy*
Indoors	40	50
Porch	90	20
Outdoors	100	0

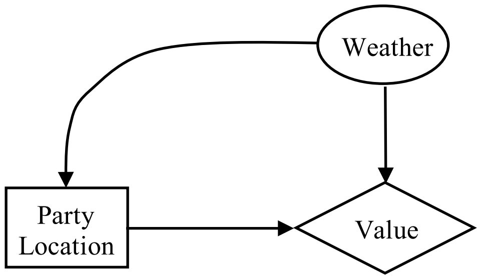
$$u(x) = \frac{4}{3}\left(1 - e^{-x/72.13}\right)$$

Decision Model with Perfect Information on Weather



- Here, "Clairvoyance" is a deterministic node because it is always equal to the Weather outcome with probability 1.

- The same result is obtained if we simply draw an information arc directly from Weather to Party Location:

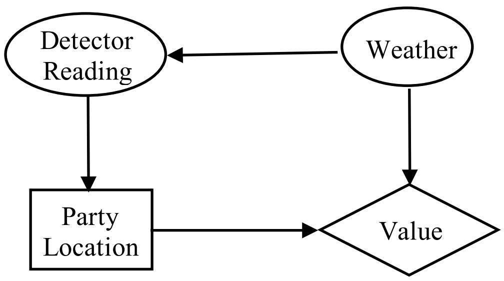


- Having clairvoyance on weather is the same as being able to "observe" the weather prior to deciding on the location.

- Hence in general, to represent the situation where we have perfect information (clairvoyance) on an uncertain event, we simply draw an information arc directly from the said event to the decision node. However, in doing so, we must not create any cycle.

Decision Model with Imperfect Information on Weather

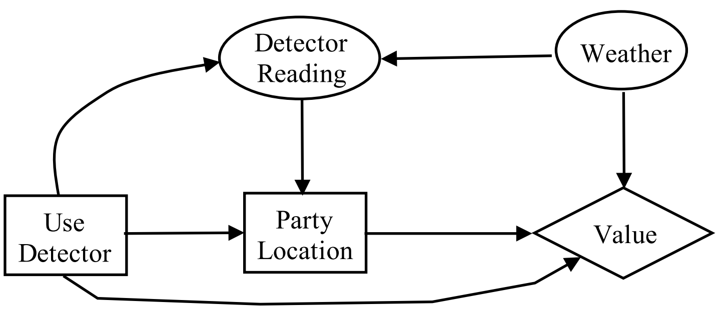
- Suppose we do not have perfect information on weather, but rather, imperfect information via a weather detection device, then the influence diagram representation for the situation is as follows:



- We input the "true-sunny detection rate" and "true-rainy detection rate" in the "Detector Reading" node.
- The detector reading is known before the location decision is carried out. Hence add an *information arc* from the observed node (detector indicator) to the location decision node.

Decision Model with Option to use Weather Detector or not

- If the decision maker has the option to use the rain detector or not we add a decision node at the front:



5.4 Evaluating Influence Diagrams

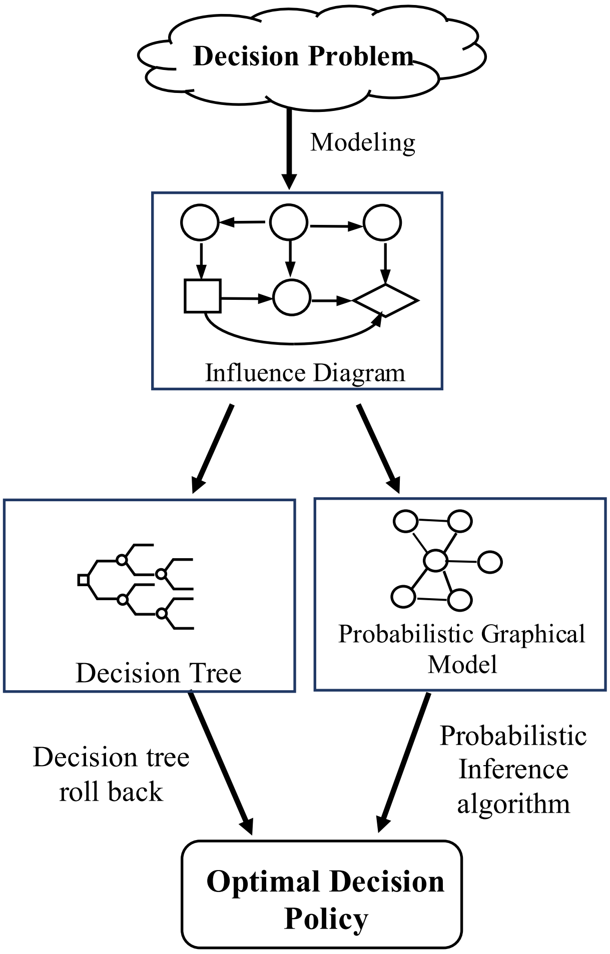
5.4.1 Approaches to Solving an Influence Diagram

- To find the optimal decision policy of a decision problem represented by an influence diagram we need to evaluate it.

- Methods for evaluating an influence diagram:

  1. Convert the influence diagram into an equivalent decision tree and perform tree roll back.

  2. Convert the influence diagram into a probabilistic graphical model (PGM) and apply probabilistic inference algorithms.



- See Shachter (1986) for the first algorithm for evaluation an ID directly without converting it into a decision tree, and books on traditional artificial intelligence for probabilistic graphical models (PGM). Bayesian network (BN) is one type of PGM.

- Note that DPL software uses an influence diagram and a decision tree to model a decision problem. It then automatically generates the optimal decision policy by expanding and solving the decision tree.

5.4.2 The Decision Network Conditions

- An influence diagram may be used to represent a decision problem, but before we can solve it by any methods to find the optimal decision policy, we must ensure that it is properly specified.

- The **Decision Network Conditions** ensure that an influence diagram is solvable, i.e., there is no ambiguity in the information.

- An influence diagram is said to satisfy the **Decision Network Conditions** if the following two conditions are true:

  1. **The Complete Decisions Ordering Condition**

     - There exists a complete chronological order for all the decision nodes.
     - This means that the decisions are made sequentially and there are no independent or parallel decisions involved.
     - This condition is also called the "Single Decision Maker" condition.

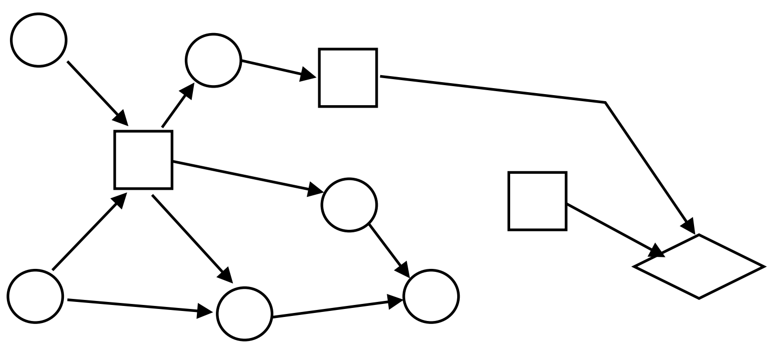
  2. **The "No-Forget" Condition**

     - All direct predecessors (parents) of a decision node are also direct predecessors (parents) of all decision nodes down the chronological order.
     - This means that information available at one decision point is always available (not forgotten) at subsequent decision points.

Example

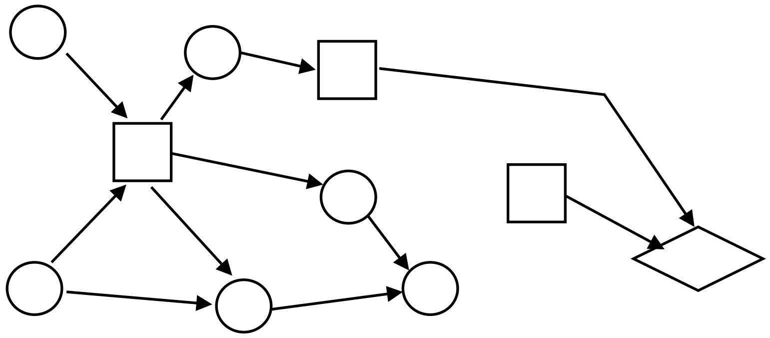
- The following ID is not a decision network. It cannot be solved.

- Can you identify the violations?

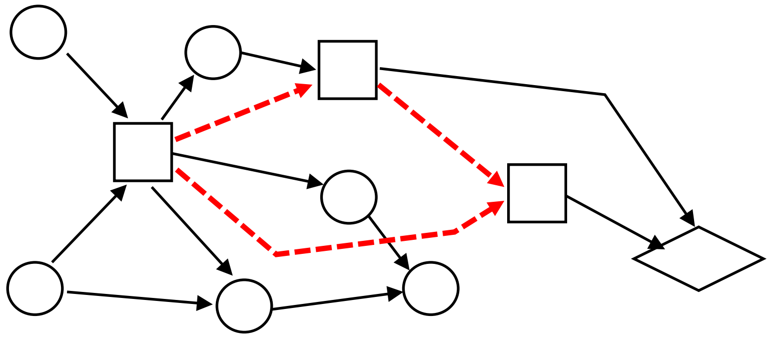


Example

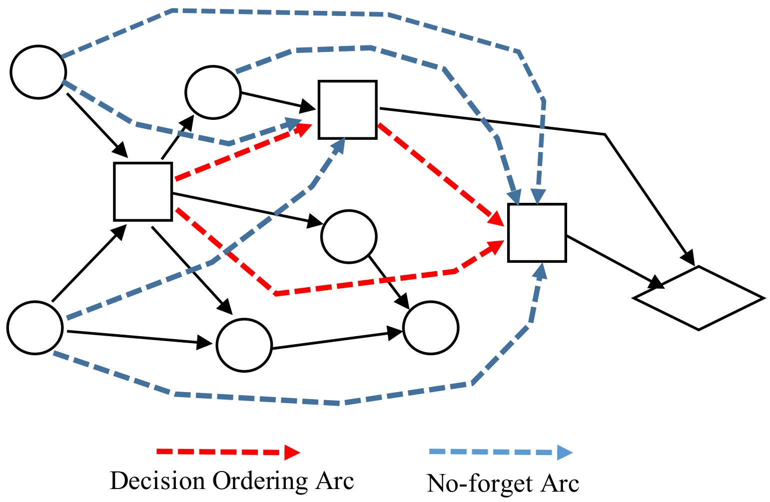
- Convert the previous influence diagram into a Decision Network:



- Adding Decision Ordering (Chorological) Arcs:



- Adding No-Forget Arcs:



- The influence diagram is now a Decision Network and can be solved without ambiguity.

5.4.3 Drawing a Decision Tree from an Influence Diagram

- We would like to convert an influence diagram (ID) into an equivalent decision tree so that we can solve it.

- The equivalent tree must satisfy the following constraints:

  1. The ID must be a decision network.

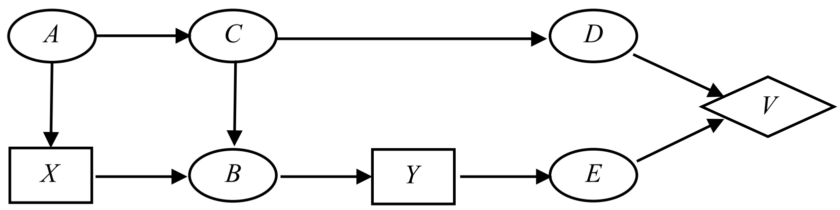
  2. The decision nodes in the DT must follow the same chorological order in the ID.

  3. A node that is a direct predecessor (parent) of a decision node in the ID should appear in the DT just before the decision node that first observed it.

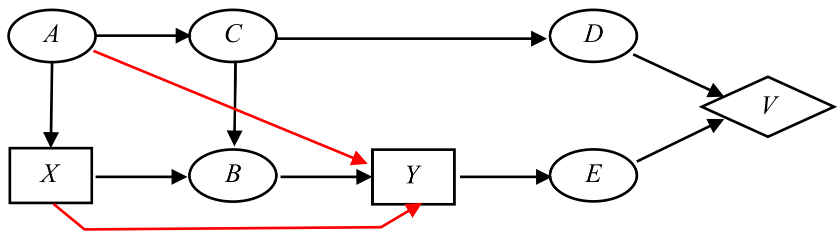
  4. A node that is not direct predecessor (parent) of any decision node in the ID should appear after the last decision node in the DT.

Example 1

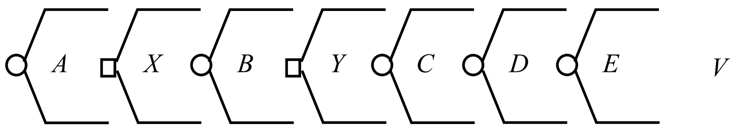
- Given the following influence diagram:



- Convert it into a decision network:



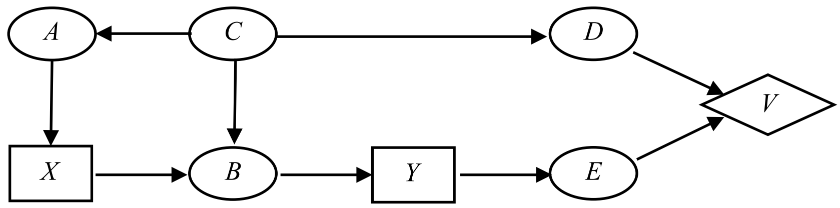
- An equivalent decision tree (in generic format) is as follows:



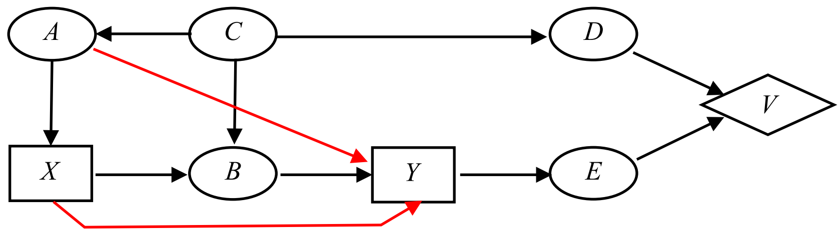
- Any permutation of the node sequence C, D, E is also valid.

Example 2

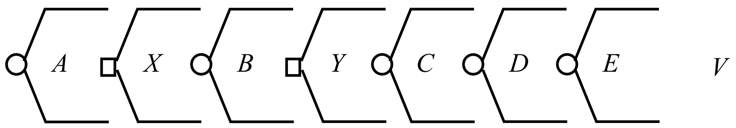
- Given the following influence diagram:



- Convert it into a decision network:

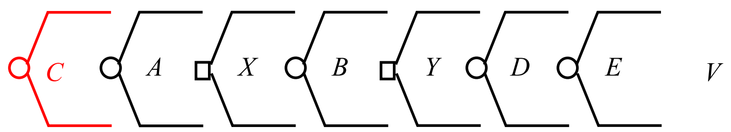


- An equivalent decision tree (in generic format) is as follows:



- Any permutation of the node sequence *C*, *D*, *E* is also valid.

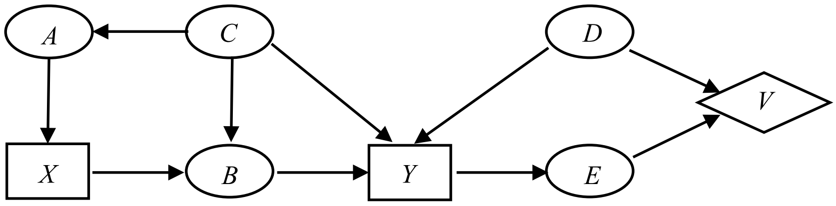
- Note that the following decision tree is **WRONG !**



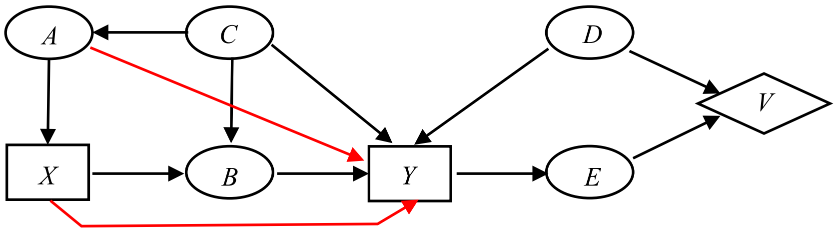
- Even though chance node *C* is a grandparent of *X* in the ID, it is not observed by either *X* or *Y* as there is no information arc from *C* to *X* and *Y*. Hence *C* cannot be drawn before the decision nodes *X* and *Y* in the DT.

Example 3

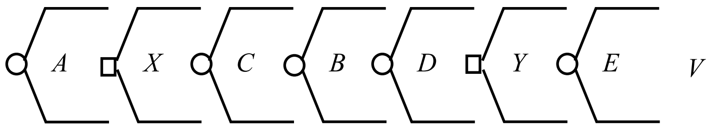
- Given the following influence diagram:



- Convert it into a decision network:



- An equivalent decision tree (in generic format) is as follows:



- Any permutation of the node sequence *C*, *B*, *D* is also valid.

5.4.4 Automatically Generating an Equivalent Decision Tree from an Influence Diagram

[*Optional Materials for Research and/or Computer Science Students*]

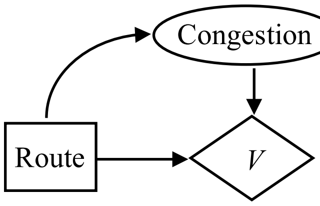
- We can automatically generate an equivalent Decision Tree from an Influence Diagram.
- Algorithm:
  1. If the ID is not a *decision network*, convert it by adding chorological and no-forget arcs.
  2. Perform arc reversal operations until the ID is a *decision tree network*.
  3. Find a valid sequence of nodes for the tree.
  4. Draw the Tree and copy the conditional probabilities directly from the ID to the Tree.
- For details, see Appendix A.

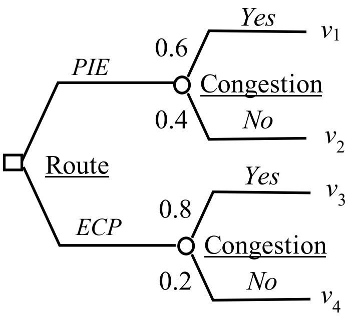
5.5 Influence Diagram in Canonical Form

Influence Diagrams with Influence Arc

- Consider the problem of getting from Changi Airport to the City. You have to decide on which route to take: PIE or ECP.

- Because the probability of encountering a congestion along the way depends on which route you took, you might model the decision problem as follows:

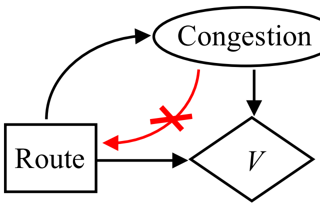


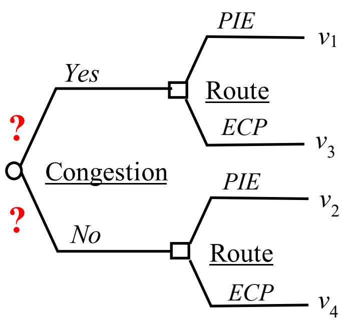


- Using this decision model, the best route can be determined.

- Suppose now you want to determine the expected value of perfect information on congestion, can it be done? Can you create a decision tree or influence diagram with perfect information?

- The value of information on congestion cannot be performed using the decision model created this way.





- If you add an information arc from chance node "Congestion" to decision node "Route" to denote perfect information, a loop will be created because there is already an arc from the "Route" to "Congestion".

- If we avoid the *influence arc* from "Route" to "Congestion", we can perform value of information analysis.

Influence Diagram in Canonical Form

Definition

- An influence diagram is said to be in ***Canonical Form*** if it does not have any chance nodes that are descendants of a decision node.

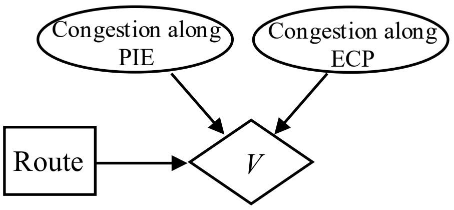
Drawing an Influence Diagram in Canonical Form

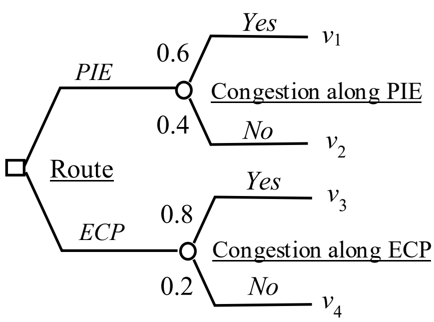
- An influence in non-canonical form can be redrawn into canonical form by replacing the chance node with multiple chance nodes.

Example

- We replace the single chance node "Congestion" by two chance nodes
  1. "Congestion along PIE"
  2. "Congestion along ECP"

- If we assume that congestion along PIE is independent of congestion along ECP, then we can model the problem as follows:



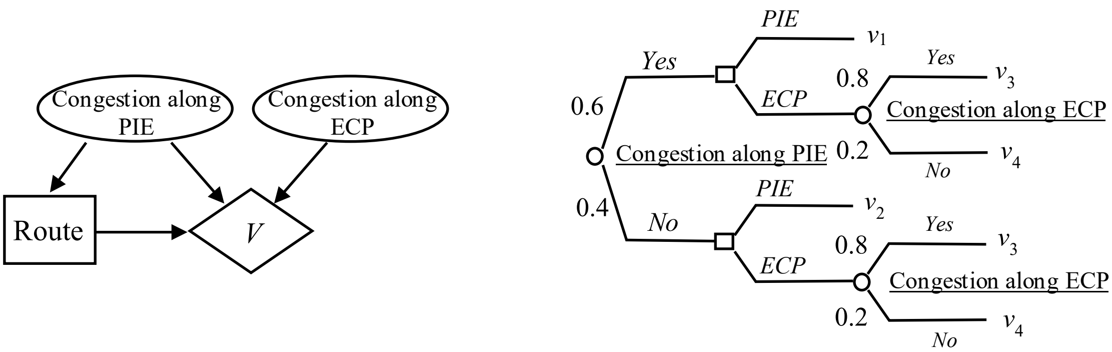


- The diagram is now in ***Canonical Form*** because the decision node does not have any descendant chance nodes.

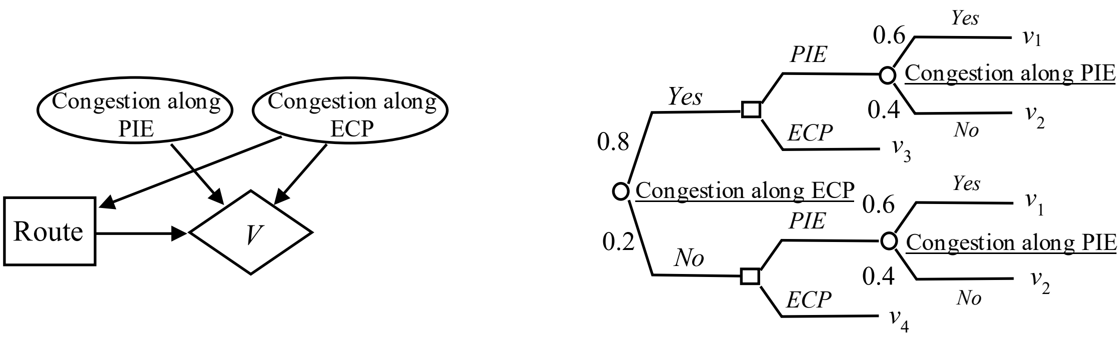
Performing Value of Information Analysis

- We can now perform value of information analysis, but the expected value of information are now computed specific to particular routes:

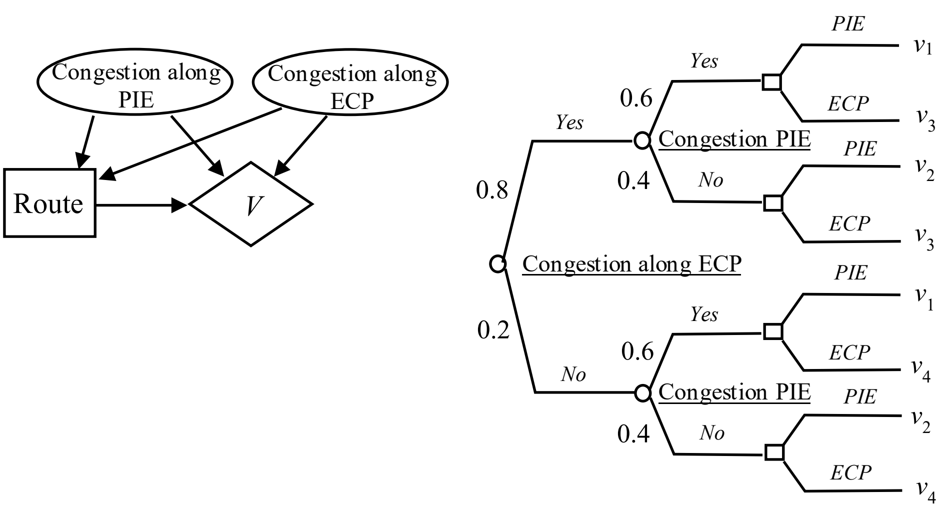
- Decision model with free perfect information on congestion along PIE:



- Decision model with free perfect information on congestion along ECP:



- Decision model with free joint perfect information on congestion along PIE and congestion along ECP:

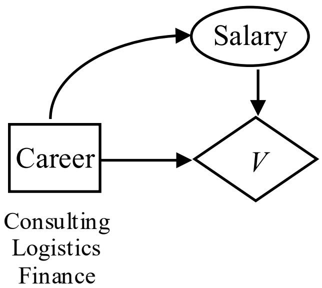


When the decision node has more than two alternatives

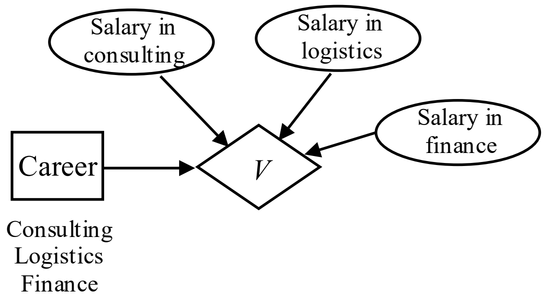
- We can usually draw an influence diagram in canonical form by defining as many chance nodes are there are number of alternatives in the decision node.

Example (Career Planning for ISyE Graduates)

- A popular ISEM graduate’s career choice decision problem:



<u>Non-canonical form</u>
Salary depends on your career choice

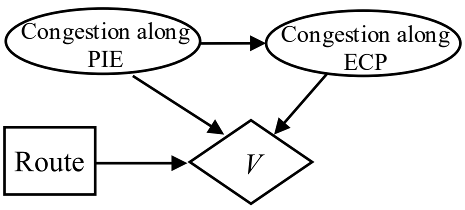


<u>Canonical form</u>
Salaries are different in each job sector and they are mutually independent

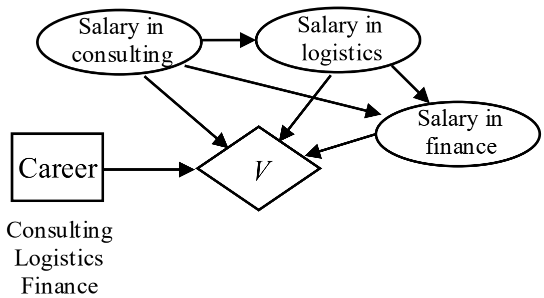
Canonical Form with Non-Mutually Independent Chance Nodes

- In general, the chance nodes do not need to be mutually independent. Relevance arcs can be added between them if needed.

Example



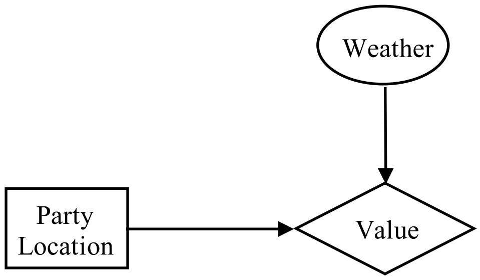
Congestion along ECP and congestion along PIE may not be independent

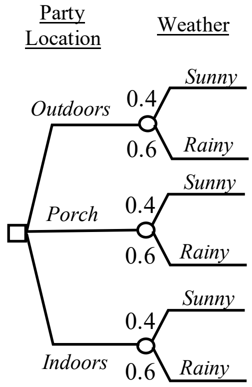


Salaries in the three job sectors may not be mutually independent

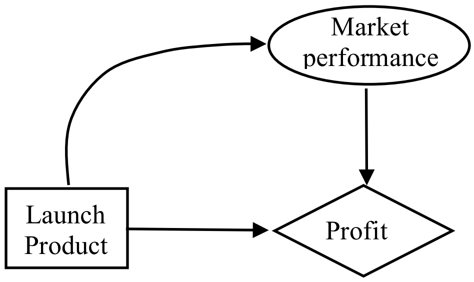
5.6 Example Influence Diagrams

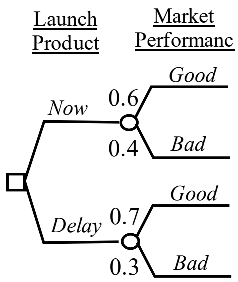
The Party Problem



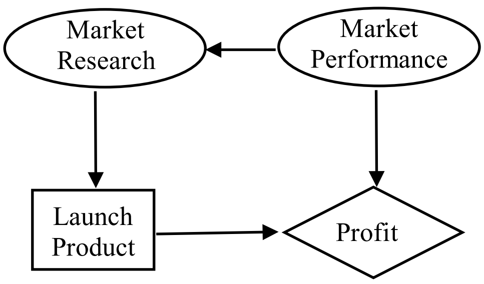


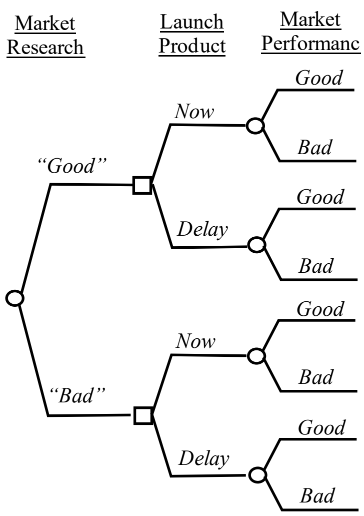
Basic Risky Decision Problem (One-Decision One-Uncertainty)



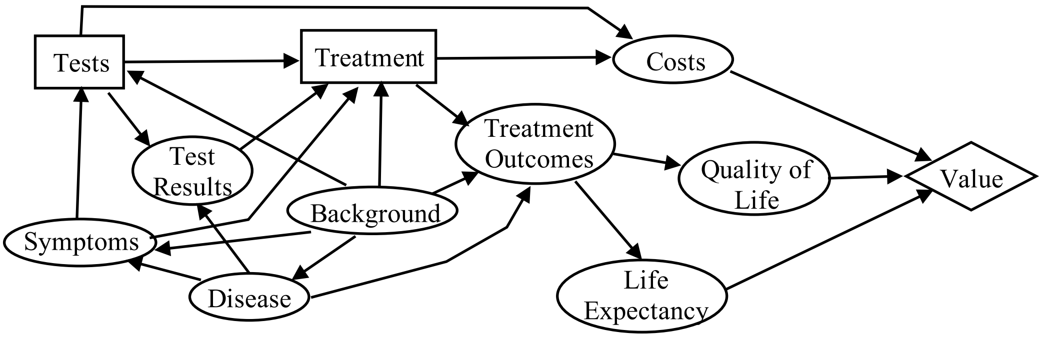


Decision Problem with Free Imperfect Information on the Uncertainty

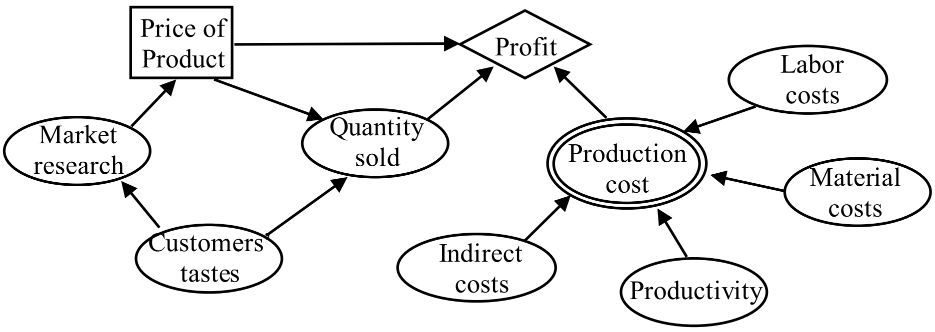




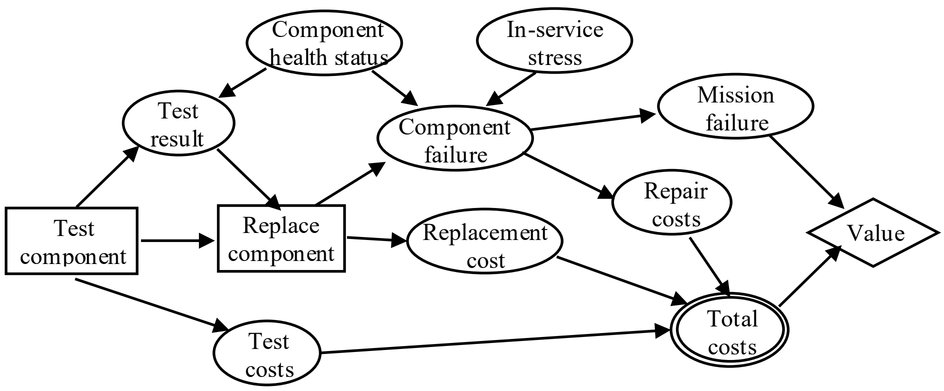
A Medical Decision Problem



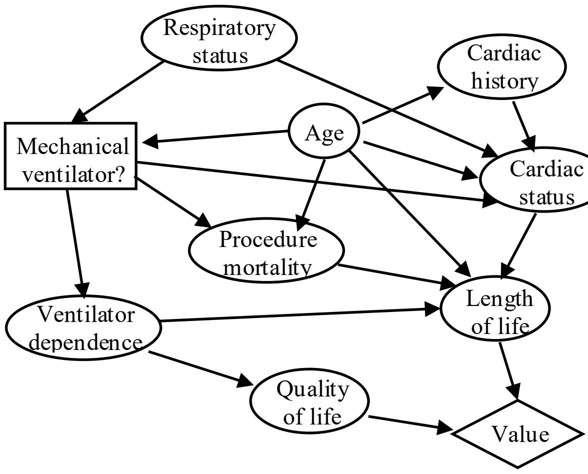
The Production/Sale Problem



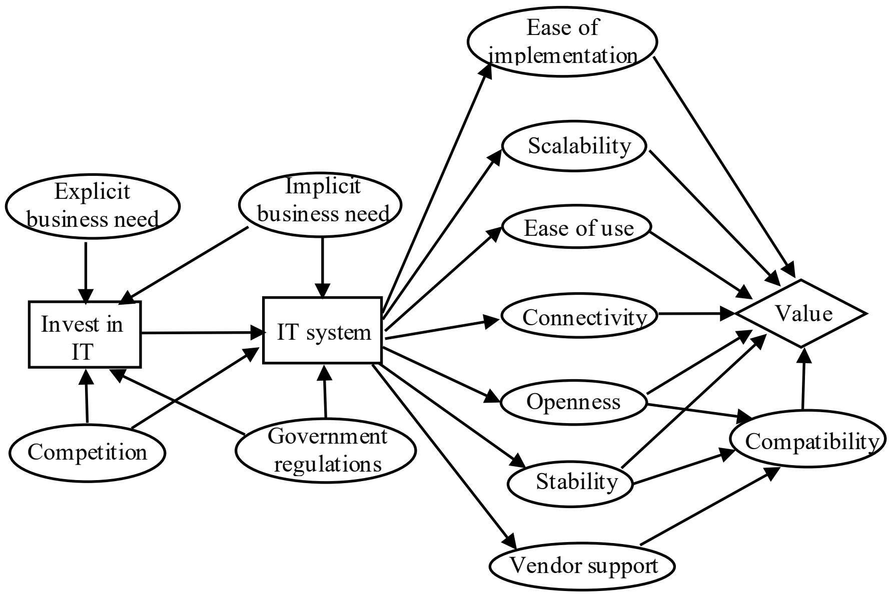
Critical Component Replacement Maintenance Decision Problem



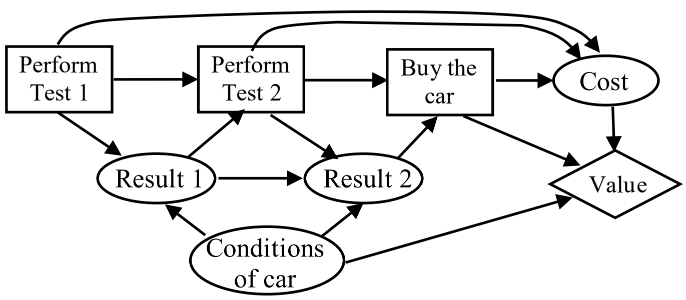
An Intelligent ICU Decision Problem



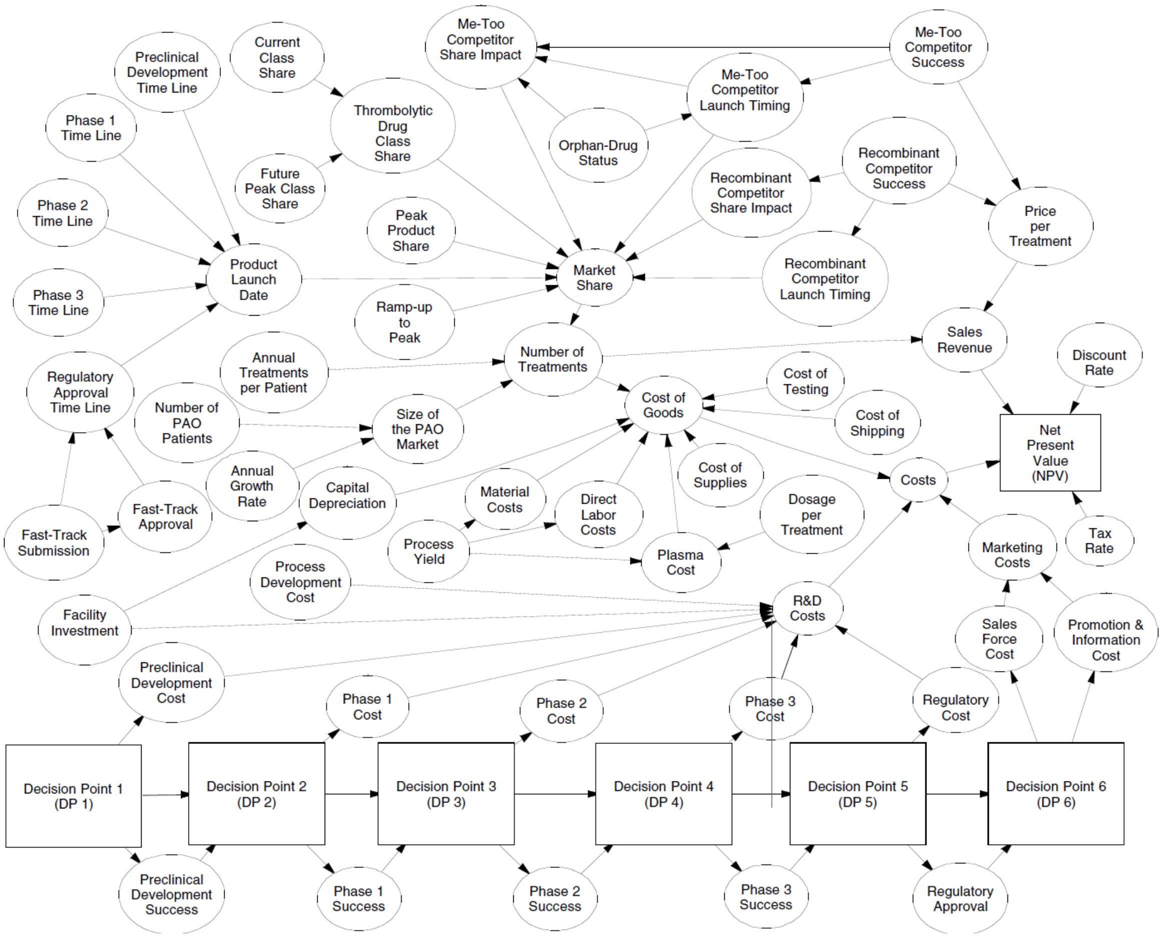
Investment Decision for an IT System



Buying a Used Car

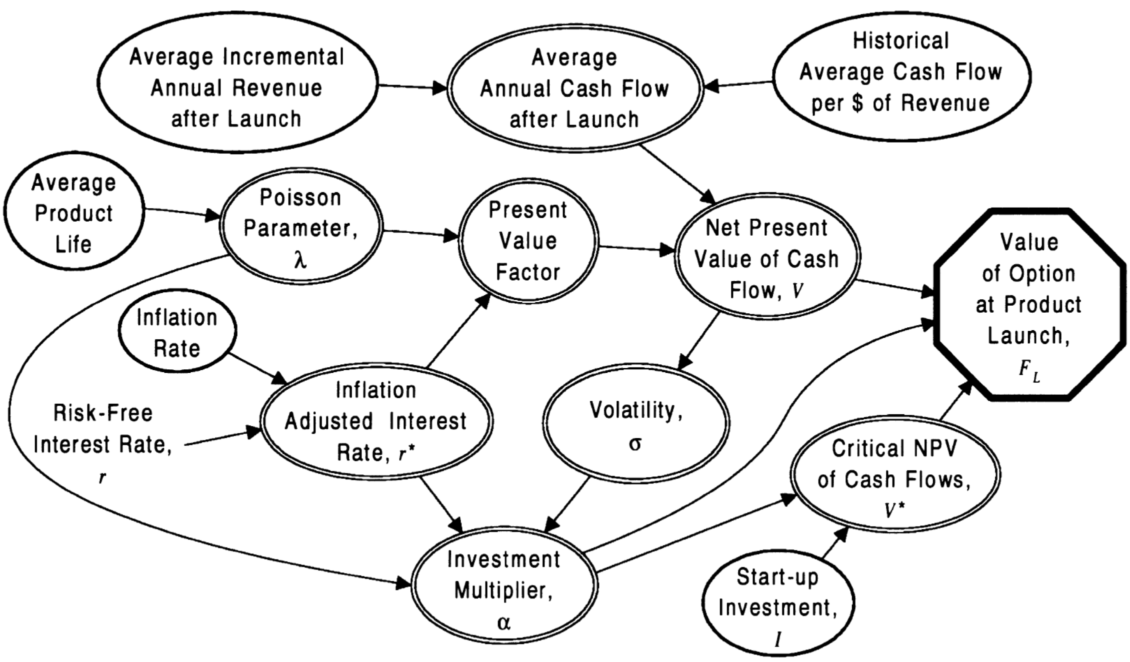


Decisions to Develop New Drugs (Stonebraker 2002)

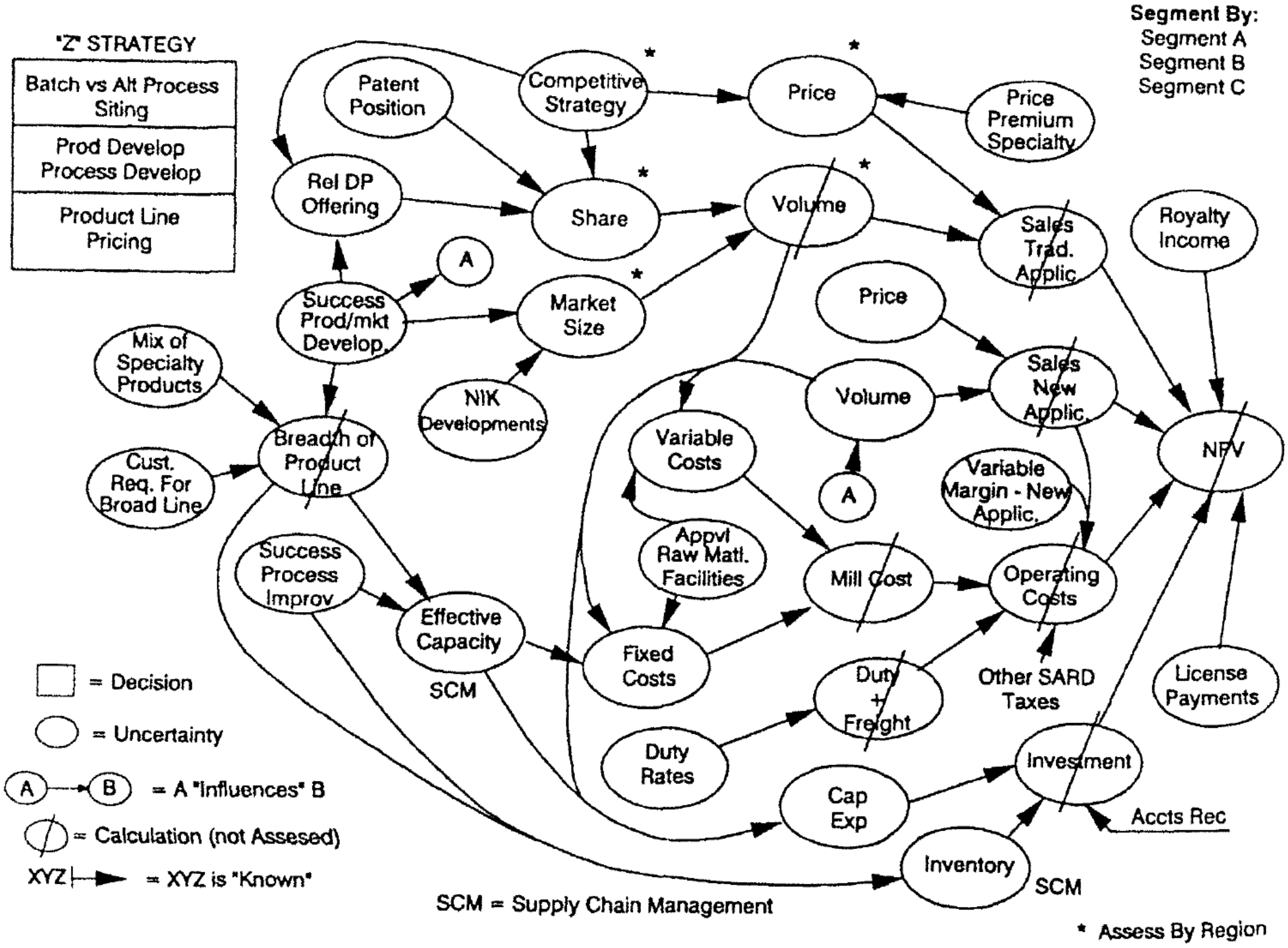


Valuation of R & D Projects using Options Pricing and Decision Analysis Models

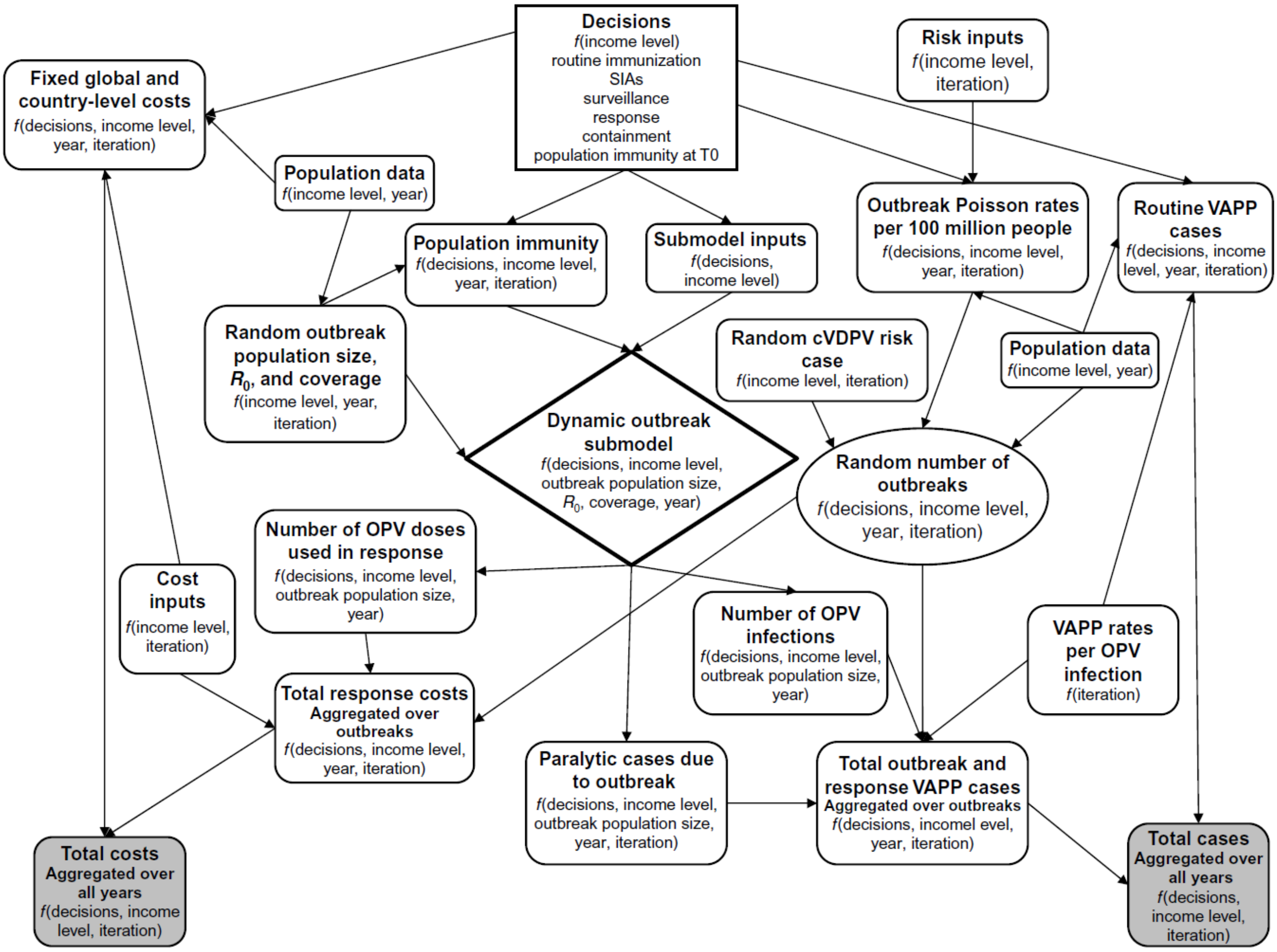
(Perdue et al,1999)



Decision and Risk Analysis in Du Pont (Krumm and Rolle, 1992)



Polio Eradicators Decision Analysis (Thompson *et al*, 2015)



References

1. RA Howard and JE Matheson (1981). Influence Diagrams. In RA. Howard and J.E. Matheson (Editors). *Readings on the Principles and Applications of Decision Analysis*, pp 720 –762. This classical paper been republished in the journal *Decision Analysis* **2**(3):127-143, 2005.

2. RA Howard (1990). From influence to relevance to knowledge. In RM. Oliver and JQ Smith (editors), *Influence diagrams, Belief Nets and Decision Analysis*, pp 3-23, John Wiley.

3. RD Shachter (1986). Evaluating Influence Diagrams, *Operations Research* **34**(6):871-882.

4. L Helm (1996). Improbable Inspiration: The future of software may lie in the obscure theories of an 18th century cleric named Thomas Bayes. *Los Angeles Times*, October 28, 1996.

5. T Verma and J Pearl (1988). Causal Networks: Semantics and Expressiveness. In *Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence*.

6. D Geiger and J Pearl (1988). On the logic of causal models. In *Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence*.

7. J.S. Stonebraker (2002). How Bayer Makes Decisions to Develop New Drugs. *Interfaces* **32**(6):77-90.

8. R.K. Perdue, W.J. McAllister, P.V. King and B.G. Berkey (1999). Valuation of R and D Projects Using Options Pricing and Decision Analysis Models. *Interfaces* **29**(6):57-74.

9. F.V. Krumm, C.F. Roll (1992). Management and Application of Decision and Risk Analysis in Du Pont. *Interfaces* **22**(6):84-93.

10. K.M. Thompson, R.J.D Tebbens, M.A. Pallansch, S.G.F Wassilak and S.L Cochi (2015). Polio Eradicators Use Integrated Analytical Models to Make Better Decisions. *Interfaces* **45**(1):5-25

Books on Bayesian Network and Influence Diagram

1. J. Pearl (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kauffman Publishers.

2. E.Castillo, JM Gutierrez and AS Hadi (1997). *Expert Systems and Probabilistic Network Models*, Springer.

3. K.B. Korb and AE Nicholson (2004). *Bayesian Artificial Intelligence*, Chapman and Hall.

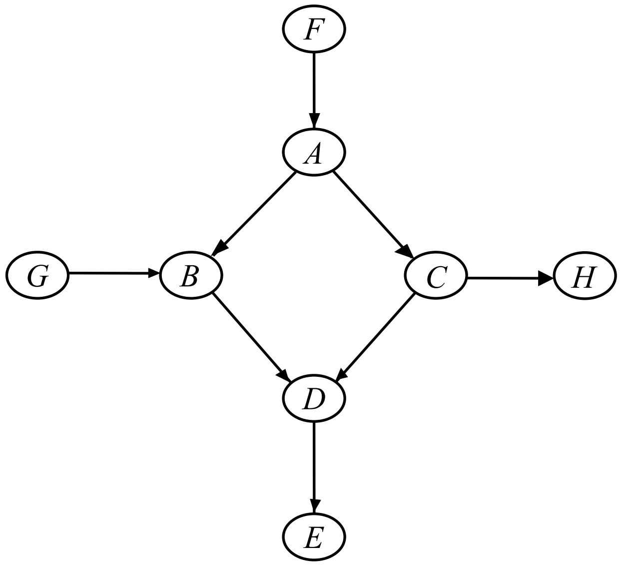
4. F.V. Jensen (2001). *Bayesian Networks and Decision Graphs*, Springer.

5. R.E. Neapolitan (2004). *Learning Bayesian Networks*, Prentice Hall.

6. M.I. Jordan (Editor) (1999). *Learning in Graphical Models*. MIT Press.

Exercises

P5.1 Given the following Bayesian network:



Use d-separation to determine whether each of the followings is true or false:

(***a***) $B \perp C \mid F$

(***b***) $B \perp C \mid A$

(***c***) $B \perp C \mid \{A, E\}$

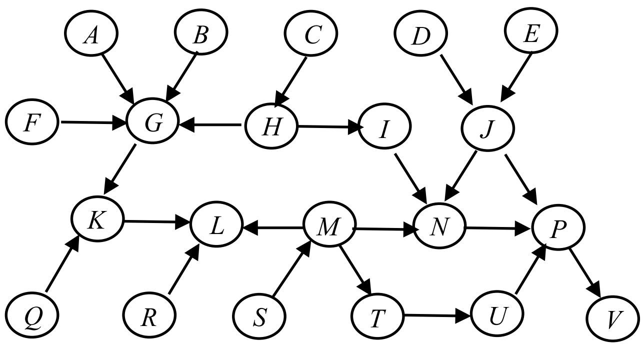
(***d***) $B \perp C \mid \{A, D\}$

(***e***) $A \perp D \mid \{B, G\}$

(***f***) $A \perp D \mid \{B, C\}$

(***g***) $E \perp F \mid \{B, C\}$

P5.2 Consider the following Bayesian network:



Indicate whether each of the following conditional independence statements is true or false:

(***a***) $\{A\} \perp \{V\} \mid \{ H, L, R \}$

(***b***) $\{A\} \perp \{V\} \mid \{ M, S, T \}$

P5.3 (Clement and Reilly 2001, Problem 3.9, p99)
A dapper young decision maker has just purchased a new suit for \$200. On the way out the door, he considers taking an umbrella. With the umbrella on hand, the suit will be protected in the event of rain. Without the umbrella, the suit will be ruined if it rains. On the other hand, if it does not rain, carrying the umbrella is an unnecessary inconvenience.

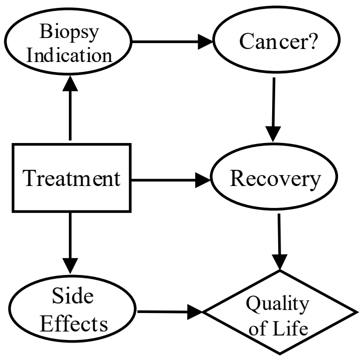
(***a***) Draw a decision tree of this situation.

(***b***) Draw an influence diagram of the situation.

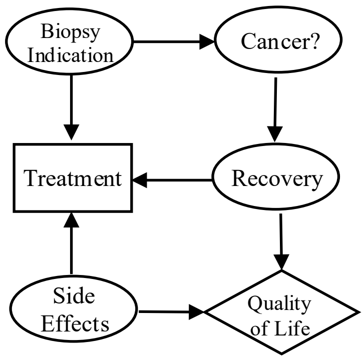
(***c***) Before deciding, the decision maker considers listening to the weather forecast on the radio. Draw an influence diagram that takes into account the weather forecast.

P5.4 You are helping a friend who might have cancer. She has just had a needle biopsy taken and will be meeting with her doctor when the laboratory report becomes available to choose a treatment. Her concerns in making the decision is whether she recovers from any cancer and some serious side effects associated with some of the treatment choices. Which of these influence diagrams correctly and best captures her situation?

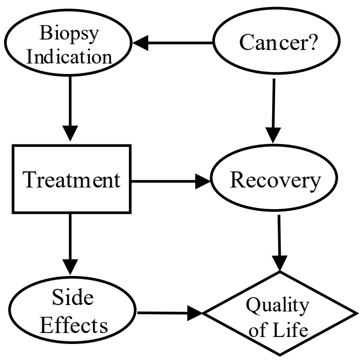
(***a***)



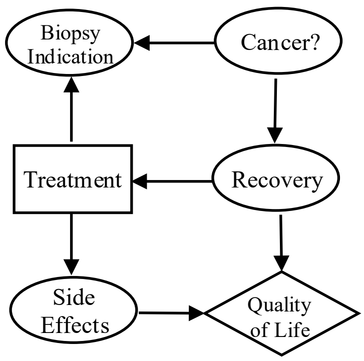
(***b***)



(***c***)



(***d***)



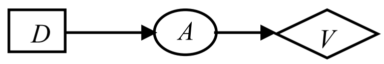
P5.5 (Clement and Reilly 2001, Problem 310, p99)

When patients suffered from hemorrhagic fever, M*A*S*H doctors replaced lost sodium by administering a saline solution intravenously. However, headquarters (HQ) sent a treatment change disallowing the saline solution. With a patient in shock and near death from a disastrously low sodium level, B.J. Hunnicut wanted to administer a low-sodium concentration saline solution as a last ditch attempt to save the patient. Colonel Potter looked at B.J. and Hawkeye and summed up the situation: "OK, let's get this straight. If we go by the new directive from HQ and don't administer saline to replace the sodium, our boy will die for sure. If we try B.J's idea, then he may survive, and we'll know how to treat the next two patients who are getting worse. If we try it and he doesn't make it, we're in trouble with HQ and may get court-martialed. I say we have no choice. Let's try it". (Source: Mr. and Mrs. Who" written by Ronny Graham, directed by Burt Metealfe, 1980)

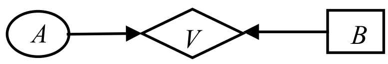
Structure the doctors' decision. What are their objectives? What risks do they face? Draw a decision tree and an influence diagram for their situation.

P5.6 For each of the influence diagram below, draw an equivalent decision tree.

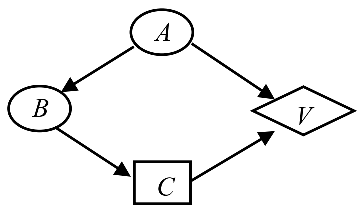
(***a***)



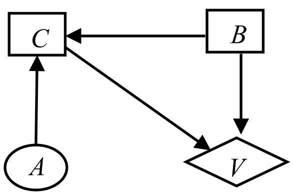
(***b***)



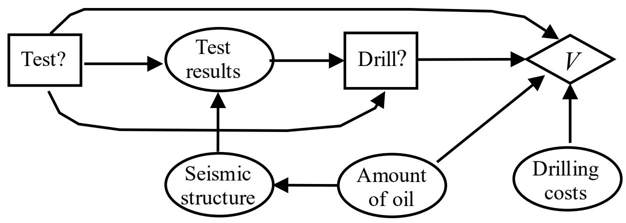
(***c***)



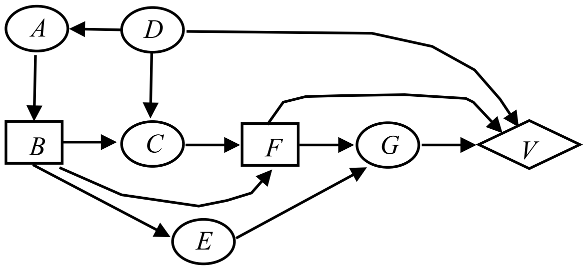
(***d***)



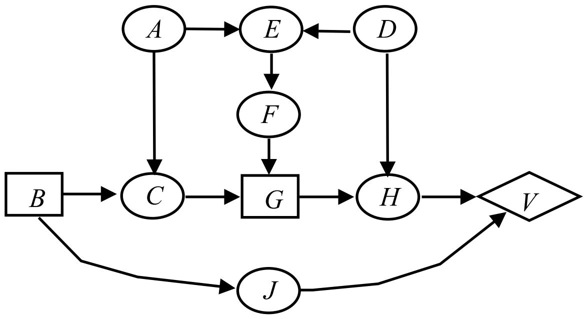
(*e*) The oil cluttering problem



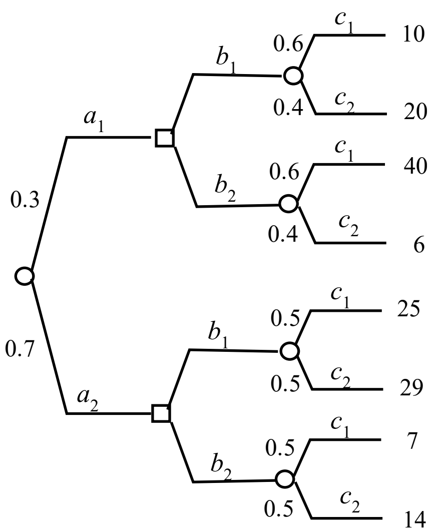
(*f*)



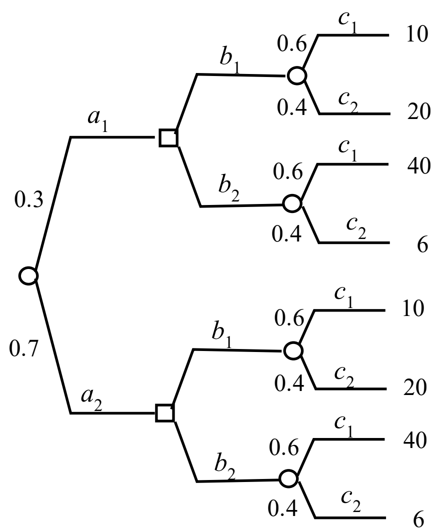
(*g*)



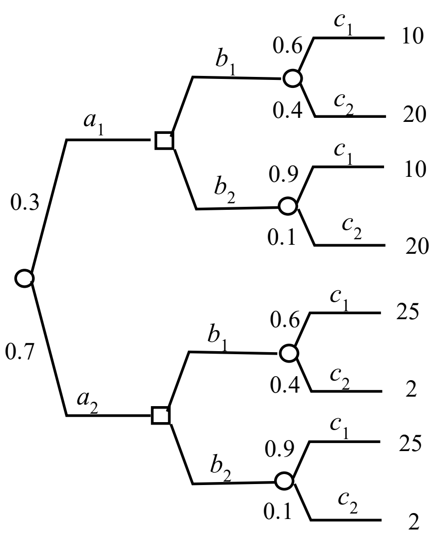
P5.7 For each of the following decision trees, draw an equivalent influence diagram representing the decision situation depicted by the tree. Include only those nodes and arcs that are necessary.



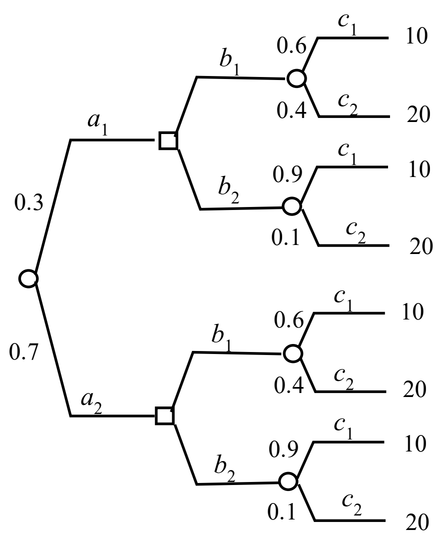
(*a*)



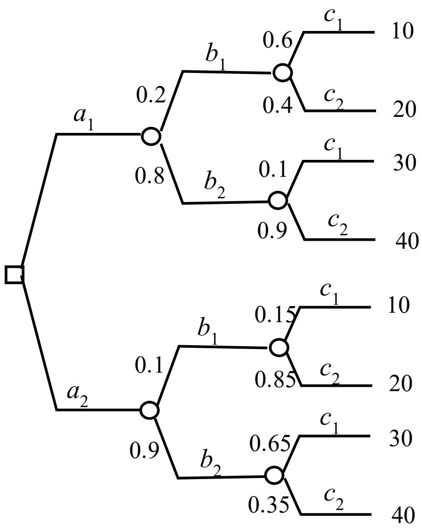
(*b*)



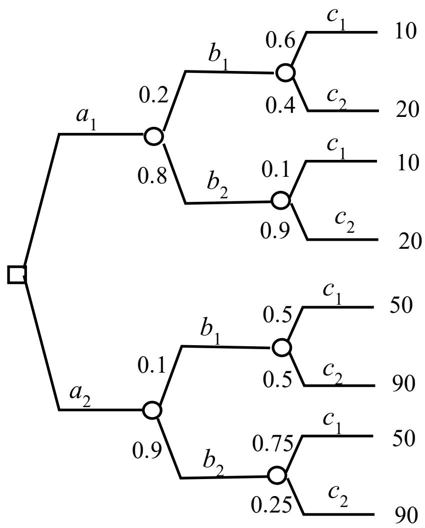
(*c*)



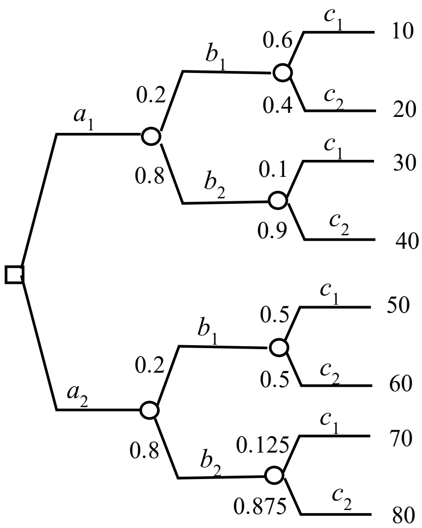
(*d*)



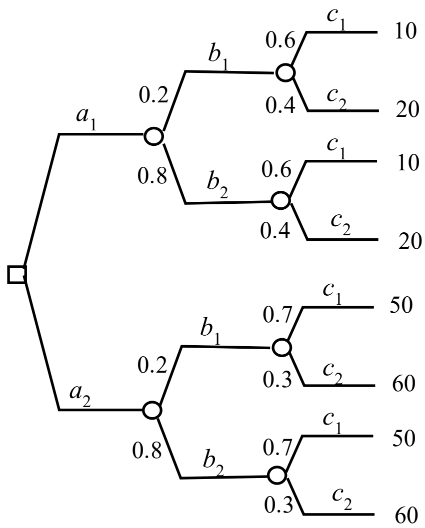
(*e*)



(*f*)



(*g*)



(*h*)

Optional Problem

P5.8 Solve Question 5.5 again using the Algorithm in Appendix A to automatically generate a sequence of nodes for the decision tree and specify the conditional probabilities stored in the tree.

Appendix A: Auto-Generating Decision Trees from Influence Diagrams

5A.1 Converting an Influence Diagram into a Decision Tree

- In Section 5.3.2, we learn how to construct a decision tree from an influence diagram by making sure that the decision tree satisfies a set of conditions.

- Suppose we want to automatically convert an influence diagram into an equivalent decision tree without human intervention, how should we do it?

- The Algorithm to automatically convert an ID into a decision tree is as follows:

  Step:

  1. If the ID is not a *decision network*, convert it by adding chorological and no-forget arcs.

  2. Perform arc reversal operations until the ID is a *Decision-Tree Network*.

  3. Find a valid sequence of nodes for the tree.

  4. Draw the Tree and copy the conditional probabilities directly from the ID to the Tree.

5A.2 Decision-Tree Networks

In Step 2 of the Algorithm, we ensure that the ID is a Decision-Tree Network.

Definition

- An Influence Diagram is said to be a **Decision-Tree Network** if it satisfies the following conditions:

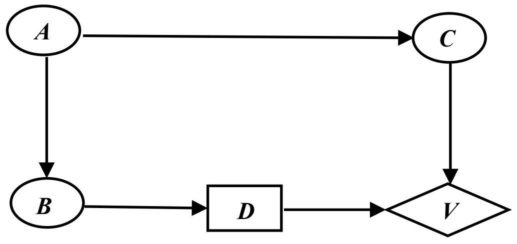
  1. It is a *Decision Network*.

  2. If there are chance nodes preceding a decision node, they must all be parent nodes. (Note that chance node can be both a parent as well as a grand-parent of a decision node).

- An influence diagram that satisfy the Decision-Tree Network conditions allows us to generate a sequence of nodes for drawing an equivalent tree and all the conditional probabilities required for the tree can be copied directly from the influence diagram. This property is exploited in Step 3 of the algorithm.

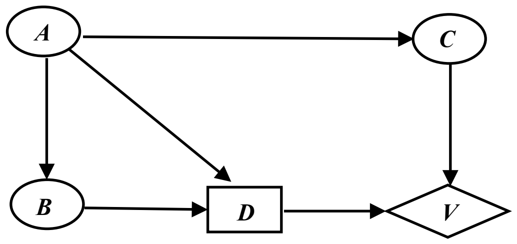
Example (Not a decision-tree network)

- The influence diagram below **is not** a decision tree network because chance node A is predecessor node of decision node D but is not a direct parent of D.



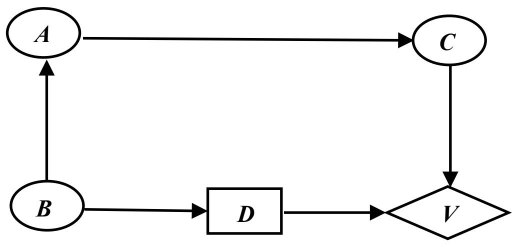
Example (A decision-tree network)

- The influence diagram below **is** a decision tree network because although the chance node A is predecessor node of decision node D, it is also a parent of D.



Example (A decision-tree network)

- The influence diagram below **is** a decision tree network as the decision node D does not have any predecessor nodes that are not direct parent of D.

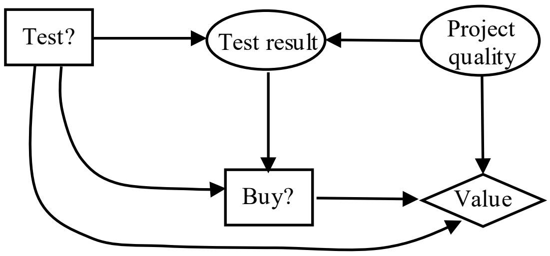


5A.3 Converting an Influence Diagram to a Decision-Tree Network by Arc Reversals

- An Influence Diagram that is already a Decision Network can be converted into a Decision-Tree Network by Arc Reversals.

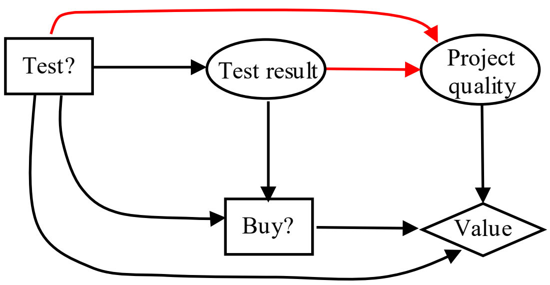
Example

- Draw an equivalent decision tree for the following influence diagram by first converting it into a decision tree network:



- The given influence diagram is already a decision network, but not yet a decision tree network. The chance node "Product Quality" is a predecessor of "Buy", but not a direct one.

- To convert it into a decision tree network, the arc between "Product quality" and "Test result" is reversed. Notice that an arc is added from "Test?" to "Product quality" as a result of the reversal operation.



- To be continued.

5A.4 Generating a Sequence of Nodes for the Decision Tree

- Given an Influence Diagram that is already a Decision-Tree Network, we can automatically generate a sequence of nodes for constructing an equivalent Decision Tree.

Algorithm for Decision Tree Sequence Generation

1. Convert the influence diagram into a decision tree network if it is not already one.

2. Create a sequence comprising all the decision nodes in chronological order.

3. Insert the chance nodes into the sequence so that the followings are satisfied:

   (*a*) A node that is a parent node of a decision node should appear just before the decision node that first observed it.

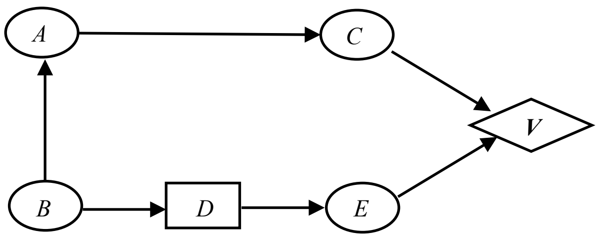
   (*b*) A node that is not a direct parent of any decision node should appear after the last decision node in the sequence.

   (*c*) Consistency with the partial ordering of the nodes in the network is maintained.

4. Insert the value node to the end of sequence.

Example

- Consider the following influence diagram.



Step 1: Influence diagram is already a decision-tree network since.

Step 2: Sequence = { D }

Step 3(a): B is a parent of D: Sequence = { B, D }

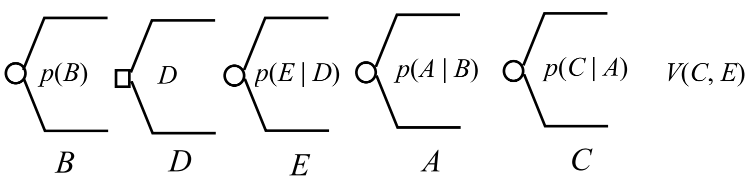
Step 3(b): A, C, and E are not parents of D: Sequence = { B, D, E, A, C }

The sequence { B, D, E, A, C } is consistent with the partial ordering in the network.

Step 4: Insert value node: Sequence = { B, D, E, A, C, V }

- Hence an equivalent decision tree can be drawn using the sequence { B, D, E, A, C, V }

- The generic tree with all the conditional probabilities is as follows:



- Note that all the conditional probabilities for the tree can be copied directly from the decision-tree network.

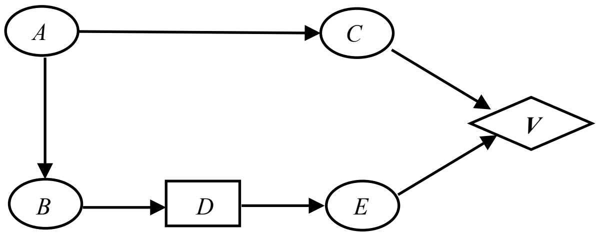
The Decision Tree is Non-Unique

- Notice that in the above example, the following sequences are also correct for drawing the decision tree:

  1. $\{ B, D, A, C, E, V \}$
  2. $\{ B, D, A, E, C, V \}$
  3. $\{ B, D, E, A, C, V \}$

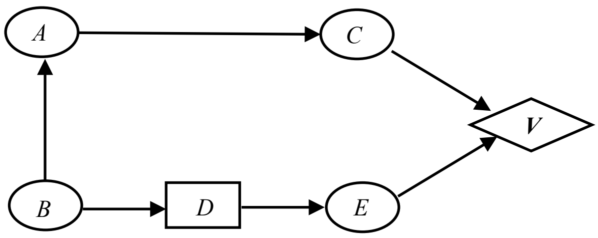
Example

- Consider now the following influence diagram.



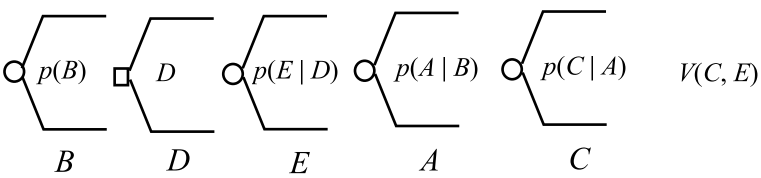
- It is not yet a decision tree network as chance node A is a predecessor of D, but is not a direct parent of D.

- It can be converted into a Decision Tree Network by reversing the arc between A and B:



- The sequence of nodes for the decision tree = $\{B, D, E, A, C, V\}$.

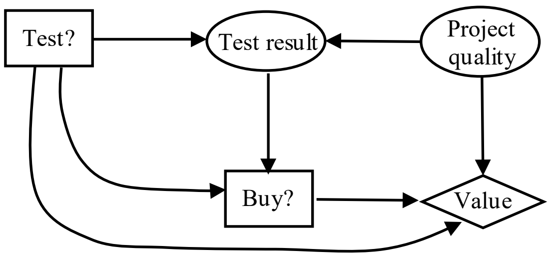
- The generic tree with all the conditional probabilities is as follows:



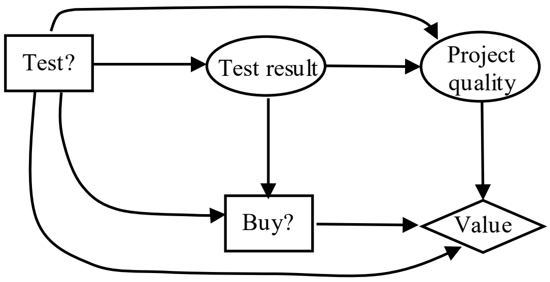
- The conditional probabilities for the tree can be copied directly from the decision-tree network.

Example

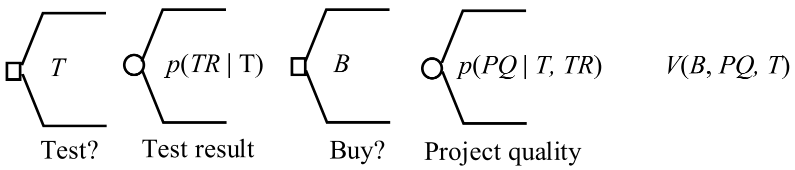
- Given the previous Influence Diagram:



- Converting it into a Decision-Tree Network:



- Sequence for the tree = { Test, Test Result, Buy, Product Quality, Value }
- The equivalent decision tree is:



- The conditional probability distributions *p*(Test result | Test) and *p*(Product Quality | Test, Test Result) can be directly copied from the decision tree network.

This page is blank.