

# Lecture 23 -- Reasoning Patterns

Prof Xiaowei Huang

<https://cgi.csc.liv.ac.uk/~xiaowei/>

(Attendance Code: **768218**)

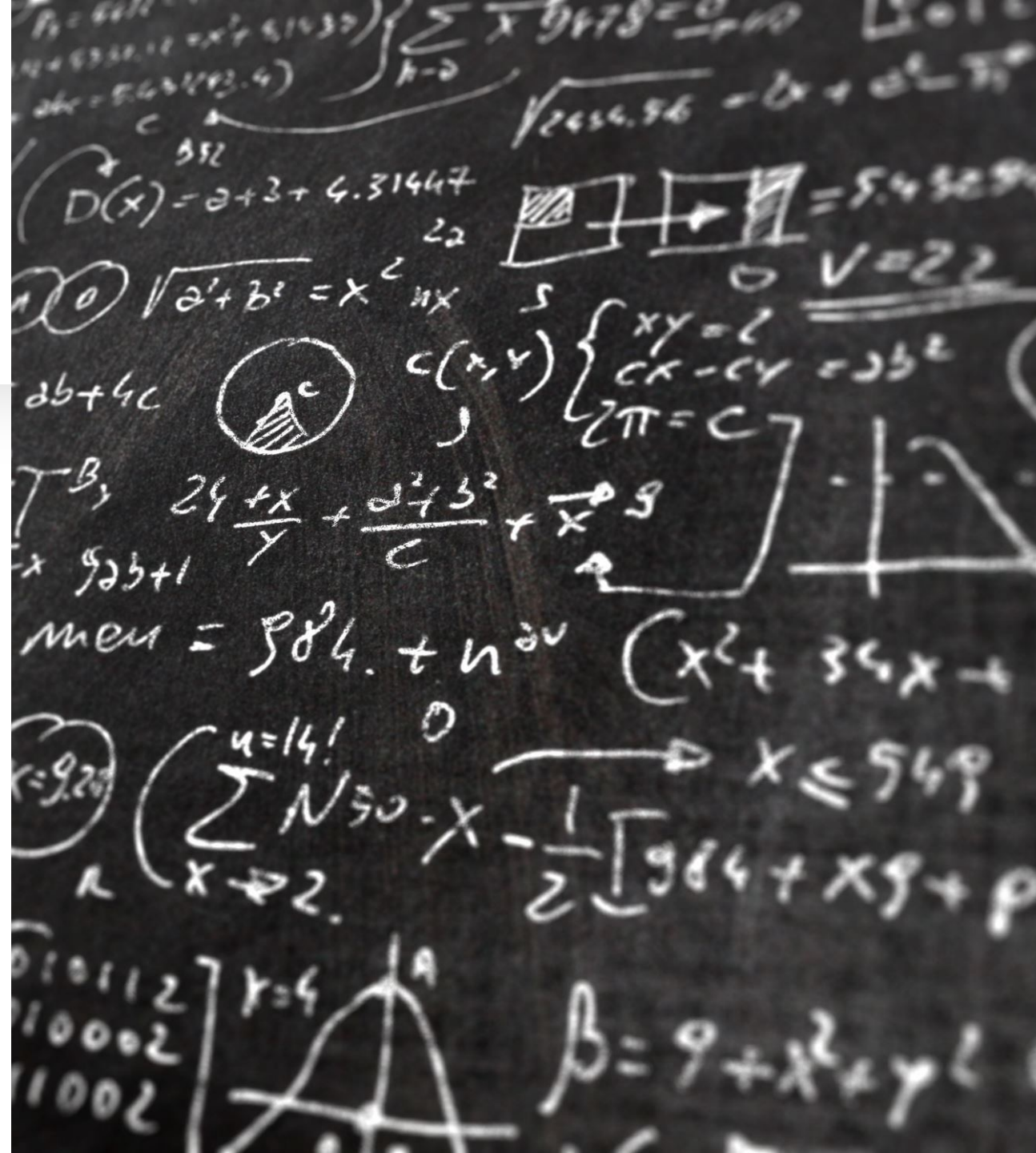
# Up to now,

- Traditional Machine Learning Algorithms
- Adversarial Attack and Defence
- Deep learning
- Probabilistic Graphical Models
  - Introduction
  - Data  $\Rightarrow$  Distribution  $\Leftrightarrow$  Graph



# Topics

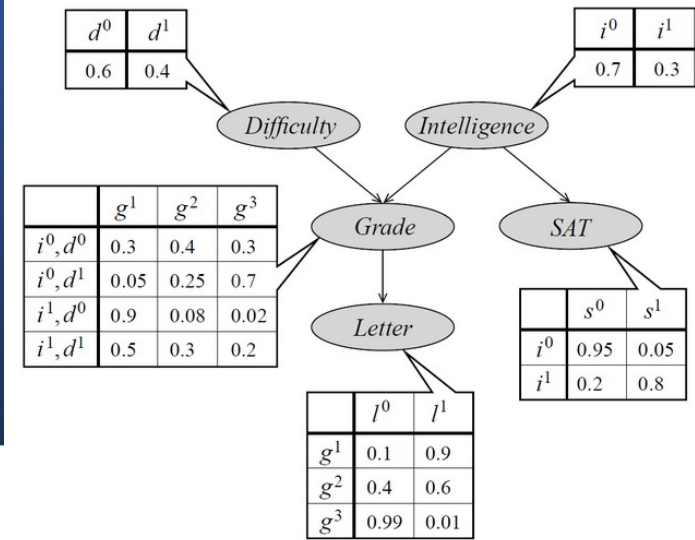
- Reasoning Patterns (i.e., do reasoning on a given graph)
  - Causal Reasoning
  - Evidential Reasoning
  - Intercausal reasoning



# Recap: Local Independencies in a BN

- A BN  $G$  is a directed acyclic graph whose nodes represent random variables  $X_1, \dots, X_n$ .
- Let  $Pa(X_i)$  denote parents of  $X_i$  in  $G$
- Let  $Non-Desc(X_i)$  denote variables in  $G$  that are not descendants of  $X_i$
- Then  $G$  encodes the following set of *conditional independence* assumptions denoted  $I(G)$ 
  - For each  $X_i$ :  $(X_i \perp Non-Desc(X_i) \mid Pa(X_i))$
- Also known as *Local Markov Independencies*

# Recap: Local Independencies



- Graph  $G$  with CPDs is equivalent to a set of independence assertions

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

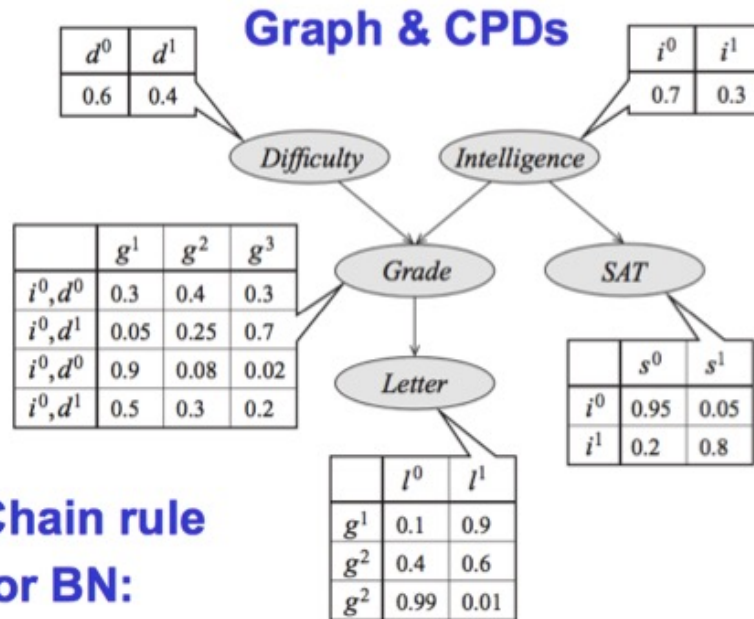
- Local Conditional Independence Assertions (starting from leaf nodes):

$$I_{local}(G) = \{ (L \perp I, D, S | G), \quad L \text{ is conditionally independent of all other nodes given parent } G \\ (S \perp D, G, L | I), \quad S \text{ is conditionally independent of all other nodes given parent } I \\ (G \perp S | D, I), \quad \text{Even given parents, } G \text{ is NOT independent of descendant } L \\ (I \perp D | \emptyset), \quad \text{Nodes with no parents are marginally independent} \\ (D \perp I, S | \emptyset) \} \quad D \text{ is independent of non-descendants } I \text{ and } S$$

- Parents of a variable shield it from probabilistic influence
  - Once value of parents known, no influence of ancestors
- Information about descendants can change beliefs about a node



# Recap: Evaluating a Joint Probability



$Val(I) = \{i^0 = \text{low intelligence}, i^1 = \text{high intelligence}\}$

$Val(D) = \{d^0 = \text{easy}, d^1 = \text{hard}\}$

$Val(G) = \{g^1 = A, g^2 = B, g^3 = C\}$

$Val(S) = \{s^0 = \text{low}, s^1 = \text{high}\}$

$Val(L) = \{l^0 = \text{weak}, l^1 = \text{strong}\}$

**Chain rule  
for BN:**

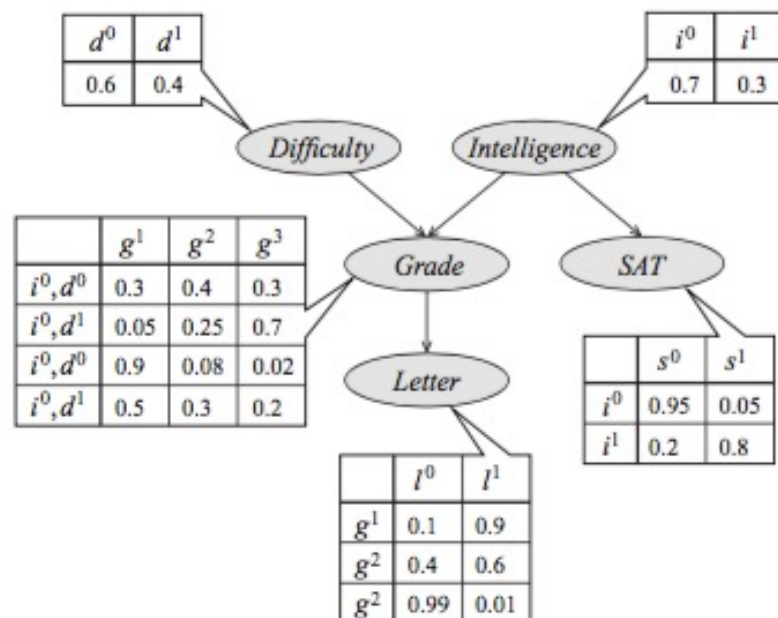
$$\begin{aligned}
 P(D, I, G, S, L) &= P(D)P(I)P(G|D, I)P(S|I)P(L|G) \\
 P(i^1, d^0, g^2, s^1, l^0) &= P(i^1)P(d^0)P(g^2|i^1, d^0)P(s^1|i^1)P(l^0|g^2) \\
 &= 0.3 \cdot 0.6 \cdot 0.08 \cdot 0.8 \cdot 0.4 = 0.004608
 \end{aligned}$$

→

$P(\text{high intelligence, easy course, grade=B, high SAT, weak letter}) = \text{very low}$

# Reasoning Patterns

- Reasoning about a student *George* using the model

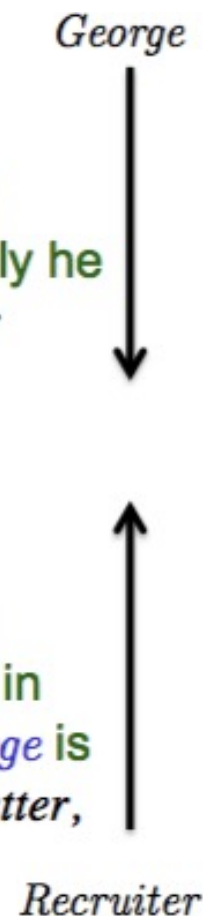


## Causal Reasoning

- George* is interested in knowing as to how likely he is to get a strong *Letter* (based on *Intelligence*, *Difficulty*)?

## Evidential Reasoning

- Recruiter* is interested in knowing whether *George* is *Intelligent* (based on *Letter*, *SAT*)

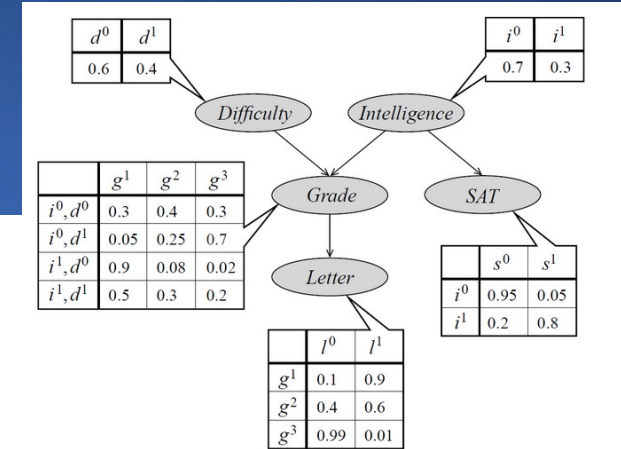




Causal Reasoning



# Causal Reasoning



- How likely *George* will get a strong *Letter* (No evidence)?

$$P(l^1) = \sum_{D,I,G,S} P(D, I, G, S, L = l^1) = \sum_{D,I,G,S} P(D)P(I)P(G|D, I)P(S|I)P(l^1|G)$$

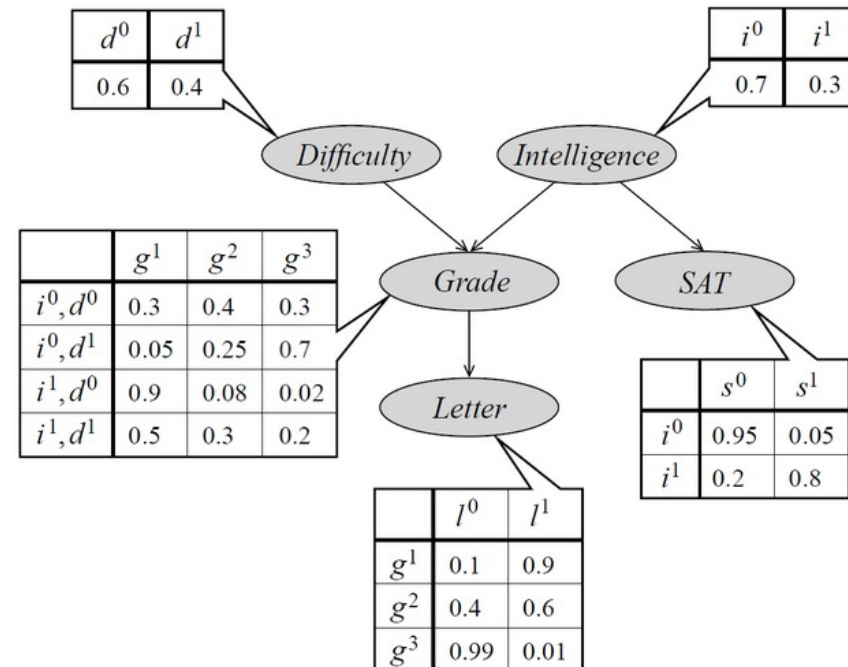
$$\begin{aligned}
 & P(D = d^0)P(I = i^0)P(g = g^1|D = d^0, I = i^0)P(S = s^0|I = i^0)P(L = l^1|g = g^1) \\
 + & P(D = d^0)P(I = i^0)P(g = g^2|D = d^0, I = i^0)P(S = s^0|I = i^0)P(L = l^1|g = g^2) \\
 + & P(D = d^0)P(I = i^0)P(g = g^3|D = d^0, I = i^0)P(S = s^0|I = i^0)P(L = l^1|g = g^3) \\
 = & + P(D = d^0)P(I = i^0)P(g = g^1|D = d^0, I = i^0)P(S = s^1|I = i^0)P(L = l^1|g = g^1) \\
 + & P(D = d^0)P(I = i^0)P(g = g^2|D = d^0, I = i^0)P(S = s^1|I = i^0)P(L = l^1|g = g^2) \\
 + & P(D = d^0)P(I = i^0)P(g = g^3|D = d^0, I = i^0)P(S = s^1|I = i^0)P(L = l^1|g = g^3) \\
 & \dots
 \end{aligned}$$

# Causal Reasoning

- How likely *George* will get a strong *Letter* (No evidence)?

$$P(l^1) = \sum_{D,I,G,S} P(D, I, G, S, L = l^1) = \sum_{D,I,G,S} P(D)P(I)P(G|D, I)P(S|I)P(l^1|G)$$

- $P(l^1)=0.502$
- Obtained by summing-out other variables in joint distribution

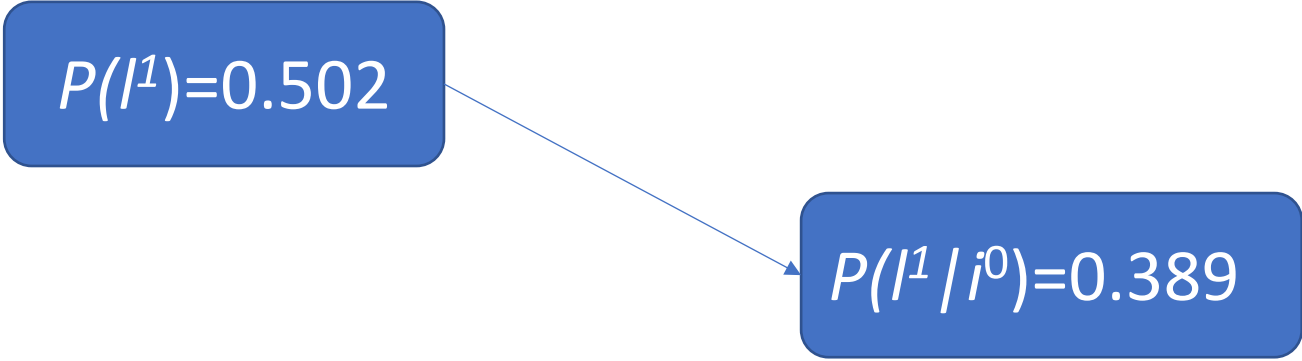


# Causal Reasoning

- Knowing *George* is not so *Intelligent* ( $i^0$ )

$$P(l^1|i^0) = \frac{P(l^1, i^0)}{P(i^0)} = \frac{\sum_{D,I,G} P(D)P(i^0)P(G|D, i^0)P(S|i^0)P(l^1|G)}{\sum_{D,G,S,L} P(D)P(i^0)P(G|D, i^0)P(S|i^0)P(L|G)}$$

- $P(l^1|i^0)=0.389$

$$P(I^1)=0.502$$


$$P(I^1 | i^0)=0.389$$

After knowing that the student is not as intelligent, we understand that the probability of getting a strong recommendation letter is lowered.



# Causal Reasoning

- Knowing COMP219 is not *Difficult* ( $d^0$ )
- $P(I^1 / i^0, d^0) = 0.513$  (Exercise!)

$$P(I^1)=0.502$$

$$P(I^1 | i^0)=0.389$$

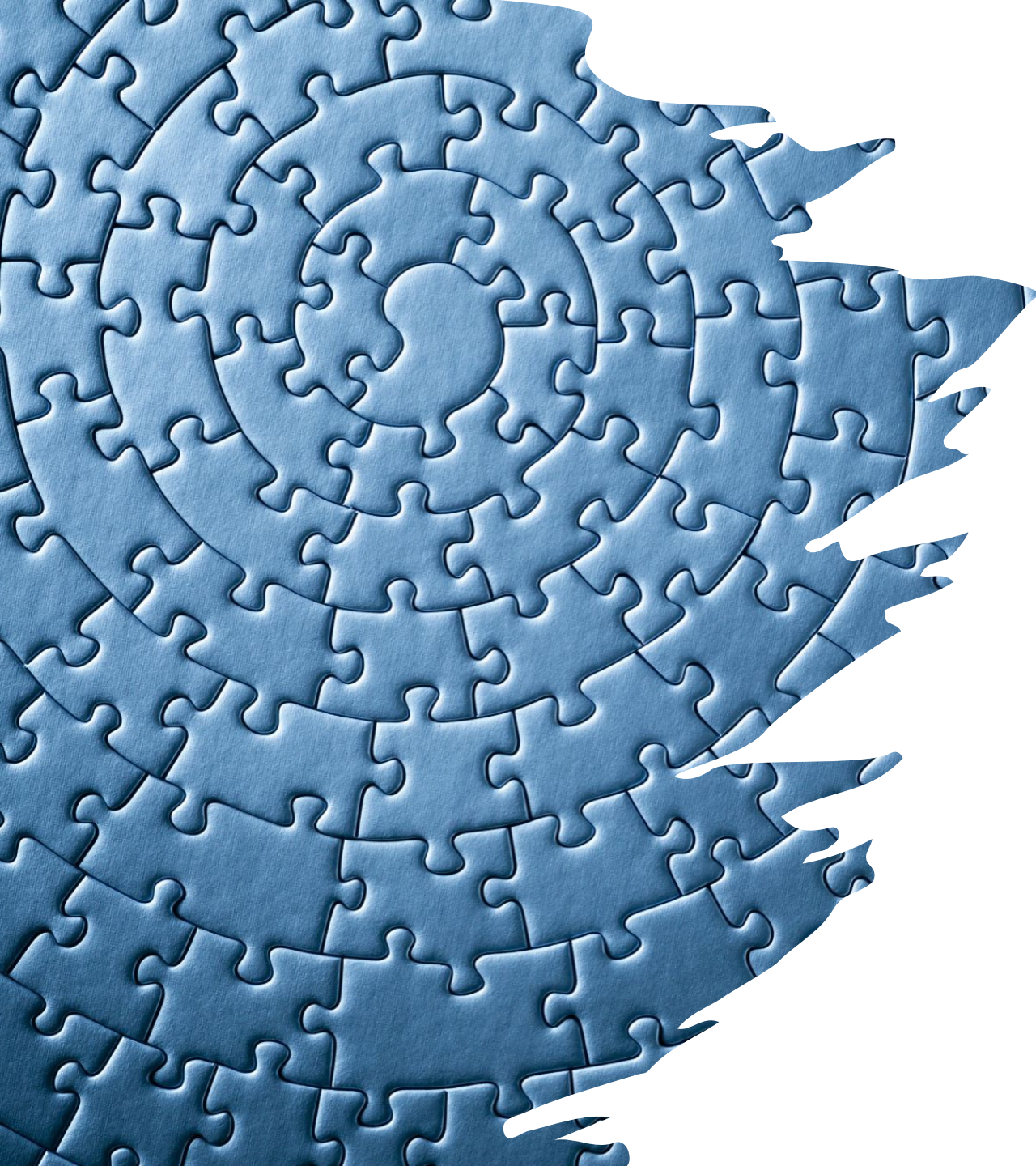
$$P(I^1 | i^0, d^0)=0.513$$

After knowing that the student is not as intelligent, we understand that the probability of getting a strong recommendation letter is lower.

After further knowing that the difficulty is low, the probability of getting a strong letter is higher.

# Causal Reasoning

- Observe how probabilities change as more evidence is obtained
- ***Causal Reasoning:***  
Predicting downstream effects of factors such as *Intelligence*



# Evidential Reasoning



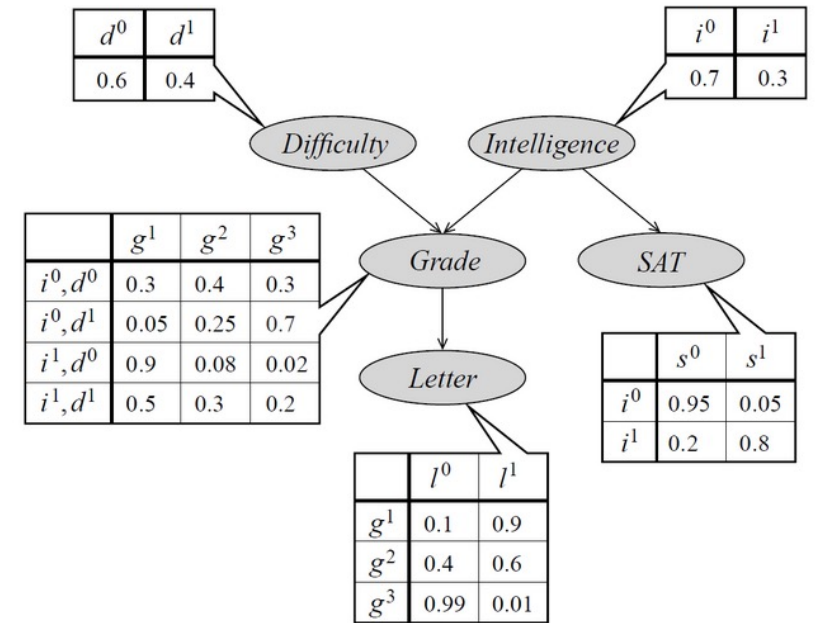
# Evidential Reasoning

- Recruiter wants to hire *Intelligent* student
- A priori *George* is 30% likely to be *Intelligent*

$$P(i^1)=0.3$$

- Finds that *George* received *Grade C* ( $g^3$ ) in *COMP219*

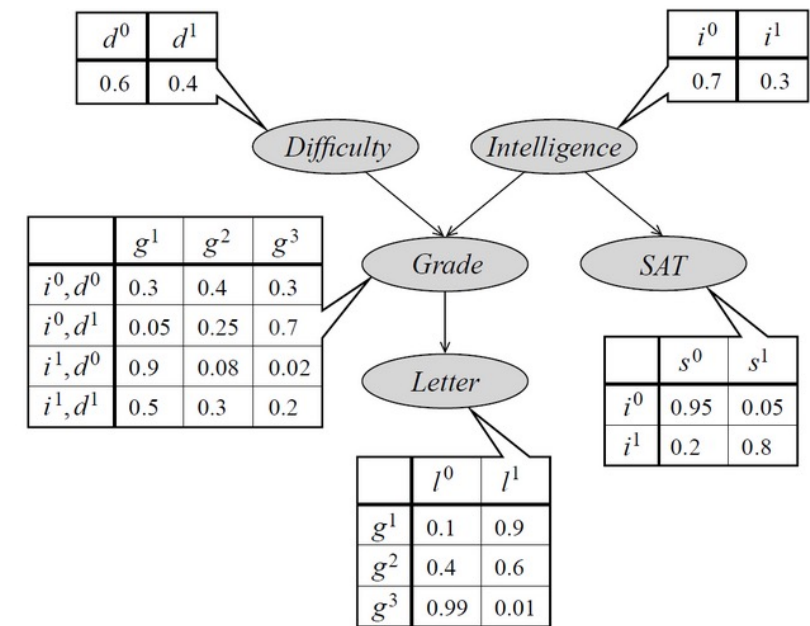
$$P(i^1 | g^3)=0.079$$

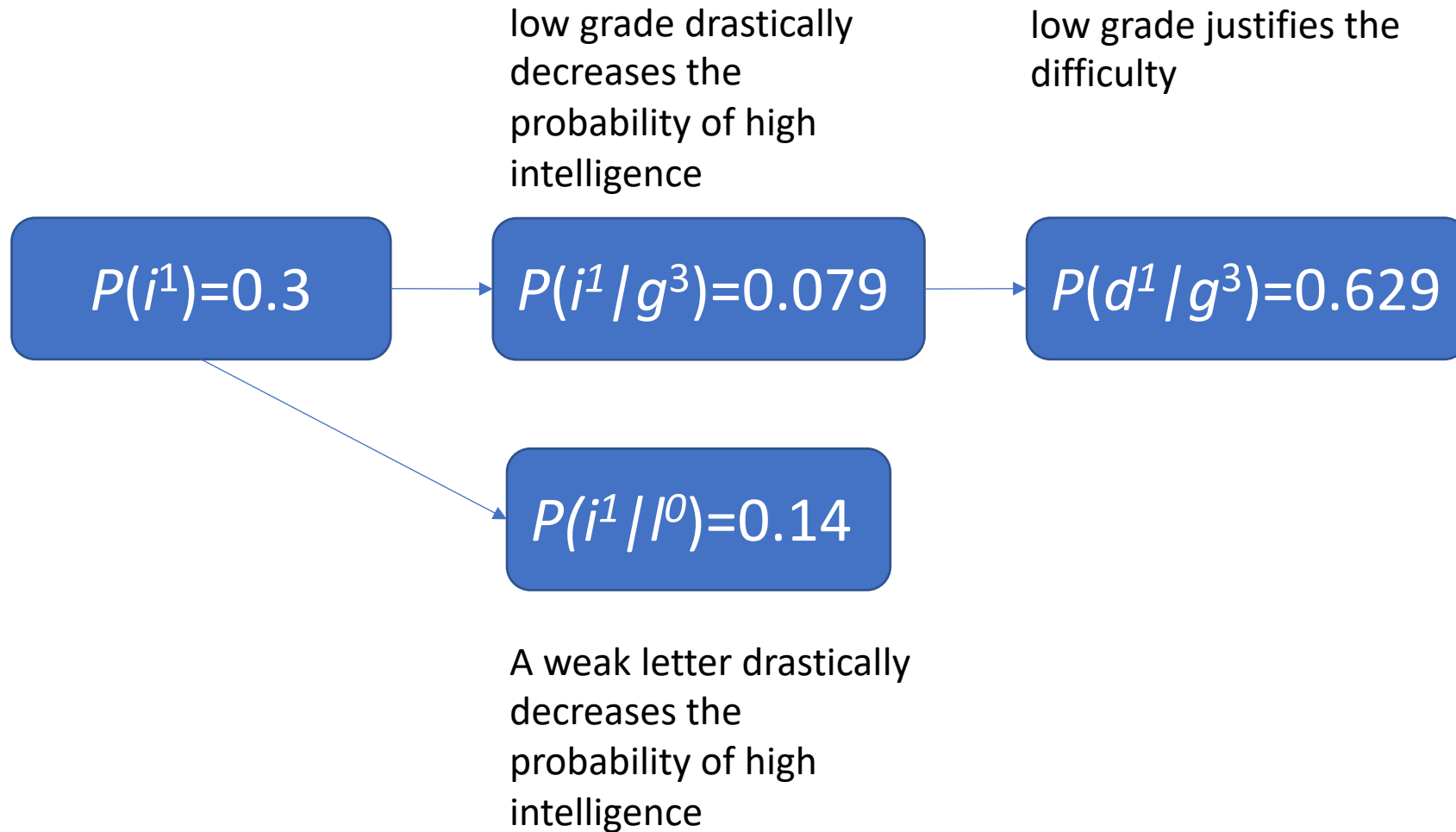


$$\begin{aligned} P(i^1 | g^3) &= \frac{P(i^1, g^3)}{P(g^3)} \\ &= \frac{\sum_{D,S,L} P(D)P(i^1)P(g^3|D, i^1)P(S|i^1)P(L|g^3)}{\sum_{D,I,S,L} P(D)P(I)P(g^3|D, I)P(S|I)P(L|g^3)} \end{aligned}$$

# Evidential Reasoning

- Recruiter wants to hire *Intelligent* student
- A priori *George* is 30% likely to be *Intelligent*  
 $P(i^1)=0.3$
- Finds that *George* received *Grade C* ( $g^3$ ) in *COMP219*  
 $P(i^1 / g^3)=0.079$
- Similarly probability of *Difficult* goes up from 0.4 to  
 $P(d^1 / g^3)=0.629$
- If recruiter has lost *Grade* but has *Letter*  
 $P(i^1 / l^0)=0.14$





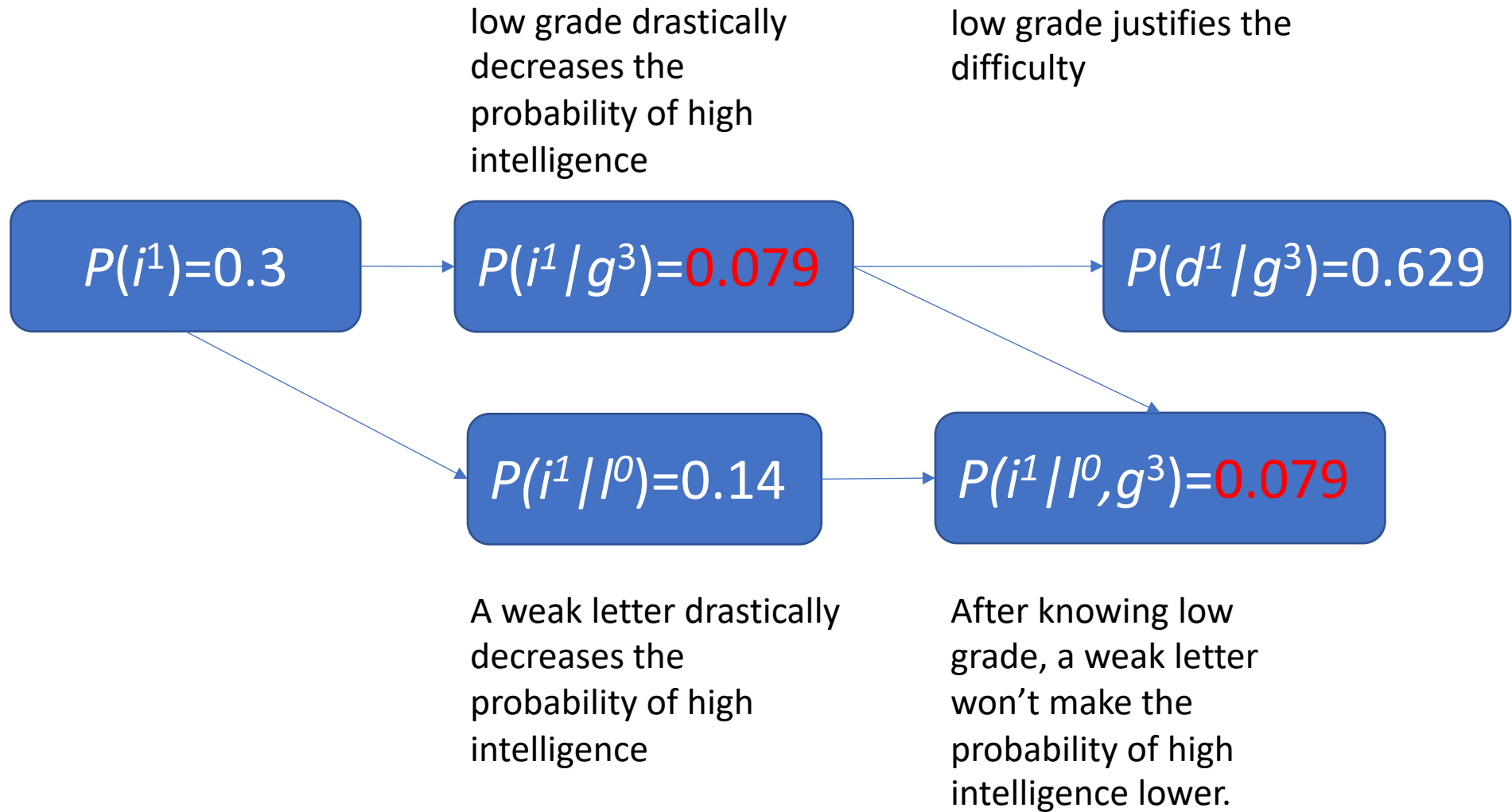


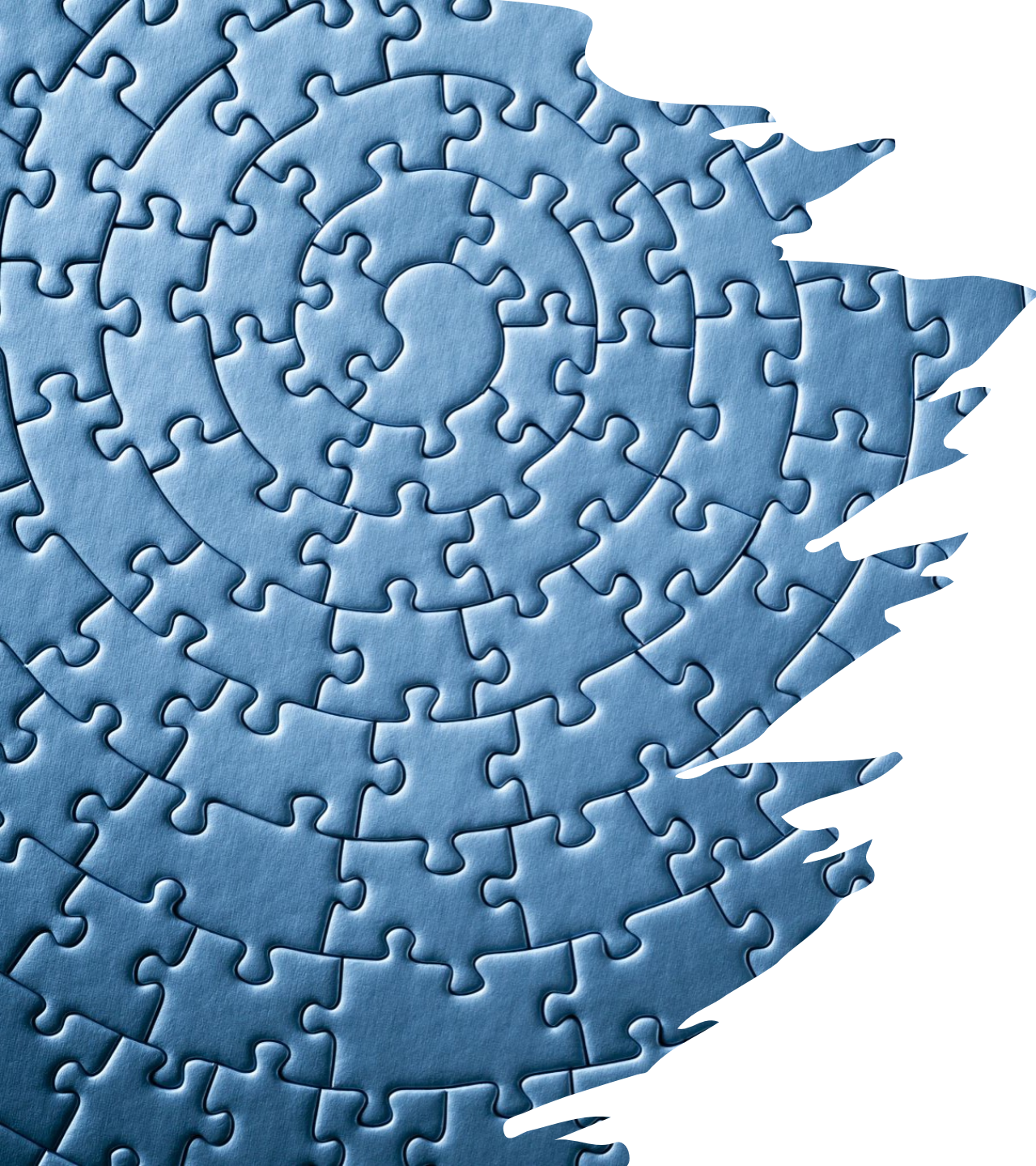
# Evidential Reasoning

- Recruiter has both *Grade* and *Letter*

$$P(i^1 / I^0, g^3) = 0.079$$

- Same as if he had only *Grade*
  - *Letter* is immaterial
- Reasoning from effects to causes is called evidential reasoning





Intercausal reasoning

# Intercausal reasoning

- Recruiter has Grade (*Letter* does not matter for *Intelligence*)

$$P(i^1 / g^3) = P(i^1 / l^0, g^3) = 0.079$$

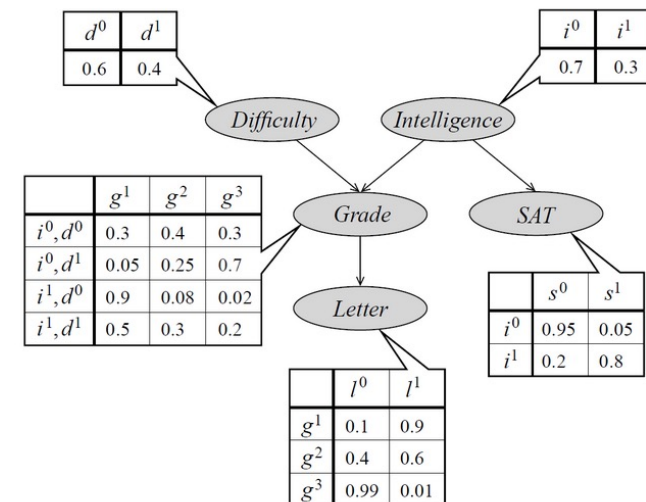
- Recruiter receives high *Score* (leads to dramatic increase)

$$P(i^1 / g^3, s^1) = 0.578$$

- Intuition:
  - High *Score* outweighs poor grade since low intelligence rarely gets good *Scores*
  - Smart students more likely to get Cs in hard classes

- At the meantime, Probability of class is difficult also goes up from

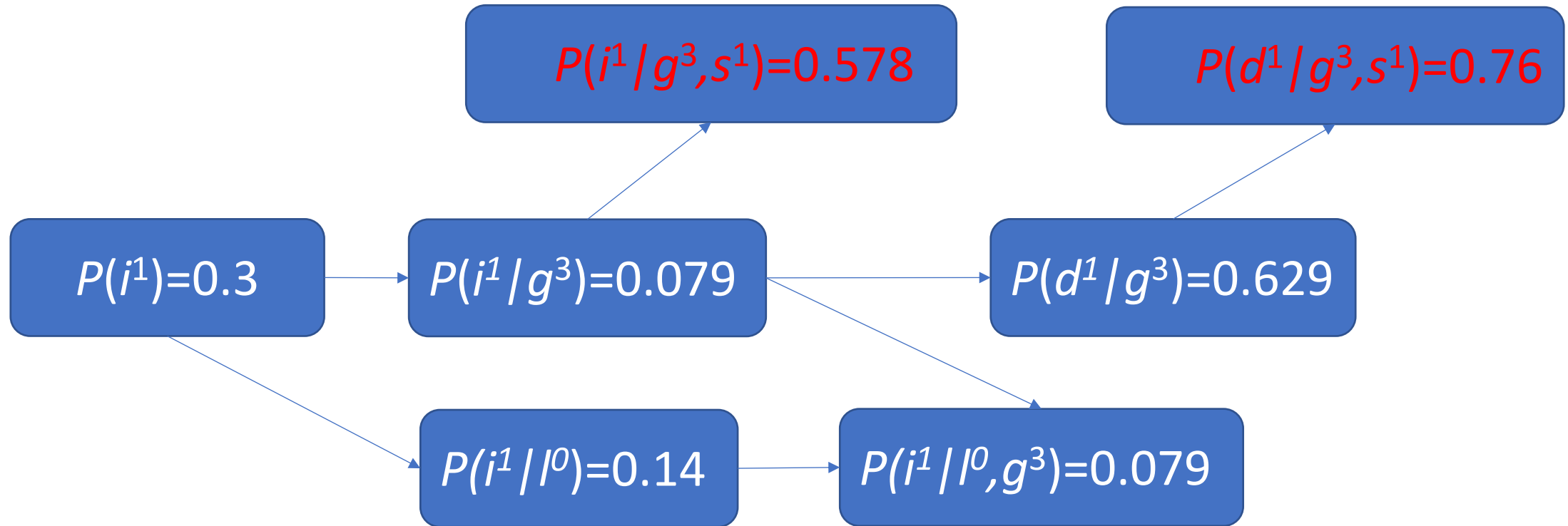
- $P(d^1 / g^3) = 0.629$  to
- $P(d^1 / g^3, s^1) = 0.76$





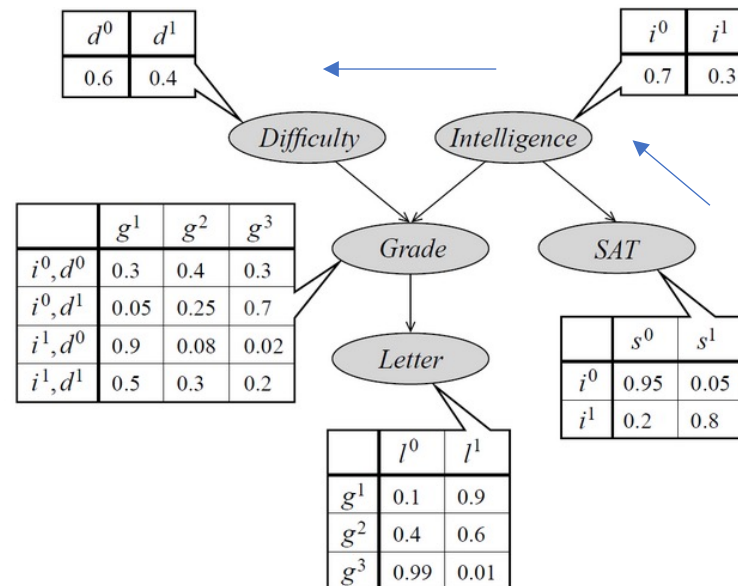
High Score outweighs poor grade since low intelligence rarely gets good Scores

Probability of class is difficult also goes up



# Intercausal reasoning

- The previous example:
  - Information about Score gave us information about Intelligence which with Grade told us about difficulty of course
  - One causal factor for **Grade**, i.e., **Intelligence**, give us information about another (**Difficulty**)



# Explaining Away

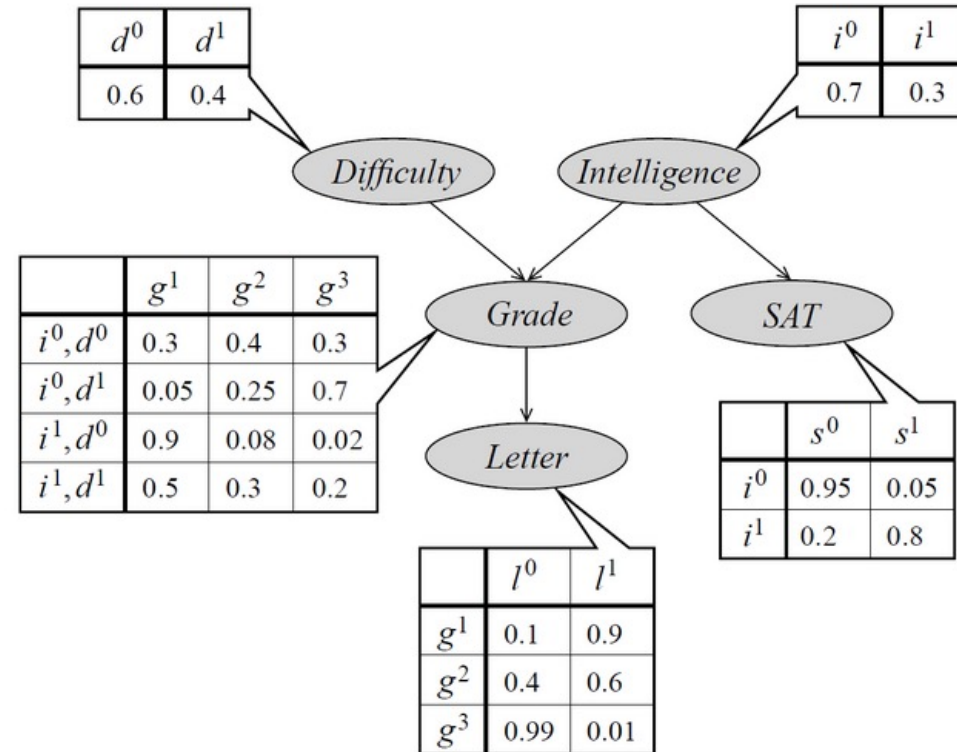
- Given  $Grade=C$ ,  $Letter=weak$

$$P(i^1 / g^3) = 0.079$$

- If we observe  $Difficulty=high$

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

- We have provided partial explanation for George's grade in COMP219



$0.11 < 0.3$  : partial explanation  
for George's grade

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

$$P(i^1) = 0.3$$

$$P(i^1 | g^3) = 0.079$$

# Explaining Away

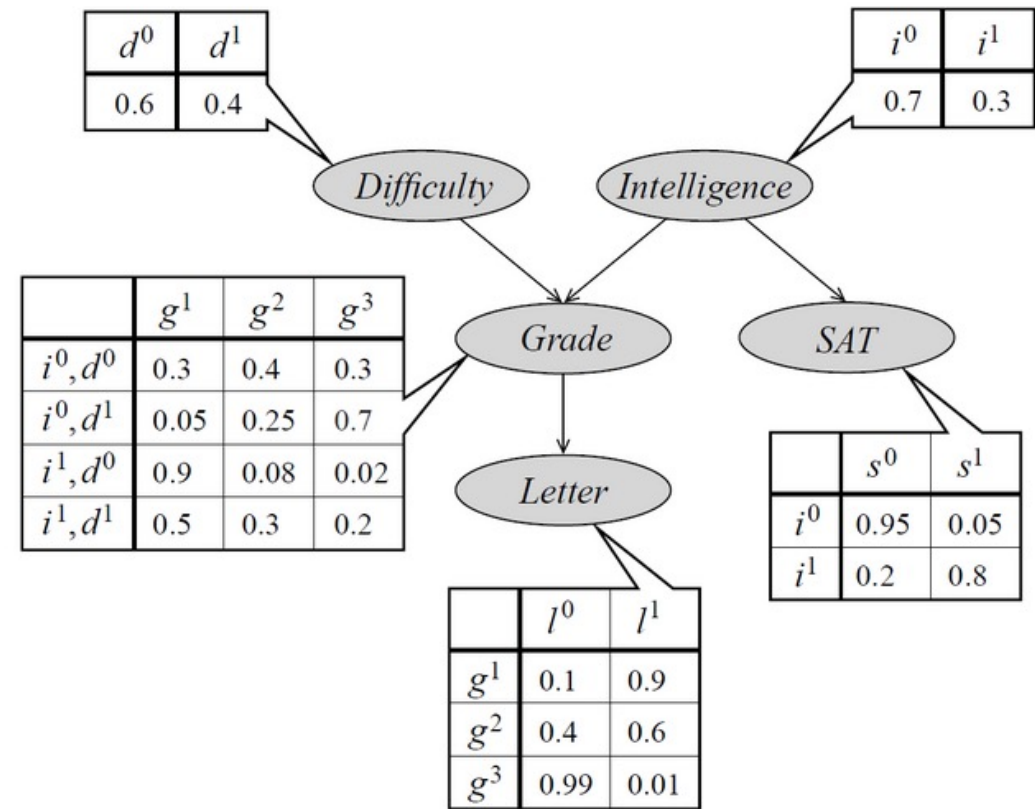
- If George gets a *B* in COMP219

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

- If we observe COMP219 is hard

$$P(i^1 / g^2, d^1) = 0.34$$

- We have **explained away the poor grade** via the difficulty of the class



partial explanation for George's  
grade

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

$$P(i^1) = 0.3$$

$$P(i^1 | g^3) = 0.079$$

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

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

$0.34 > 0.3$ :  
explained away the  
poor grade via the  
difficulty of the class



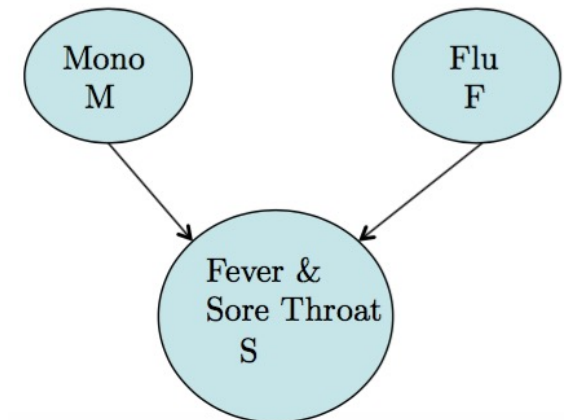
# Explaining Away

- Explaining away is one type of intercausal reasoning
- Different causes of the same effect can interact
- All determined by probability calculation rather than heuristics

# Common in Human Reasoning

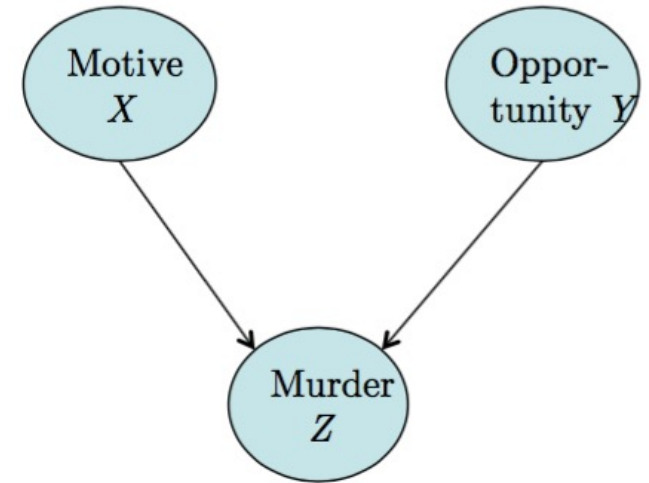
- Binary Variables
- Fever & Sore Throat can be caused by mono and flu
- When flu is diagnosed probability of mono is reduced (although mono could still be present)
- It provides an alternative explanation of symptoms

$$P(m^1/s^1) > P(m^1/s^1, f^1)$$



# Another Type of Intercausal Reasoning

- Binary Variables
  - Murder (leaf node)
  - Motive and Opportunity are causal nodes
- Binary Variables  $X, Y, Z$
- $X$  and  $Y$  both increase the probability of Murder
  - $P(z^1 | x^1) > P(z^1)$
  - $P(z^1 | y^1) > P(z^1)$
- Each of  $X$  and  $Y$  increase probability of the other
  - $P(x^1 | z^1) < P(x^1 | y^1, z^1)$
  - $P(y^1 | z^1) < P(y^1 | x^1, z^1)$



Can go in any direction  
Different from Explaining  
Away