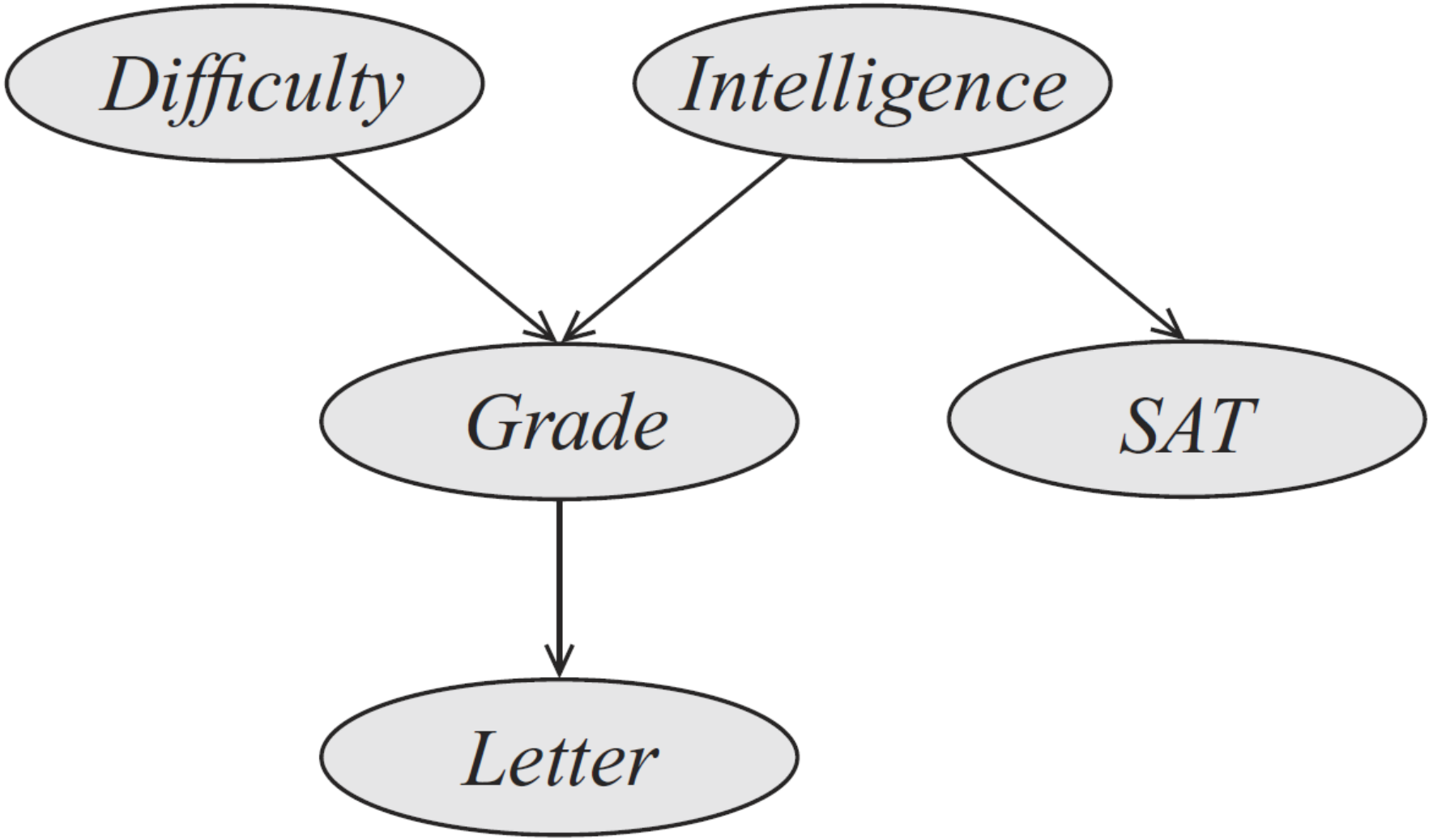
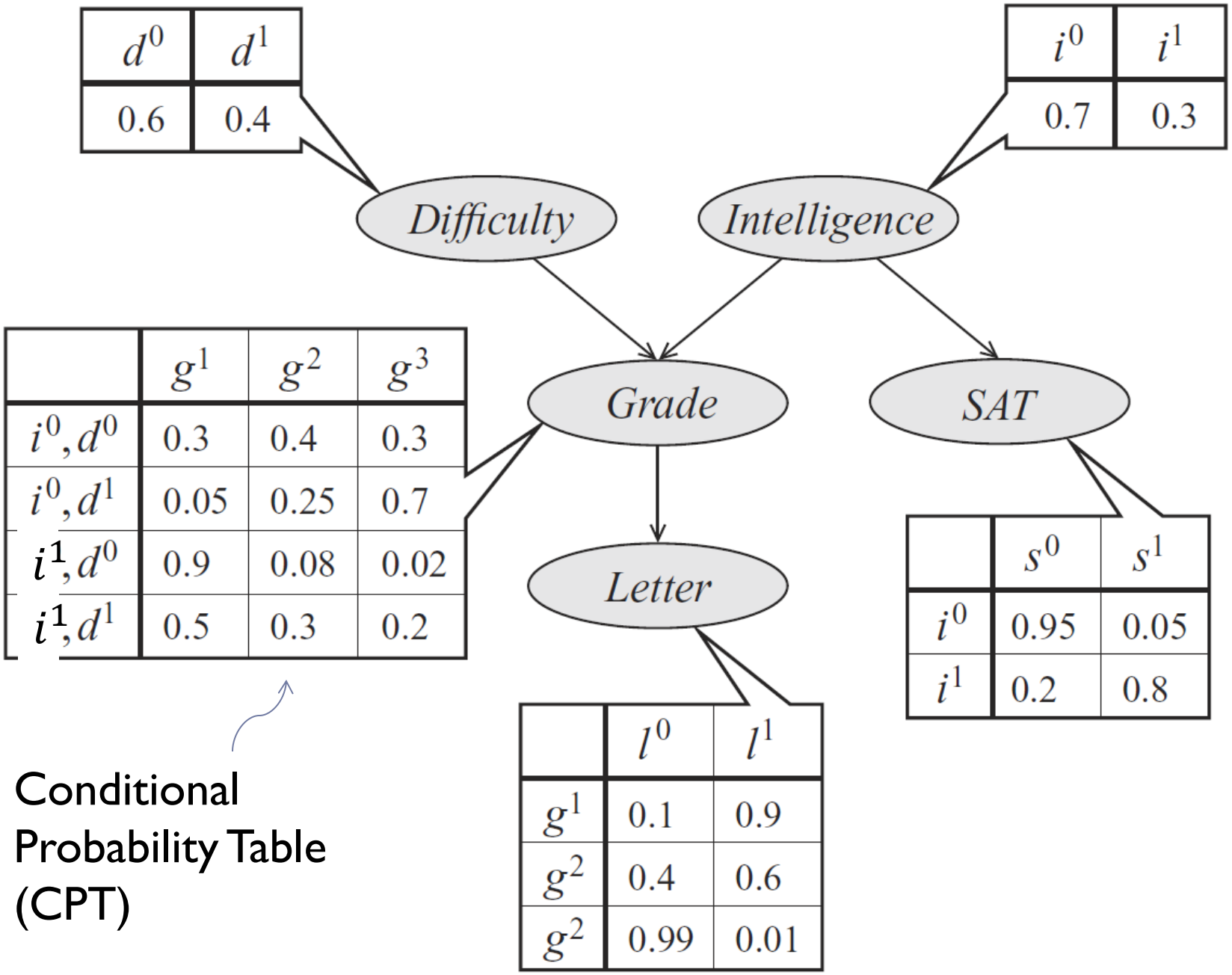
Bayesian Networks

Doug Downey

EECS 395/495 Probabilistic Graphical Models

- Begin with a graph
  - Random variables as nodes
  - Causal relationships as directed edges


Difficulty
Intelligence
Grade
SAT
Letter



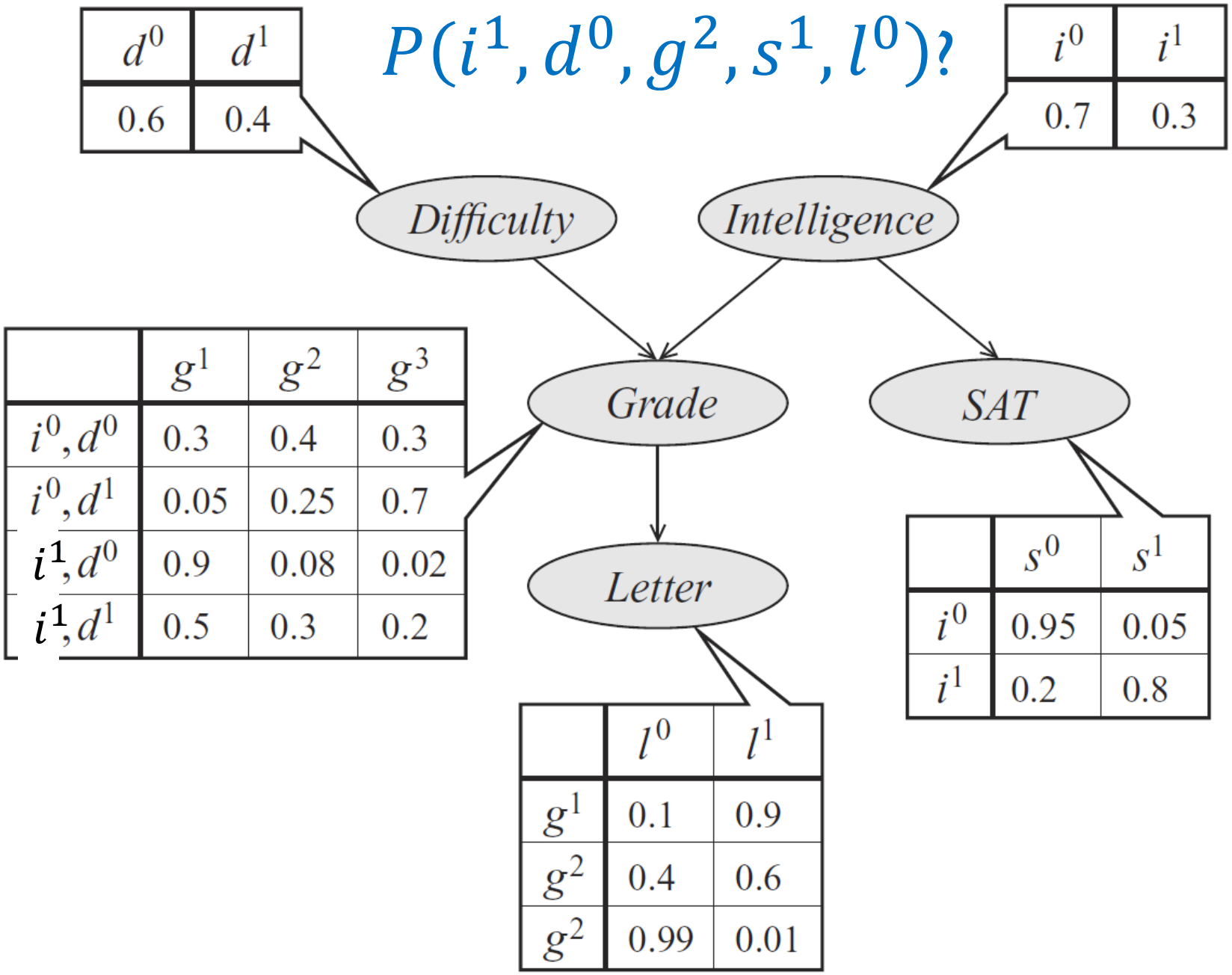
Conditional Probability Table (CPT)

What does this wacky thing do?

- BNs represent the joint distribution compactly
- You can obtain the BN's probabilities for an event by multiplying the relevant values from each CPT:

$$P(i^1, d^0, g^2, s^1, l^0) = \cdots$$

$P(i^1, d^0, g^2, s^1, l^0)$?



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1	g^2	g^3
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^2	0.99	0.01

What does this wacky thing do?

- BNs represent the joint distribution compactly
- You can obtain the BN's probabilities for an event by multiplying the relevant values from each CPT:

$$P(i^1, d^0, g^2, s^1, l^0)$$
$$= P(i^1)P(d^0)P(g^2|i^1, d^0)P(s^1|i^1)P(l^0|g^2)$$
$$= 0.3 \cdot 0.6 \cdot 0.08 \cdot 0.8 \cdot 0.4 = 0.004608$$

Building a Bayes Net

- Create a node for each important variable in domain
- Connect nodes with causal edges
  - How? Domain knowledge
    (or learn from data – more on this later)
- Obtain CPTs
  - How? Use **data**, or write from domain knowledge

Bayes Net Advantages

- Compactness
  - Our "student" network has **15** independent parameters
  - Vs. how many for a full joint distribution table?
- Ease of inference
  - (more on this later)

From Graphs to Independencies

- The Bayes Net encodes **independencies**
  - Independencies are what allow BN compactness

- Question:

  Which independencies are encoded in a given BN graph?

Global Semantics

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \mathrm{Pa}(X_i))$$

Local Independences

- Each node is conditionally independent of its non-descendants given its parents.

- Theorem:
  Local Independences ⇔ Global Semantics

What does the graph look like…

- No independence?
- All variables independent?
- Common Cause? Common Effect?
  - Correlation != causation
  - “Explaining away”

Active Trails and D-separation

- Two nodes in *G* are d-separated unless there is an active trail between them

- An *Active Trail* between nodes *X* and *Y* given evidence nodes ***E*** is any path between *X* and *Y* such that
  - For any v-structure (*A* => *C* <= *B*) on the path, either *C* or one of its descendents is in ***E***
  - No other nodes on the path are in ***E***