Lecture 6.2

Bayesian Network Fundamentals

Machine Learning

Ivan Smetannikov

09.04.2021

Lecture plan

- Semantics and Factorization
- Reasoning Patterns
- Flow of Probabilistic influence

Lecture plan

- Semantics and Factorization
- Reasoning Patterns
- Flow of Probabilistic influence

Uncertainty

- Partial knowledge of state of the world
- Noisy observations
- Phenomena not covered by our model
- Inherent stochasticity

P (G,D,I,S,L)

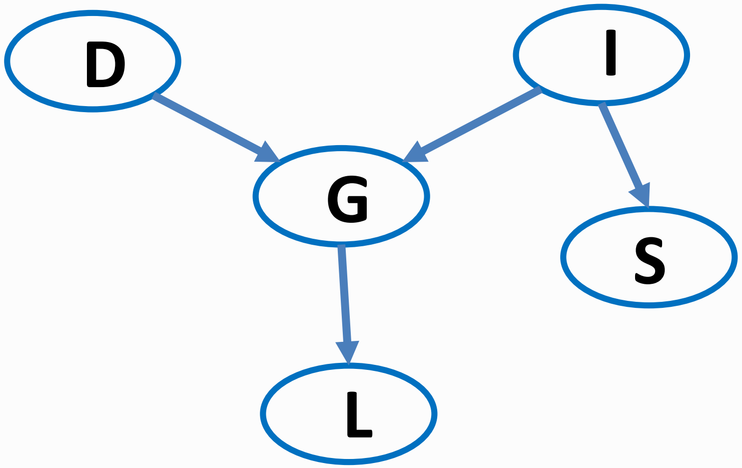
- **G**rade

P (G,D,I,S,L)

- **Grade**
- Course **D**ifficulty
- Student **I**ntelligence
- Student **S**AT
- Reference **L**etter

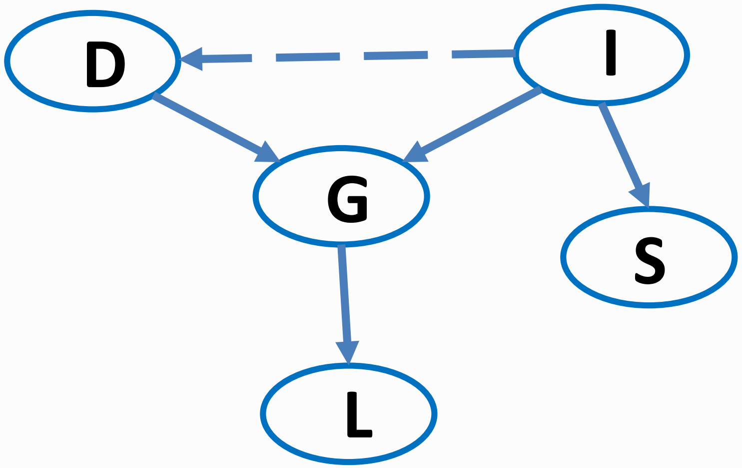
P (G,D,I,S,L)

- **G**rade
- Course **D**ifficulty
- Student **I**ntelligence
- Student **S**AT
- Reference **L**etter

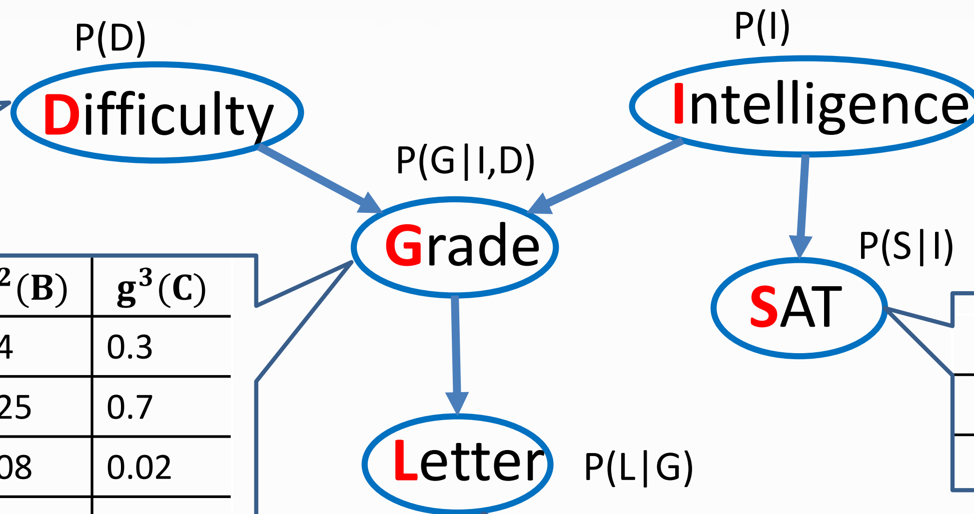


P (G,D,I,S,L)

- **G**rade
- Course **D**ifficulty
- Student **I**ntelligence
- Student **S**AT
- Reference **L**etter



CPD= Condition, Probably, Distribution



d^0	d^1
0.6	0.4

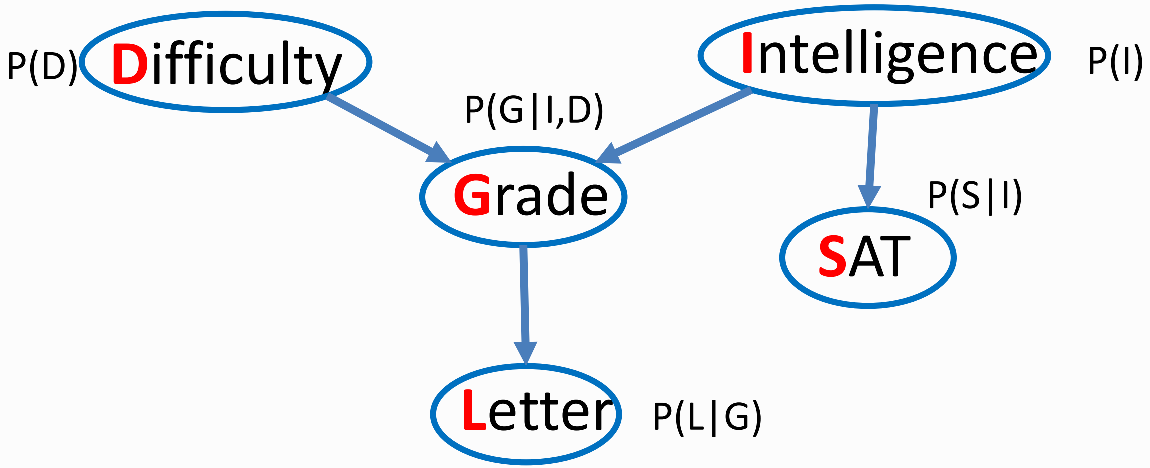
i^0	i^1
0.7	0.3

	g^1(A)	g^2(B)	g^3(C)
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

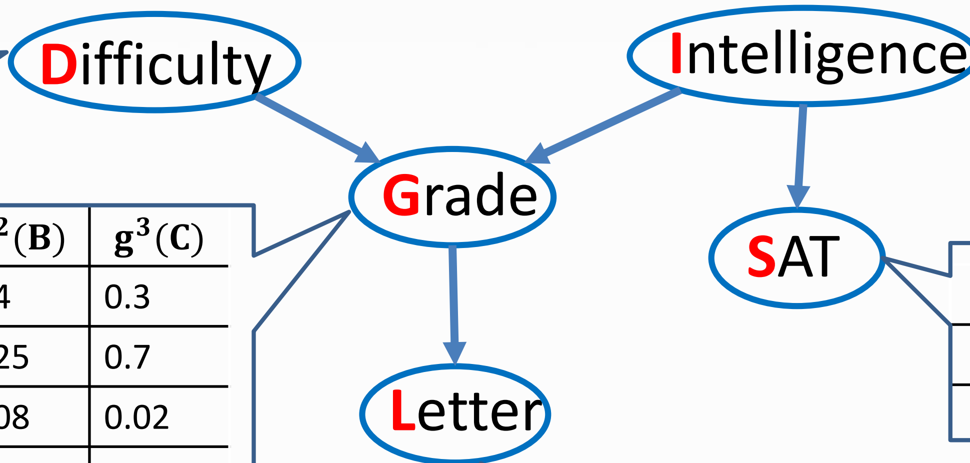
	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Chain Rule for Bayesian Networks



$$P(D,I,G,S,L)=P(D)P(I)P(G|I,D)\,P(S|I)\,P(L|G)$$

Distribution defined as a product



d^0	d^1
0.6	0.4

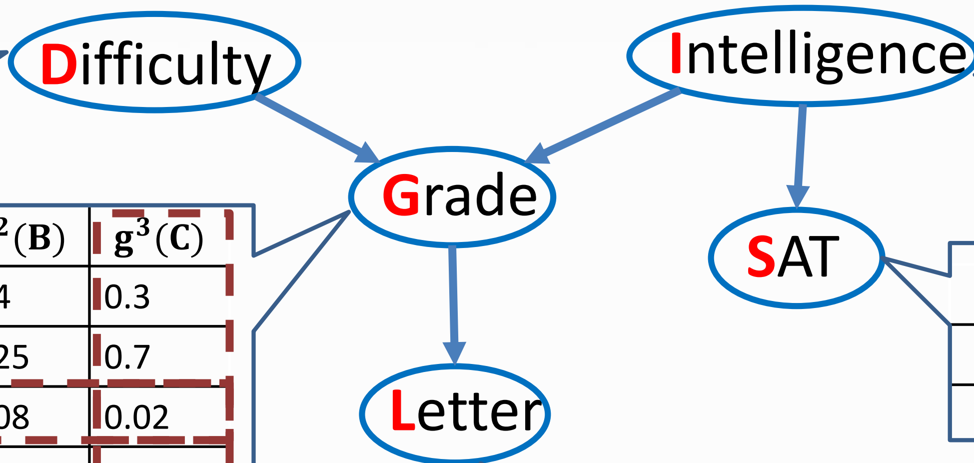
i^0	i^1
0.7	0.3

	g^1(A)	g^2(B)	g^3(C)
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

$P(d^0, i^1, g^3, s^1, l^1)=$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1(A)	g^2(B)	g^3(C)
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

$P(d^0, i^1, g^3, s^1, l^1)=$

$0.6 \times 0.3 \times 0.02 \times 0.01 \times 0.8$

Bayesian Network

- A Bayesian network is
  — A direct acyclic graph (DAG) G whose nodes represent the
  random variables $(X_1,\ldots, X_n)$
  —For each node X_i a CPD $P(X_i|Par_G(X_i))$

Bayesian Network

- A Bayesian network is

  — A direct acyclic graph (DAG) G whose nodes represent the

  random variables $(X_1,\dots, X_n)$

  —For each nose X_i a CPD $P(X_i|Par_G(X_i))$

- The BN represents a joint distribution

  via the chain rule for Bayesian networks

  $P(X_1,\dots, X_n) = \Pi\, P(X_i|Par_G(X_i))$

BN Is a Legal Distribution: $P \geq 0$

- P is a product of CPDs
- CPDs are non-negative

BN Is a Legal Distribution: ∑ P = 1

$$\sum_{D,I,G,S,L} P(D,I,G,S,L) = \sum_{D,I,G,S,L} \underbrace{P(D)P(I)P(G|I,D)P(S|I)P(L|G)}_{\text{chain rule}}$$

$$= \sum_{D,I,G,S} P(D)P(I)P(G|I,D)P(S|I) \underbrace{\cancel{\sum_L P(L|G)}}_{=1}$$

$$= \sum_{D,I,G,S} P(D)P(I)P(G|I,D)P(S|I)$$

$$= \sum_{D,I,G} P(D)P(I)P(G|I,D) \underbrace{\cancel{\sum_S P(S|I)}}_{=1} = \sum_{D,I} P(D)P(I) \sum_G P(G|I,D)$$

$$= \cdots$$

P Factorizes over G

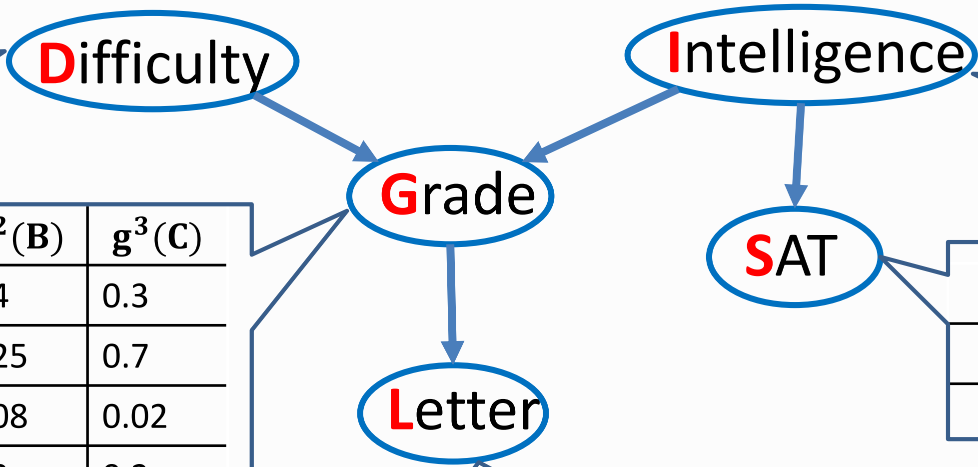
- Let G be a graph over $X_1,\ldots, X_n$.
- P factorizes over G if

$$P(X_1,\ldots, X_n) = \underbrace{\prod_i P(X_i \mid Par_G(X_i))}_{\text{chain rule}}$$

Lecture plan

- Semantics and Factorization
- Reasoning Patterns
- Flow of Probabilistic influence

The Student Network



d^0	d^1
0.6	0.4

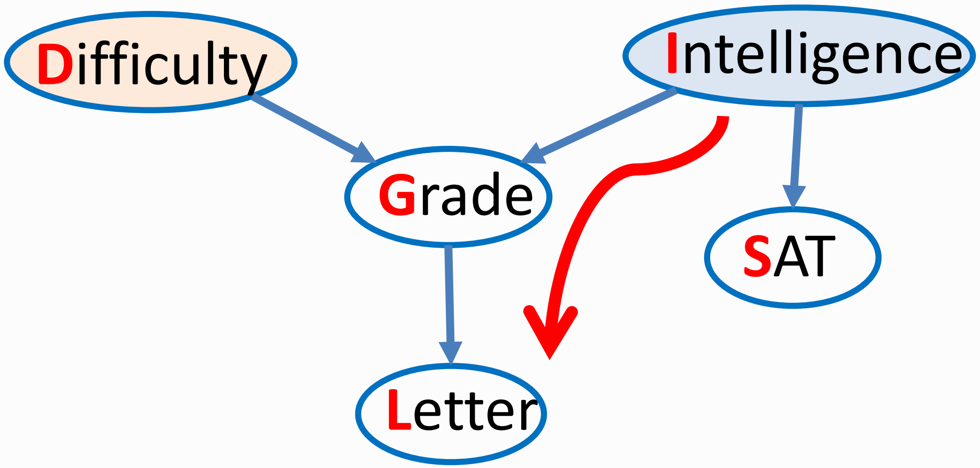
i^0	i^1
0.7	0.3

	g^1(A)	g^2(B)	g^3(C)
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Causal Reasoning

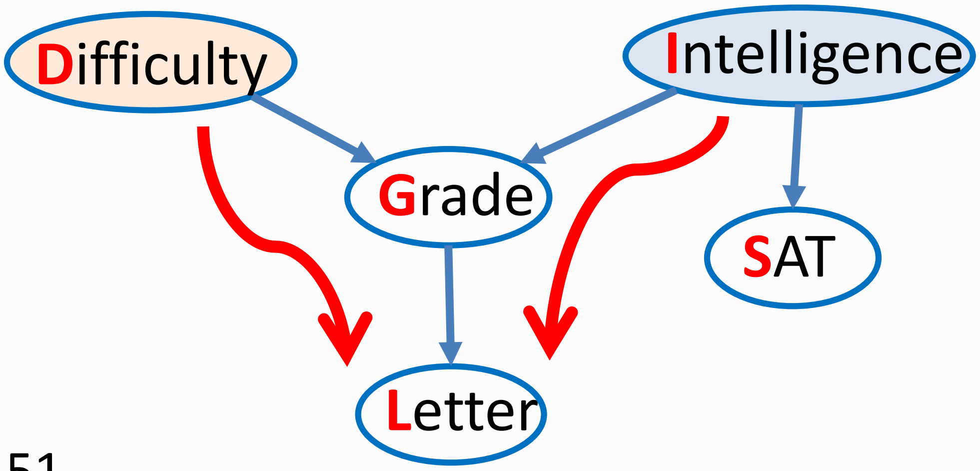


P(l^1)=0.5

P($l^1|i^0$)=0.39

P($l^1|i^0, d^0$)=?

Causal Reasoning



P(l^1)=0.5

P($l^1|i^0$)=0.39

P($l^1|i^0, d^0$)=0.51

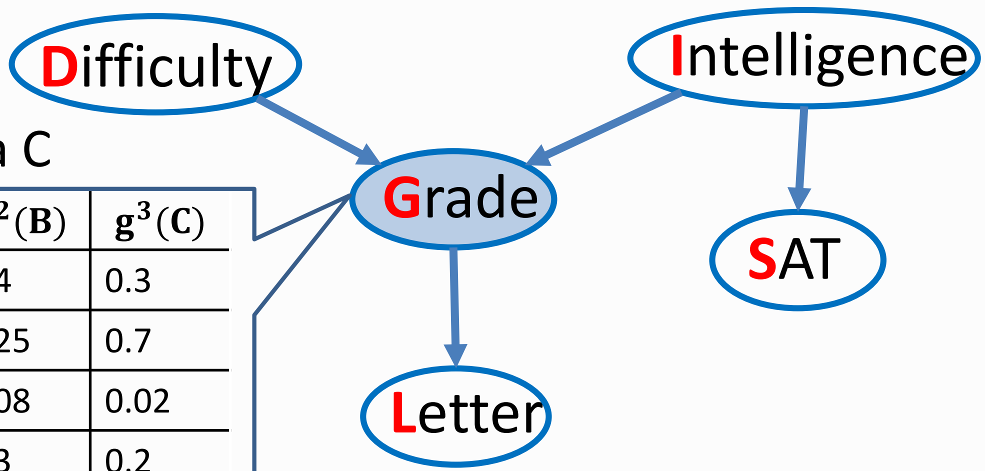
Intercausal Reasoning

$P(d^1)=0.4$

$P(d^1|g^3)=?$

$P(i^1)=0.3$

$P(i^1|g^3)=?$



Student gets a C

	$\mathbf{g^1(A)}$	$\mathbf{g^2(B)}$	$\mathbf{g^3(C)}$
$\mathbf{i^0, d^0}$	0.3	0.4	0.3
$\mathbf{i^0, d^1}$	0.05	0.25	0.7
$\mathbf{i^1, d^0}$	0.9	0.08	0.02
$\mathbf{i^1, d^1}$	0.5	0.3	0.2

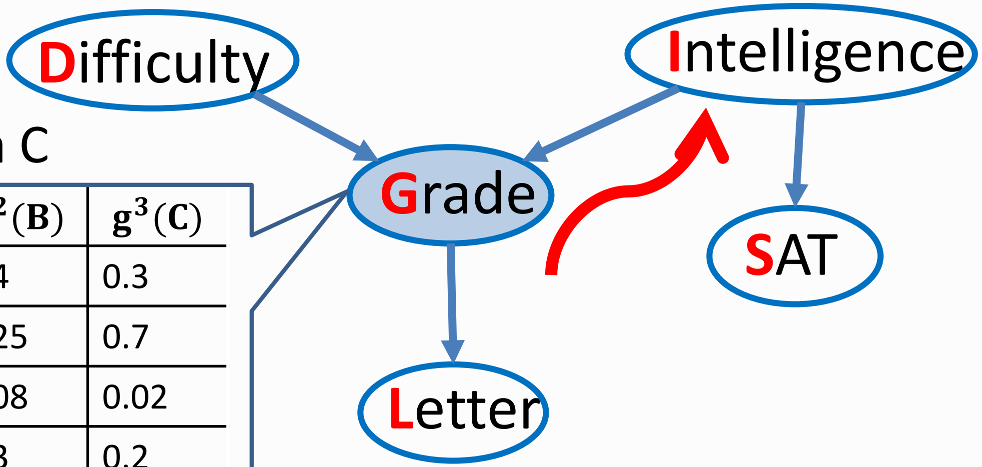
Intercausal Reasoning

P(d^1)=0.4

P($d^1|g^3$)=0.63

P(i^1)=0.3

P($i^1|g^3$)=0.08



Student gets a C

	$\mathbf{g^1}$(A)	$\mathbf{g^2}$(B)	$\mathbf{g^3}$(C)
$\mathbf{i^0, d^0}$	0.3	0.4	0.3
$\mathbf{i^0, d^1}$	0.05	0.25	0.7
$\mathbf{i^1, d^0}$	0.9	0.08	0.02
$\mathbf{i^1, d^1}$	0.5	0.3	0.2

Intercausal Reasoning

P(d^1)=0.4

P($d^1|g^3$)=0.63

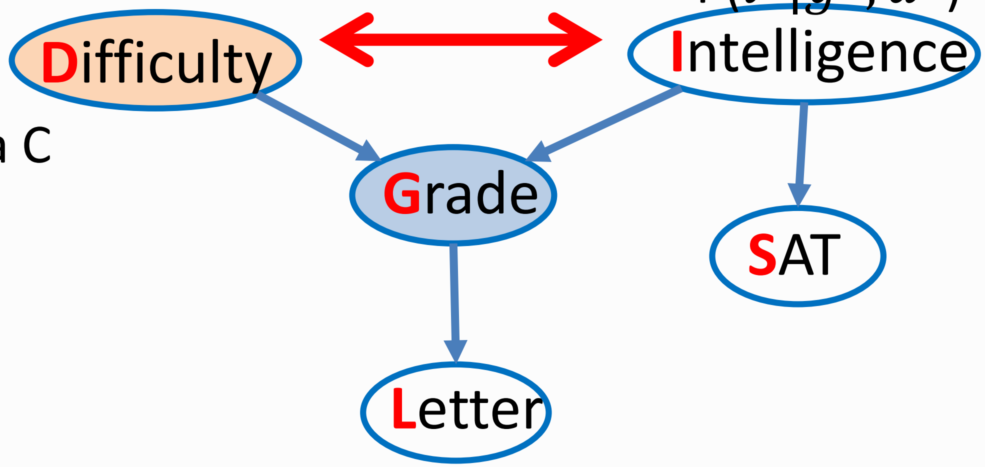
P(i^1)=0.3

P($i^1|g^3$)=0.08

P($i^1|g^3,d^1$)=?

Class is hard

Student gets a C



Intercausal Reasoning

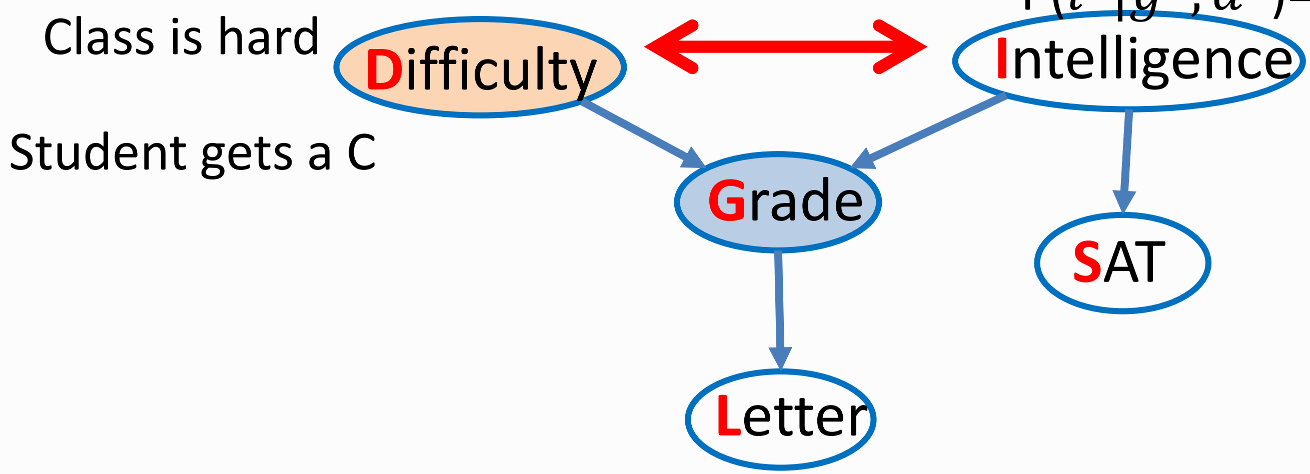
P(d^1)=0.4

P($d^1|g^3$)=0.63

P(i^1)=0.3

P($i^1|g^3$)=0.08

P($i^1|g^3, d^1$)=0.11

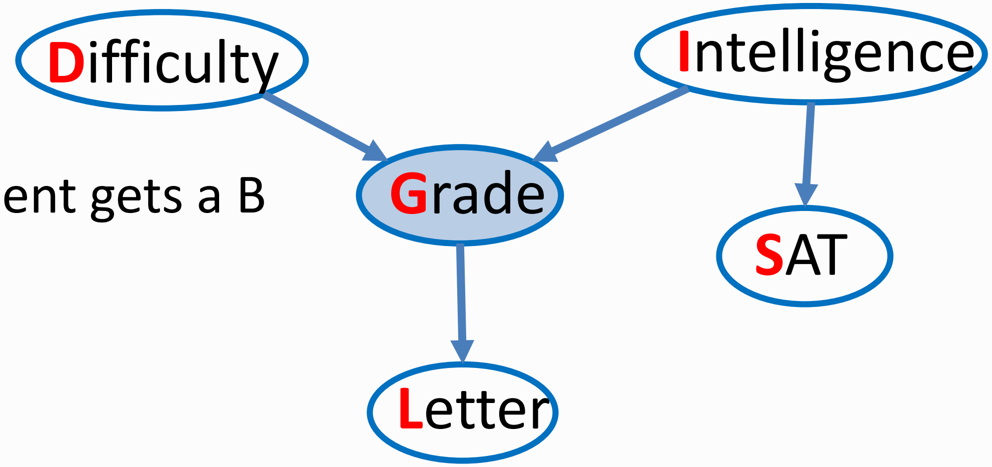


Intercausal Reasoning

P(i^1)=0.3

P($i^1|g^2$)=0.175

Student gets a B

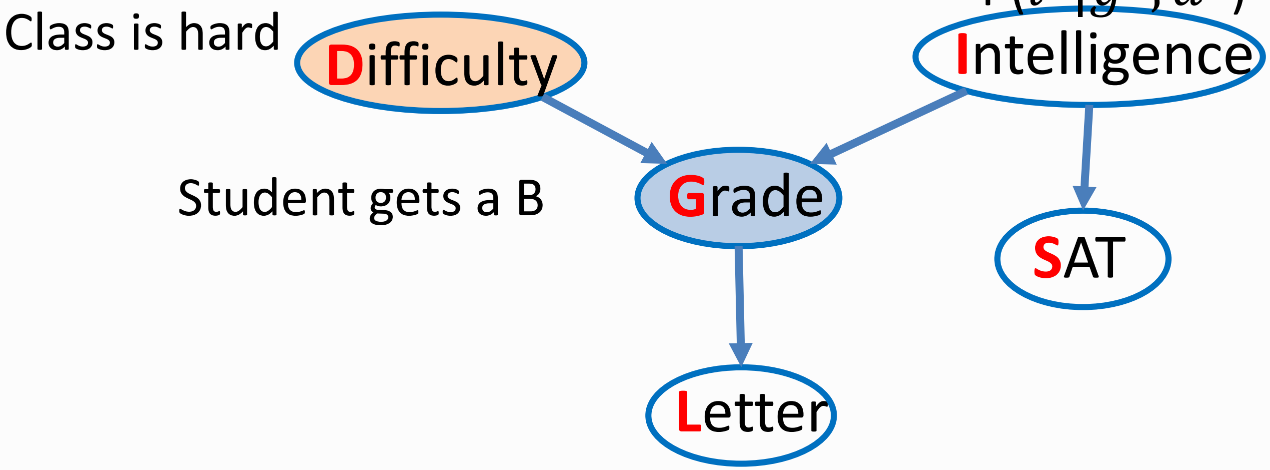


Intercausal Reasoning

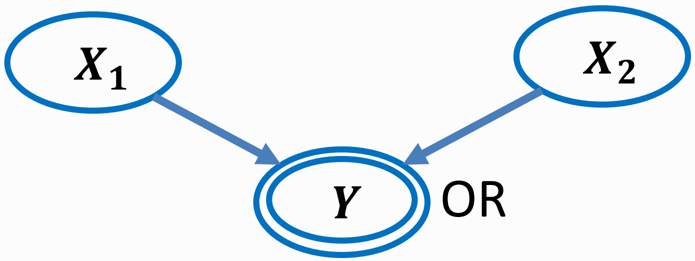
P(i^1)=0.3

P($i^1|g^2$)=0.175

P($i^1|g^2, d^1$)=0.34 > 0.3

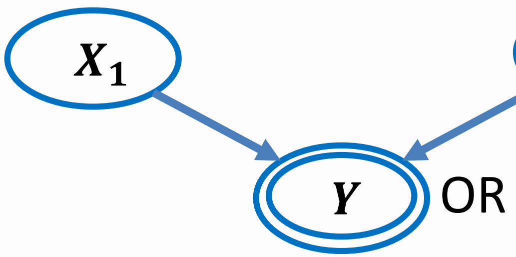


Intercausal Reasoning Explained



X_1	X_2	Y	Prob.
0	0	0	0.25
0	1	1	0.25
1	0	1	0.25
1	1	1	0.25

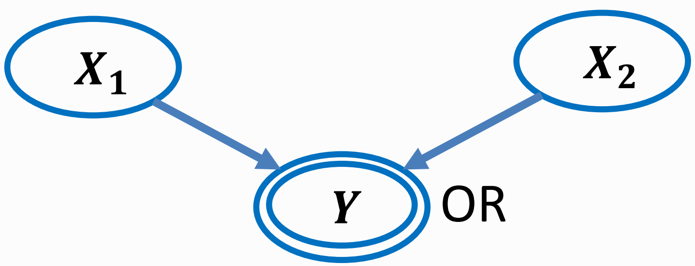
Intercausal Reasoning Explained



Evidence: $Y = 1$

X_1	X_2	Y	Prob.
0	0	0	0.25
0	1	1	0.25
1	0	1	0.25
1	1	1	0.25

Intercausal Reasoning Explained

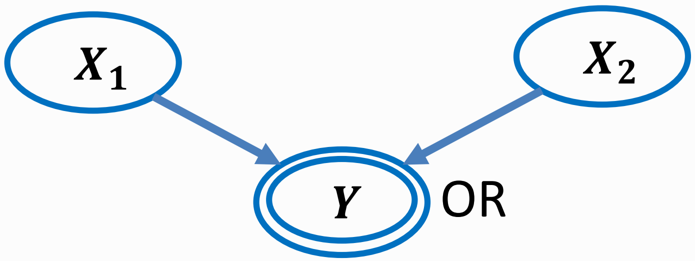


Evidence: $Y = 1$

X_1	X_2	Y	Prob.	
~~0~~	~~0~~	~~0~~	~~0.25~~	
0	1	1	0.25	0.33
1	0	1	0.25	0.33
1	1	1	0.25	0.33

$P(X_1 = 1) = \frac{2}{3}$ $\quad$ $P(X_2 = 1) = \frac{2}{3}$

Intercausal Reasoning Explained



Evidence: $Y = 1$

Explaining away:

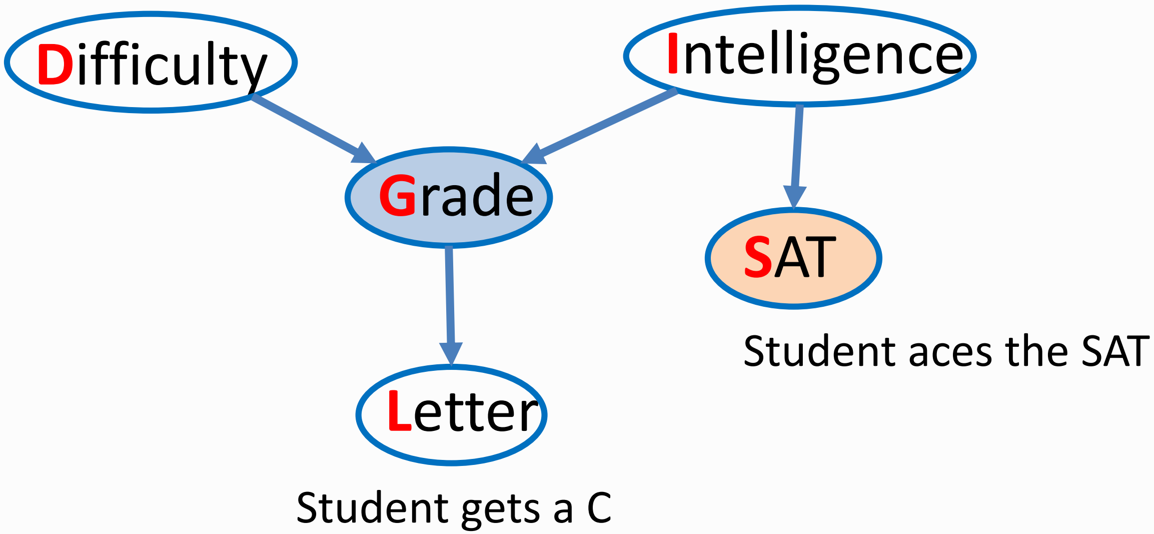
X_1	X_2	Y	Prob.		
~~0~~	~~0~~	~~0~~	~~0.25~~		
~~0~~	~~1~~	~~1~~	~~0.25~~	0.33	
1	0	1	0.25	0.33	**0.5**
1	1	1	0.25	0.33	**0.5**

$P(X_1 = 1) = \frac{2}{3}$ $\quad P(X_2 = 1) = \frac{2}{3}$

Condition $X_1 = 1$: $\quad \boldsymbol{P(X_2 = 1 | X_1 = 1) = 0.5}$

Student Aces the SAT

- What happens to the posterior probability that the class is hard?



Student Aces the SAT

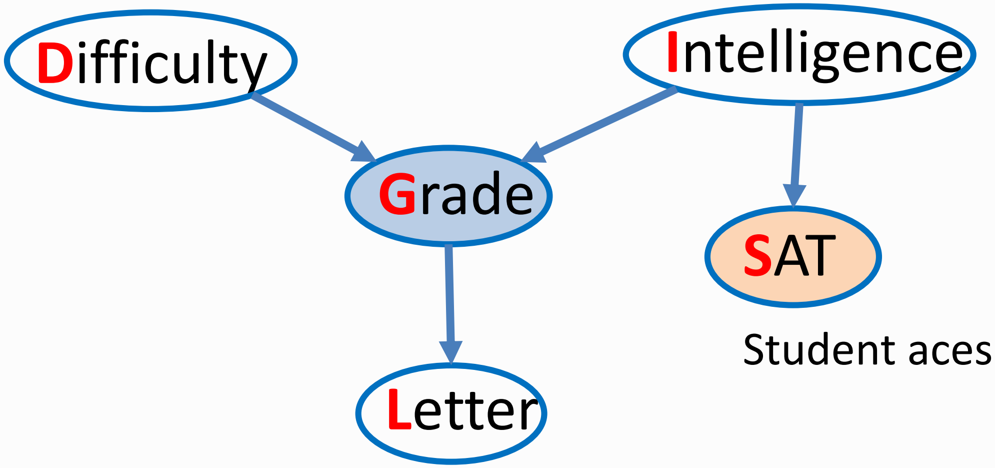
P(d^1)=0.4

P($d^1|g^3$)=0.63

P(i^1)=0.3

P($i^1|g^3$)=0.08

P($i^1|g^3, s^1$)=?



Student aces the SAT

Student gets a C

Student Aces the SAT

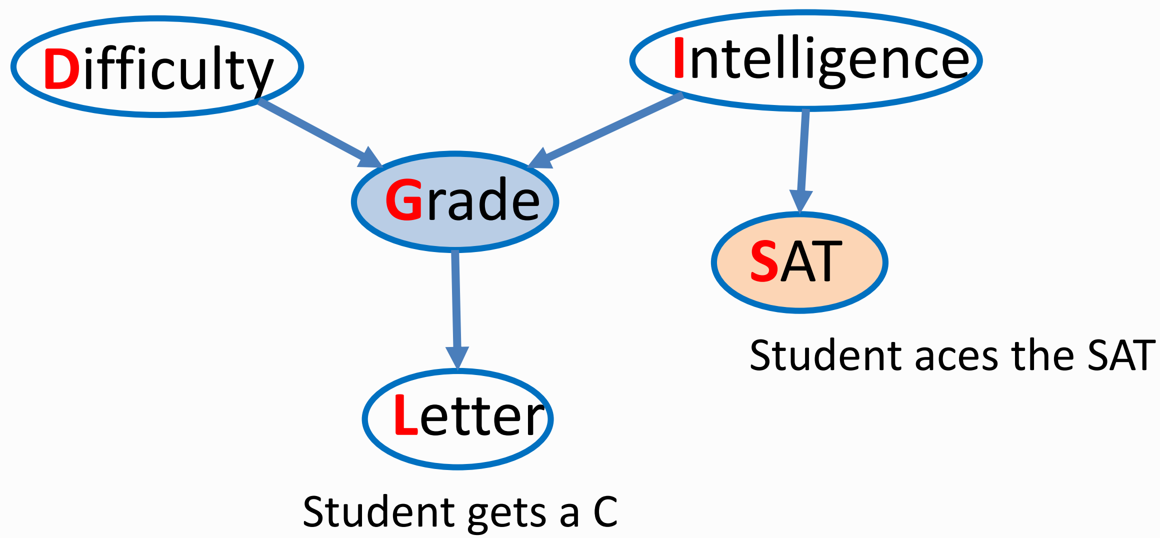
$P(d^1)=0.4$

$P(d^1|g^3)=0.63$

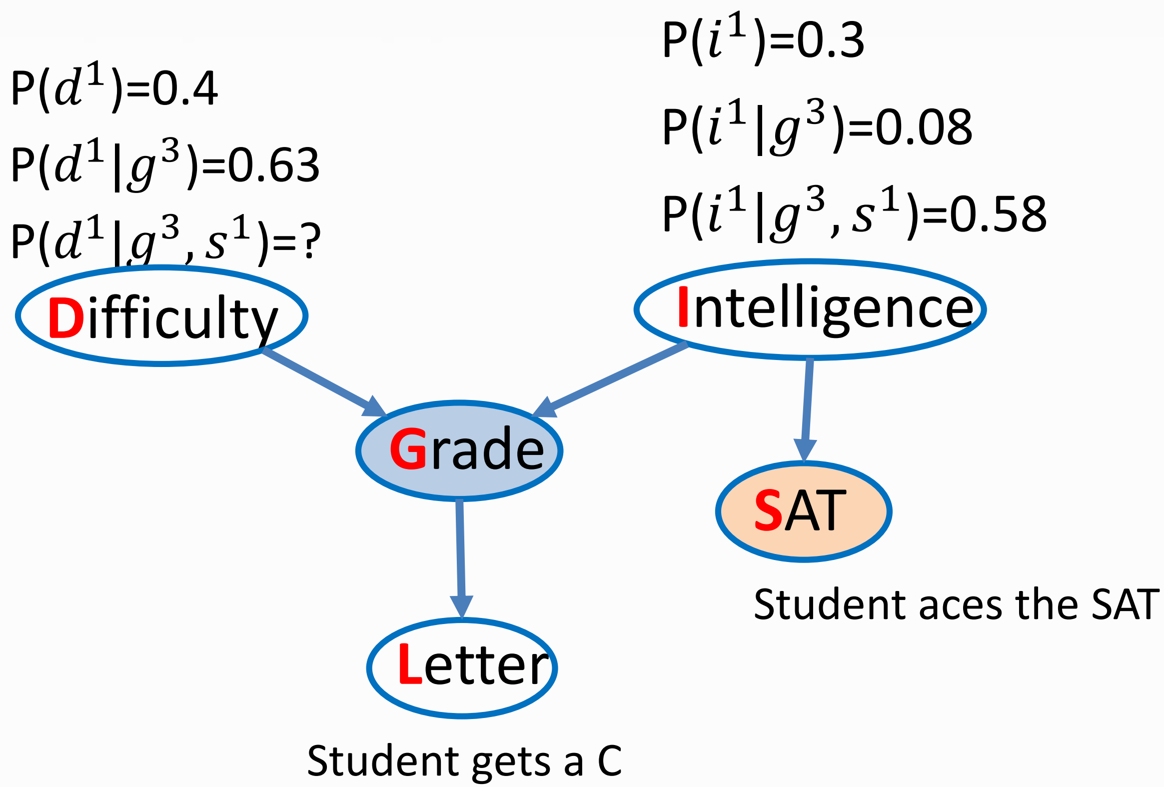
$P(i^1)=0.3$

$P(i^1|g^3)=0.08$

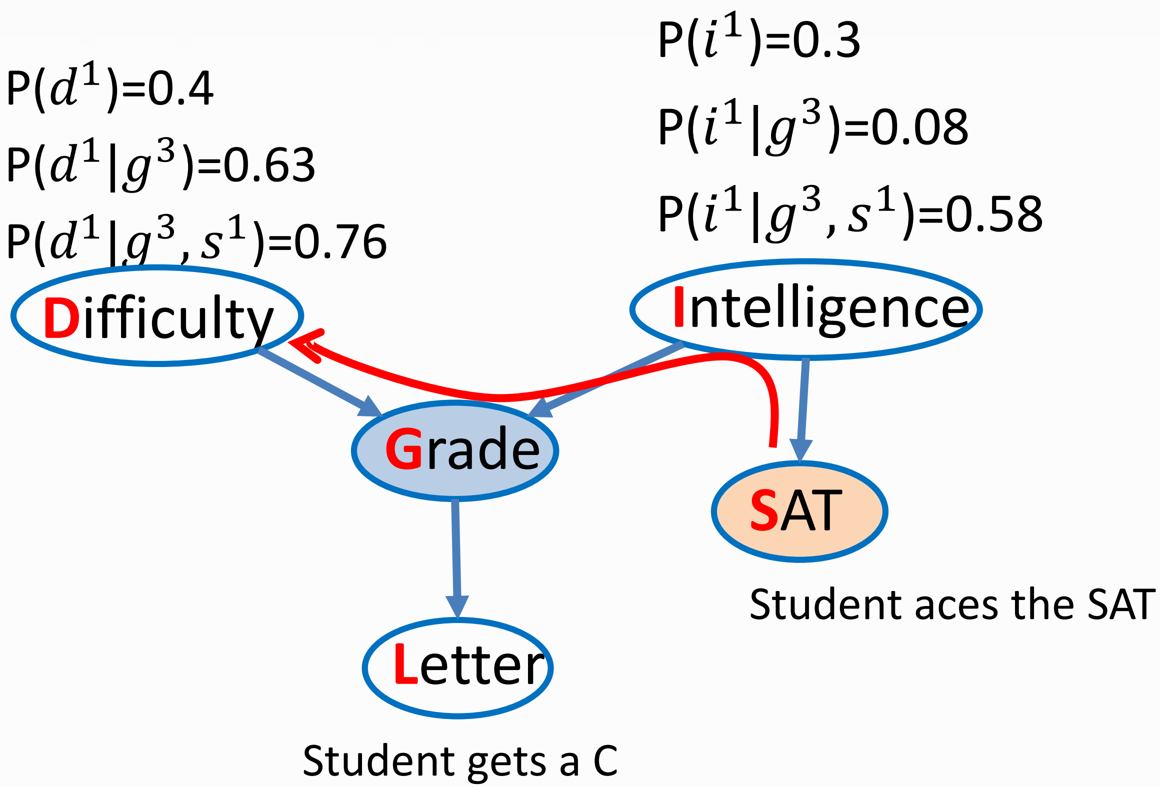
$P(i^1|g^3,s^1)=0.58$



Student Aces the SAT



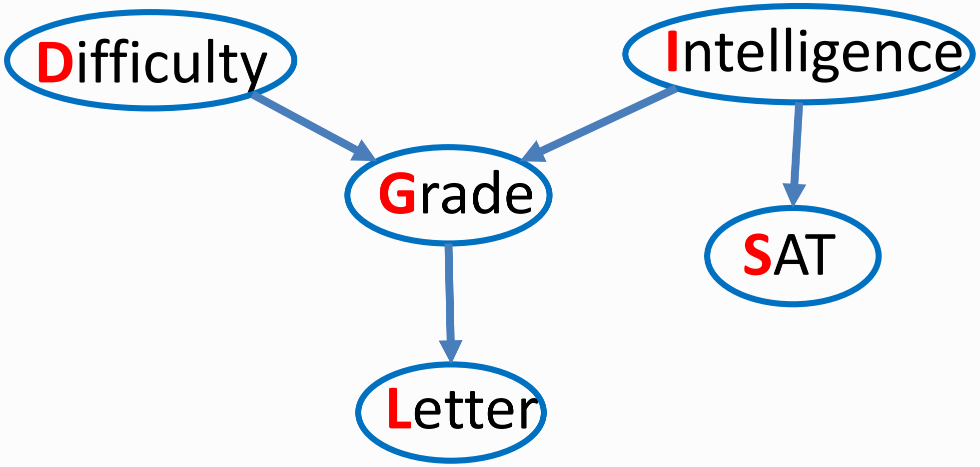
Student Aces the SAT



Lecture plan

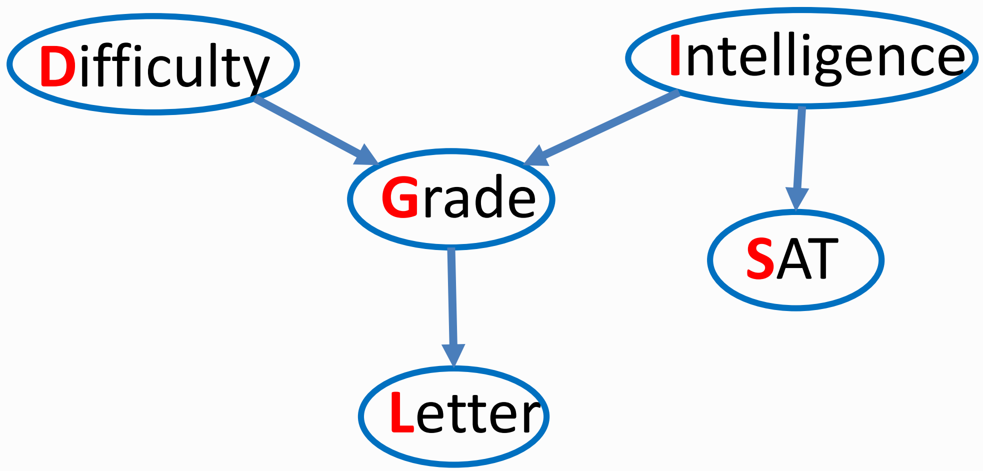
- Semantics and Factorization
- Reasoning Patterns
- Flow of Probabilistic influence

When can X influence Y?



When can X influence Y?

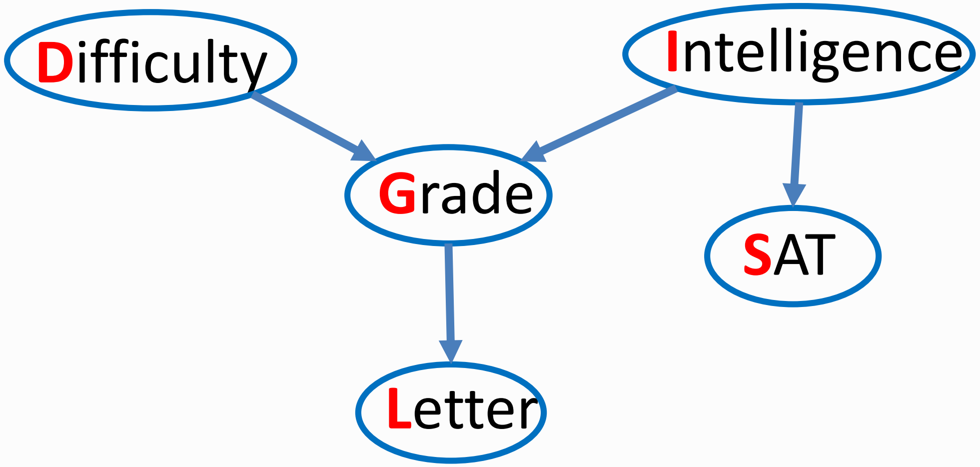
X Influence Y = X can change probability distribution of Y



When can X influence Y?

X influence Y = X can change probability distribution of Y

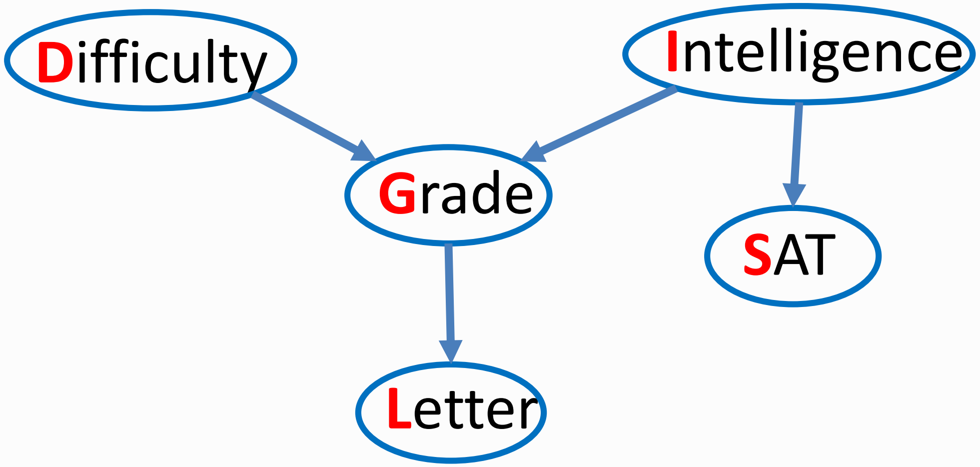
- $X \rightarrow Y$
- $X \leftarrow Y$



When can X influence Y?

X influence Y = X can change probability distribution of Y

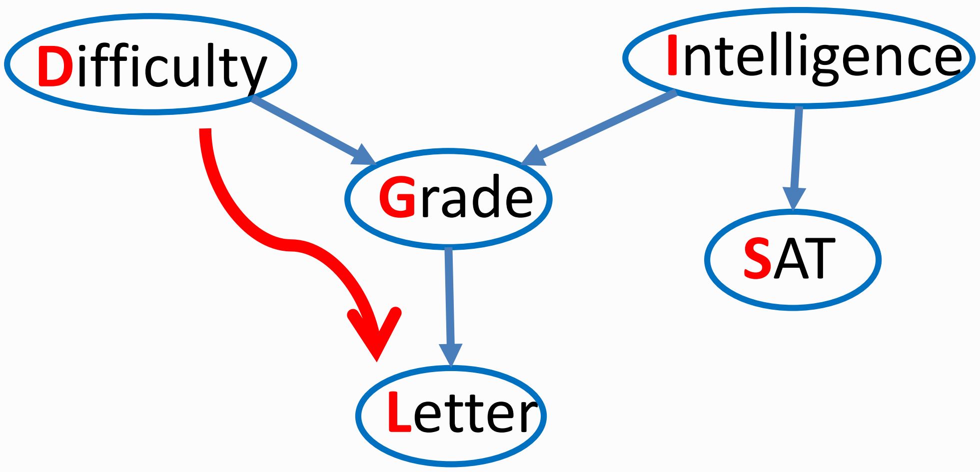
- $X \rightarrow Y$ ✓
- $X \leftarrow Y$ ✓



When can X influence Y?

X influence Y = X can change probability distribution of Y

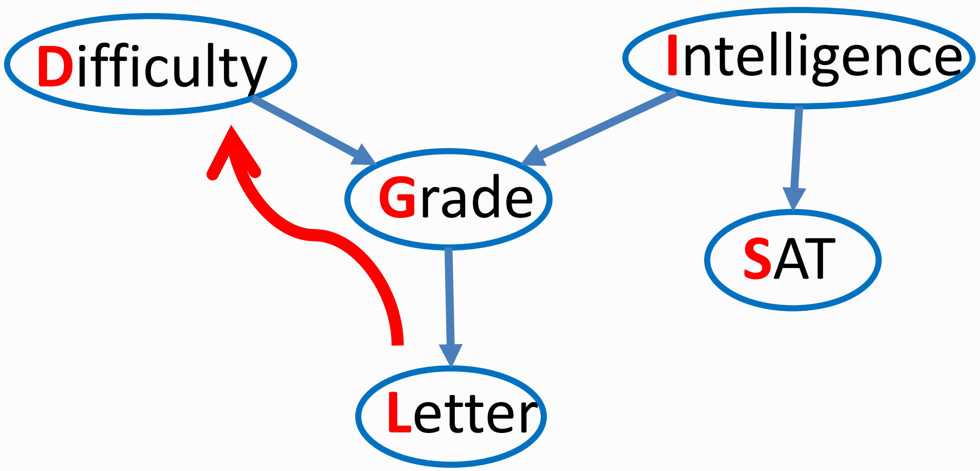
- $X \rightarrow Y$
- $X \leftarrow Y$
- $X \rightarrow W \rightarrow Y$



When can X influence Y?

X influence Y = X can change probability distribution of Y

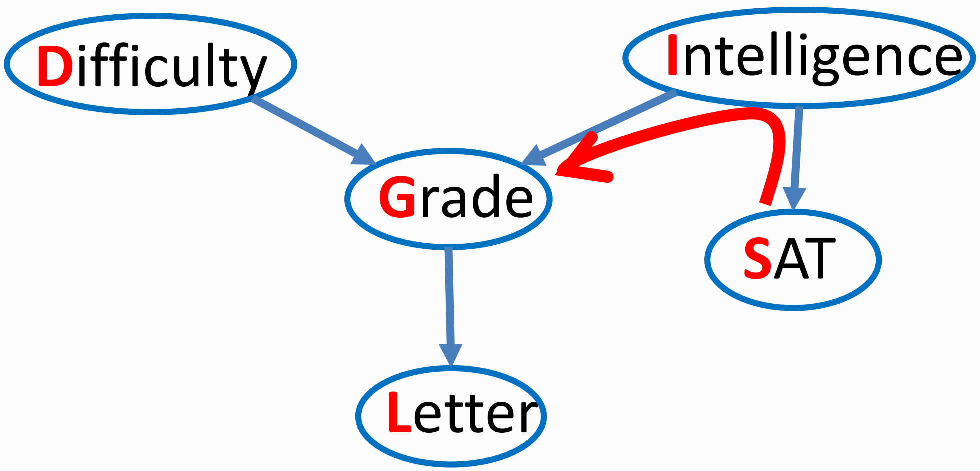
- $X \rightarrow Y$ ✓
- $X \leftarrow Y$ ✓
- $X \rightarrow W \rightarrow Y$ ✓
- $X \leftarrow W \leftarrow Y$



When can X influence Y?

X influence Y = X can change probability distribution of Y

- $X \rightarrow Y$ ✓
- $X \leftarrow Y$ ✓
- $X \rightarrow W \rightarrow Y$ ✓
- $X \leftarrow W \leftarrow Y$ ✓
- $X \leftarrow W \rightarrow Y$

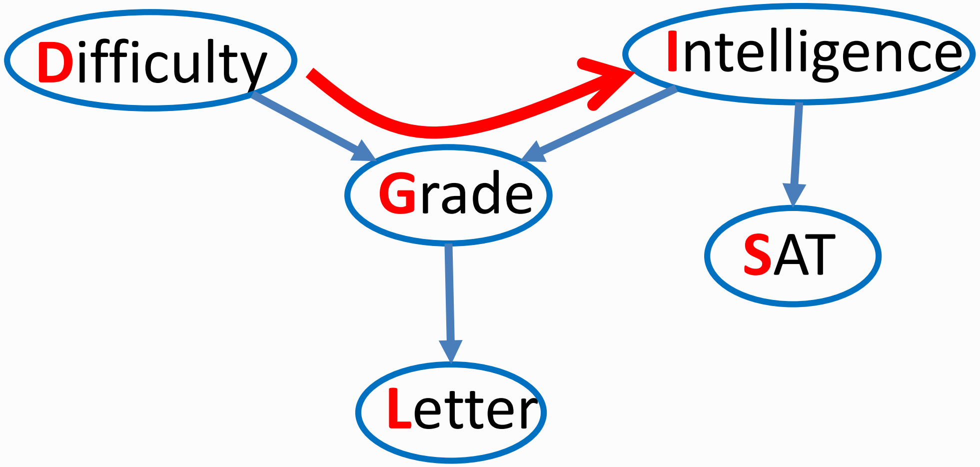


When can X influence Y?

X influence Y = X can change probability distribution of Y

- $X \rightarrow Y$ ✓
- $X \leftarrow Y$ ✓
- $X \rightarrow W \rightarrow Y$ ✓
- $X \leftarrow W \leftarrow Y$ ✓
- $X \leftarrow W \rightarrow Y$ ✓
- $X \rightarrow W \leftarrow Y$

V-structure

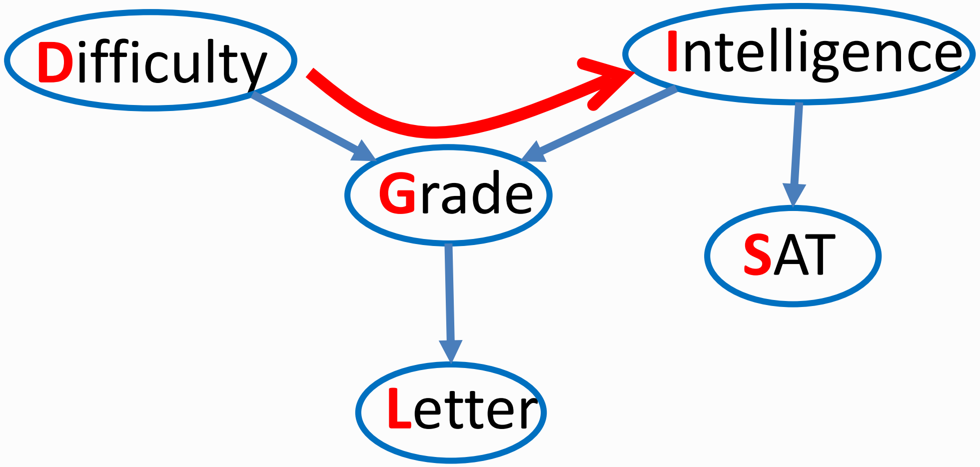


When can X influence Y?

X influence Y = X can change probability distribution of Y

- $X \rightarrow Y$ ✓
- $X \leftarrow Y$ ✓
- $X \rightarrow W \rightarrow Y$ ✓
- $X \leftarrow W \leftarrow Y$ ✓
- $X \leftarrow W \rightarrow Y$ ✓
- $X \rightarrow W \leftarrow Y$ ✗

V-structure



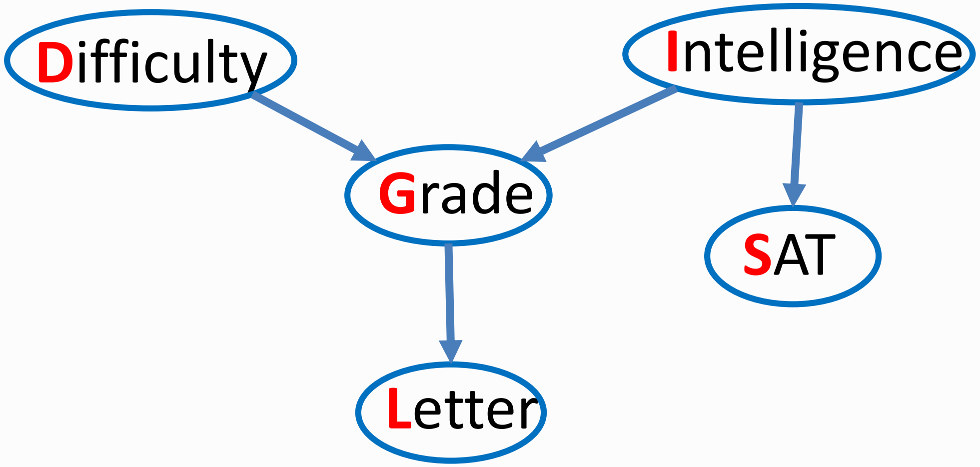
Active Trails

A trail $\mathrm{X}_1 - \cdots - \mathrm{X}_\mathrm{n}$ is active if

it has no v-structures $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$

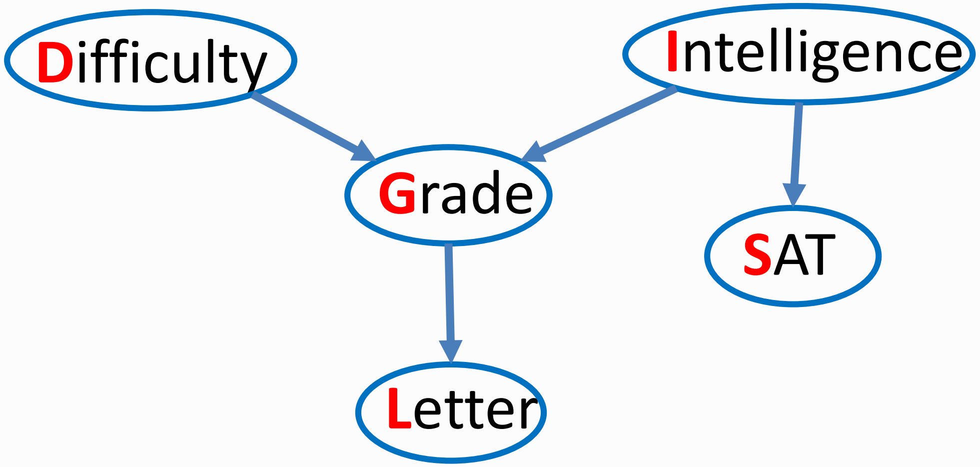
When can X influence Y Given evidence about Z?

- $X \rightarrow Y$
- $X \leftarrow Y$



When can X influence Y Given evidence about Z?

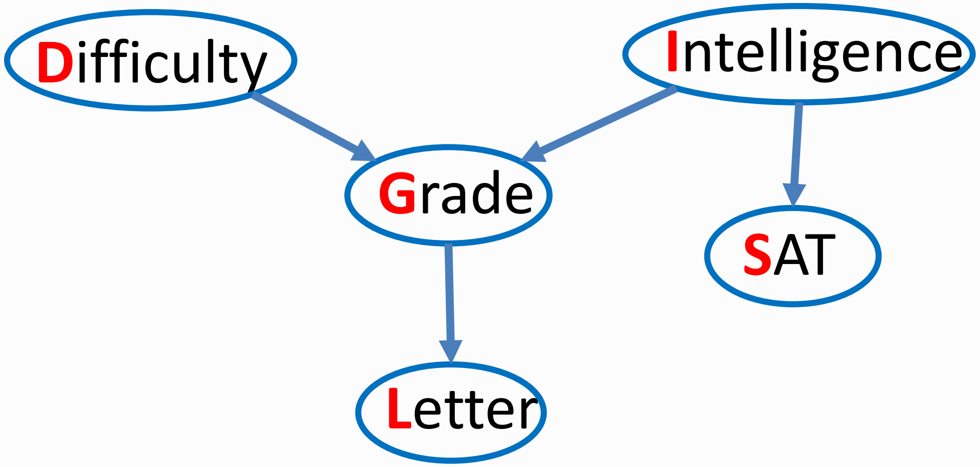
- $X \rightarrow Y$ ✓
- $X \leftarrow Y$ ✓
- $X \rightarrow W \rightarrow Y$
- $X \leftarrow W \leftarrow Y$
- $X \leftarrow W \rightarrow Y$
- $X \rightarrow W \leftarrow Y$



When can X influence Y Given evidence about Z?

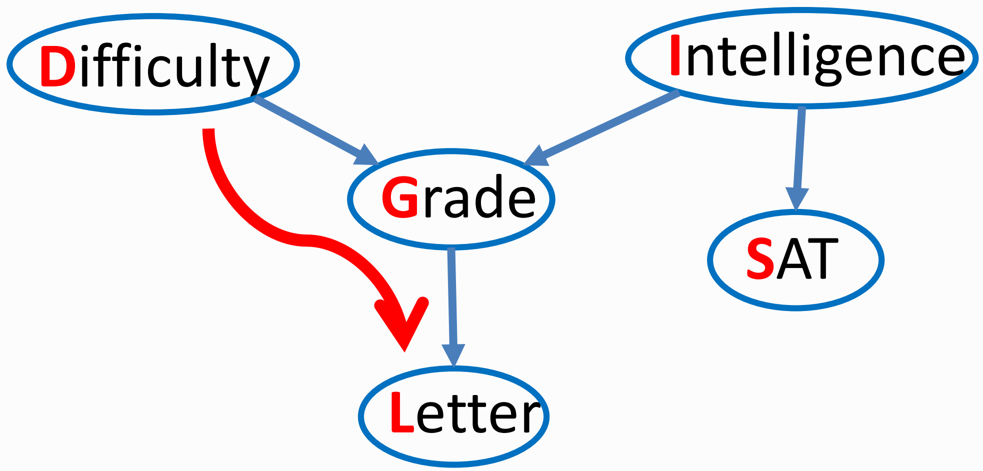
- $X \rightarrow Y$ ✓
- $X \leftarrow Y$ ✓

	$W \notin Z$	$W \in Z$
$X \rightarrow W \rightarrow Y$	???	
$X \leftarrow W \leftarrow Y$	???	
$X \leftarrow W \rightarrow Y$	???	
$X \rightarrow W \leftarrow Y$	???	



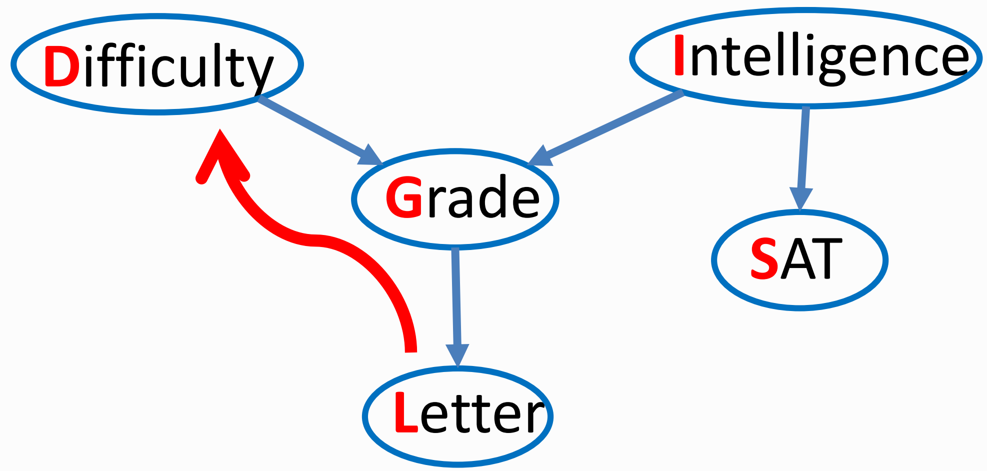
When can X influence Y Given evidence about Z?

	$W \notin Z$	$W \in Z$
$X \rightarrow Y$ ✓		
$X \leftarrow Y$ ✓		
$X \rightarrow W \rightarrow Y$	✓	???
$X \leftarrow W \leftarrow Y$	✓	
$X \leftarrow W \rightarrow Y$	✓	
$X \rightarrow W \leftarrow Y$		



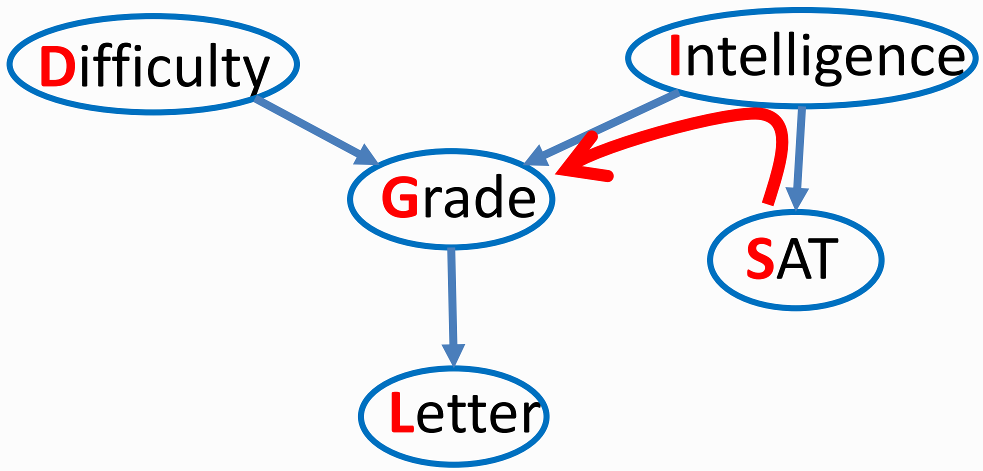
When can X influence Y Given evidence about Z?

	$W \notin Z$	$W \in Z$
• $X \to Y$ ✓		
• $X \leftarrow Y$ ✓		
• $X \to W \to Y$	✓	✗
• $X \leftarrow W \leftarrow Y$	✓	???
• $X \leftarrow W \to Y$	✓	
• $X \to W \leftarrow Y$		



When can X influence Y Given evidence about Z?

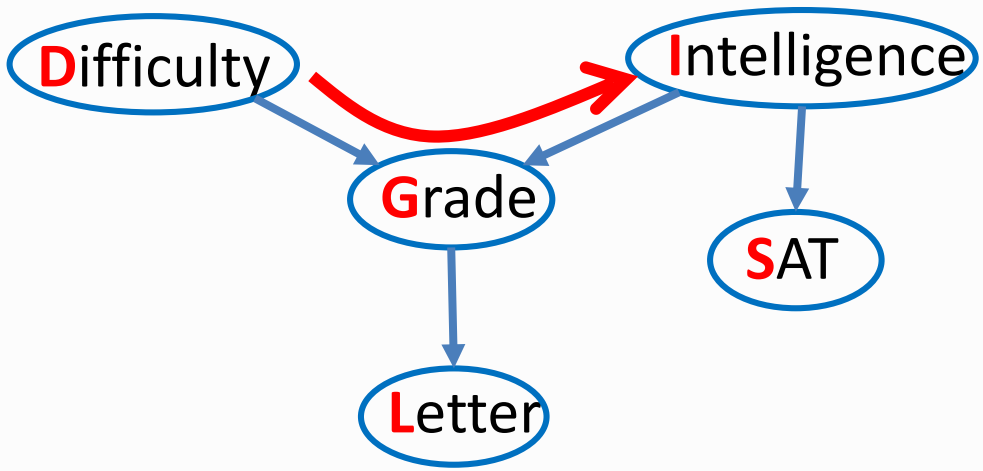
	$W \notin Z$	$W \in Z$
$X \rightarrow Y$ ✓		
$X \leftarrow Y$ ✓		
$X \rightarrow W \rightarrow Y$	✓	✗
$X \leftarrow W \leftarrow Y$	✓	✗
$X \leftarrow W \rightarrow Y$	✓	???
$X \rightarrow W \leftarrow Y$		



When can X influence Y Given evidence about Z?

- $X \rightarrow Y$ ✓
- $X \leftarrow Y$ ✓

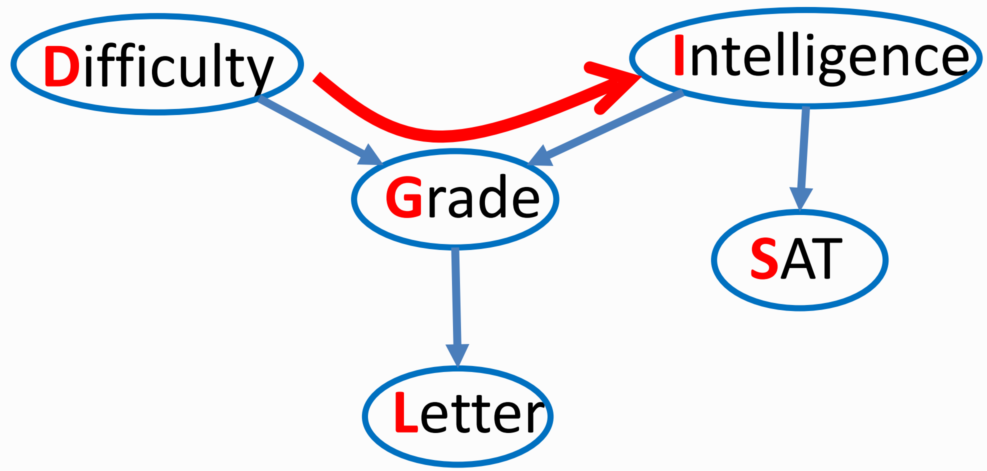
	$W \notin Z$	$W \in Z$
$X \rightarrow W \rightarrow Y$	✓	✗
$X \leftarrow W \leftarrow Y$	✓	✗
$X \leftarrow W \rightarrow Y$	✓	✗
$X \rightarrow W \leftarrow Y$		???



When can X influence Y Given evidence about Z?

- $X \rightarrow Y$ ✓
- $X \leftarrow Y$ ✓

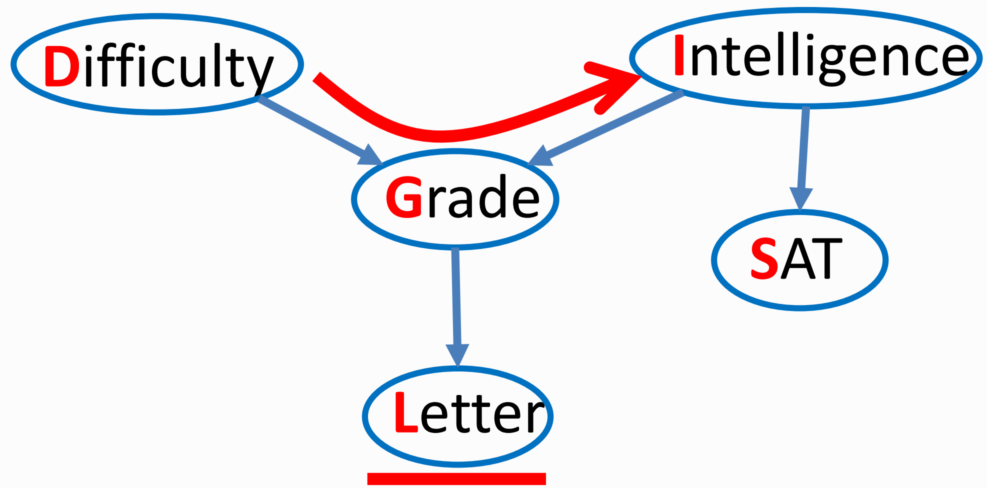
	$W \notin Z$	$W \in Z$
$X \rightarrow W \rightarrow Y$	✓	✗
$X \leftarrow W \leftarrow Y$	✓	✗
$X \leftarrow W \rightarrow Y$	✓	✗
$X \rightarrow W \leftarrow Y$	???	✓



When can X influence Y Given evidence about Z?

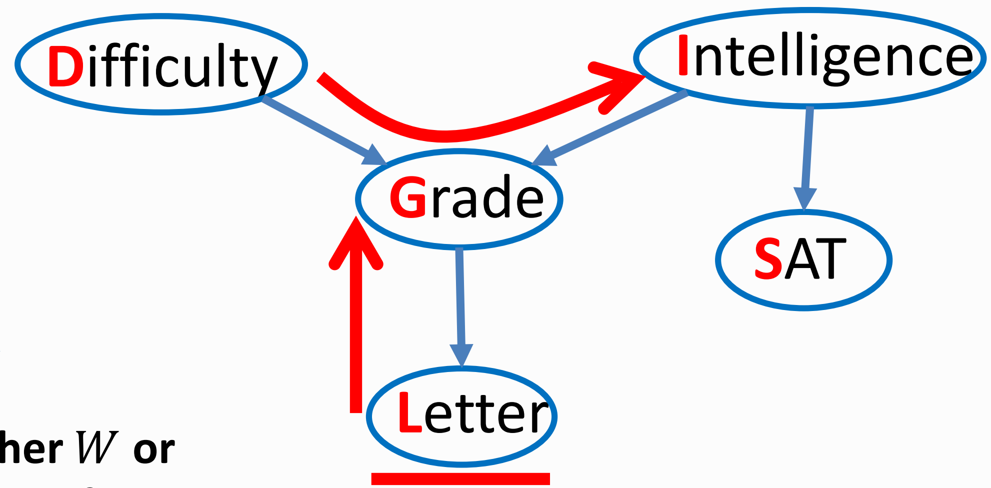
- $X \rightarrow Y$ ✓
- $X \leftarrow Y$ ✓

	$W \notin Z$	$W \in Z$
$X \rightarrow W \rightarrow Y$	✓	✗
$X \leftarrow W \leftarrow Y$	✓	✗
$X \leftarrow W \rightarrow Y$	✓	✗
$X \rightarrow W \leftarrow Y$	✗ ???	✓



When can X influence Y Given evidence about Z?

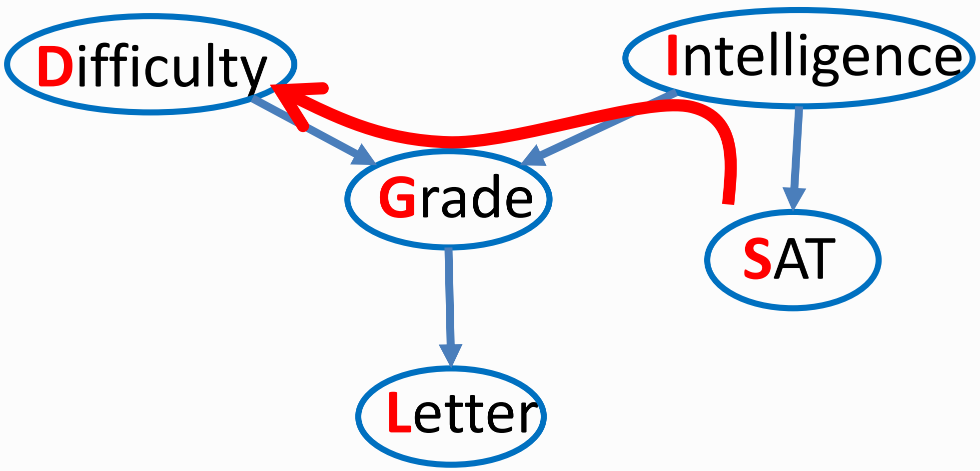
	$W \notin Z$	$W \in Z$
• $X \rightarrow Y$ ✓		
• $X \leftarrow Y$ ✓		
• $X \rightarrow W \rightarrow Y$	✓	✗
• $X \leftarrow W \leftarrow Y$	✓	✗
• $X \leftarrow W \rightarrow Y$	✓	✗
• $X \rightarrow W \leftarrow Y$	✗ if W and all of its descents not in Z	✓ either W or one of its descents is in Z



When can X influence Y Given evidence about Z?

$S - I - G - D$ allows influence to flow when:

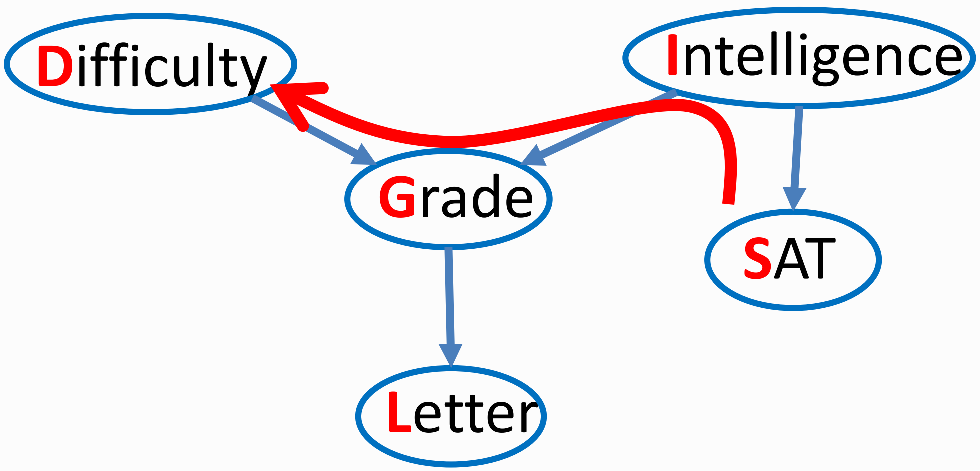
- I observed?



When can X influence Y Given evidence about Z?

$S - I - G - D$ allows influence to flow when:

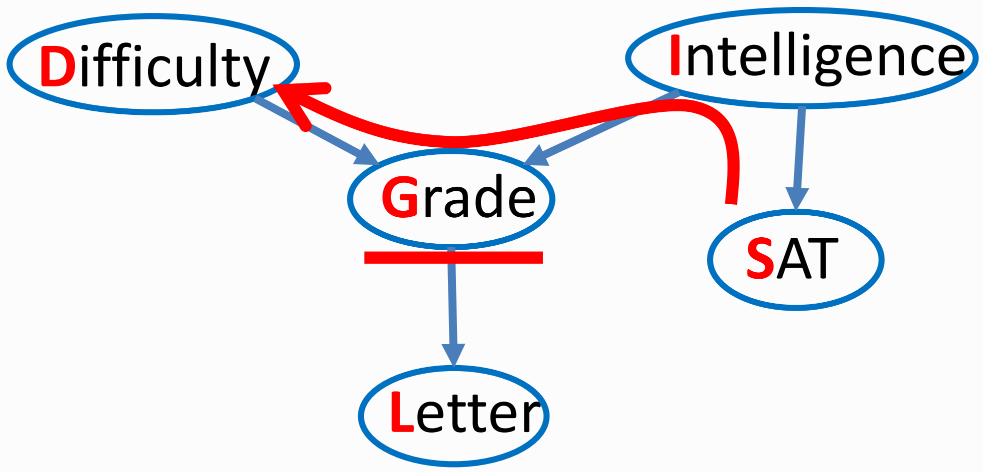
- I observed? ✗
- I not observed, nothing else?



When can X influence Y Given evidence about Z?

$S - I - G - D$ allows influence to flow when:

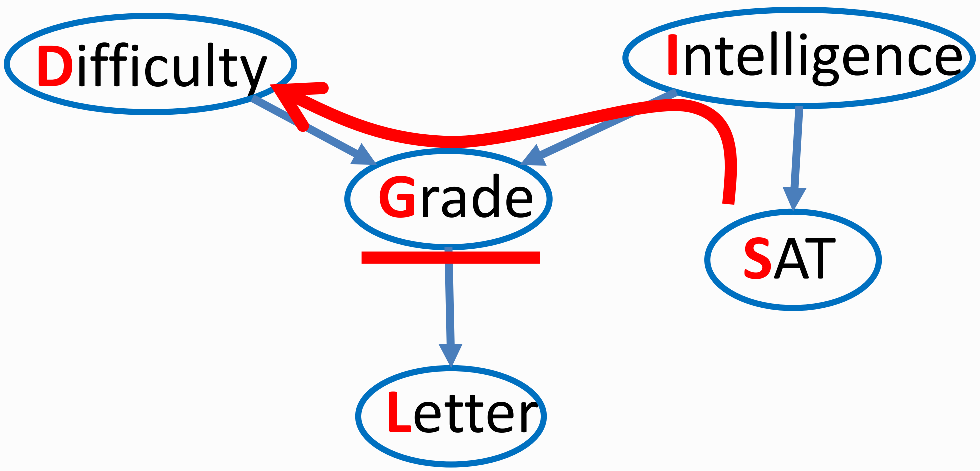
- I observed? ✗
- I not observed, nothing else? ✗
- I not observed G is observed?



When can X influence Y Given evidence about Z?

$S - I - G - D$ allows influence to flow when:

- I observed? ✗
- I not observed, nothing else? ✗
- I not observed G is observed? ✓



Active Trails

A trail $X_1 - \cdots - X_k$ is active if given Z if:

- for any v-structure $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$ we have that X_i or one of its descendants $\in Z$ activate v-structure
- no other X_i is in Z not in v-structure