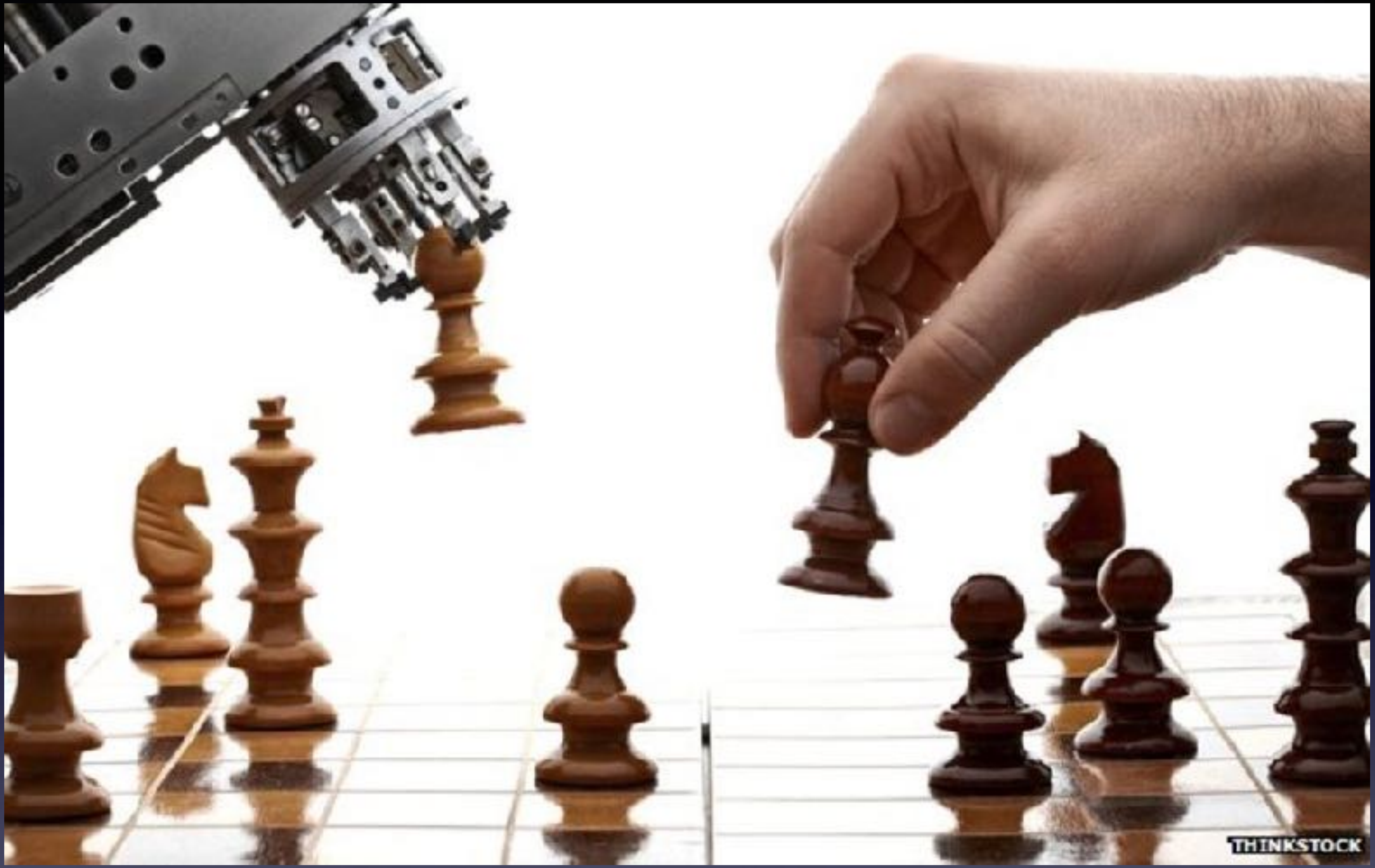
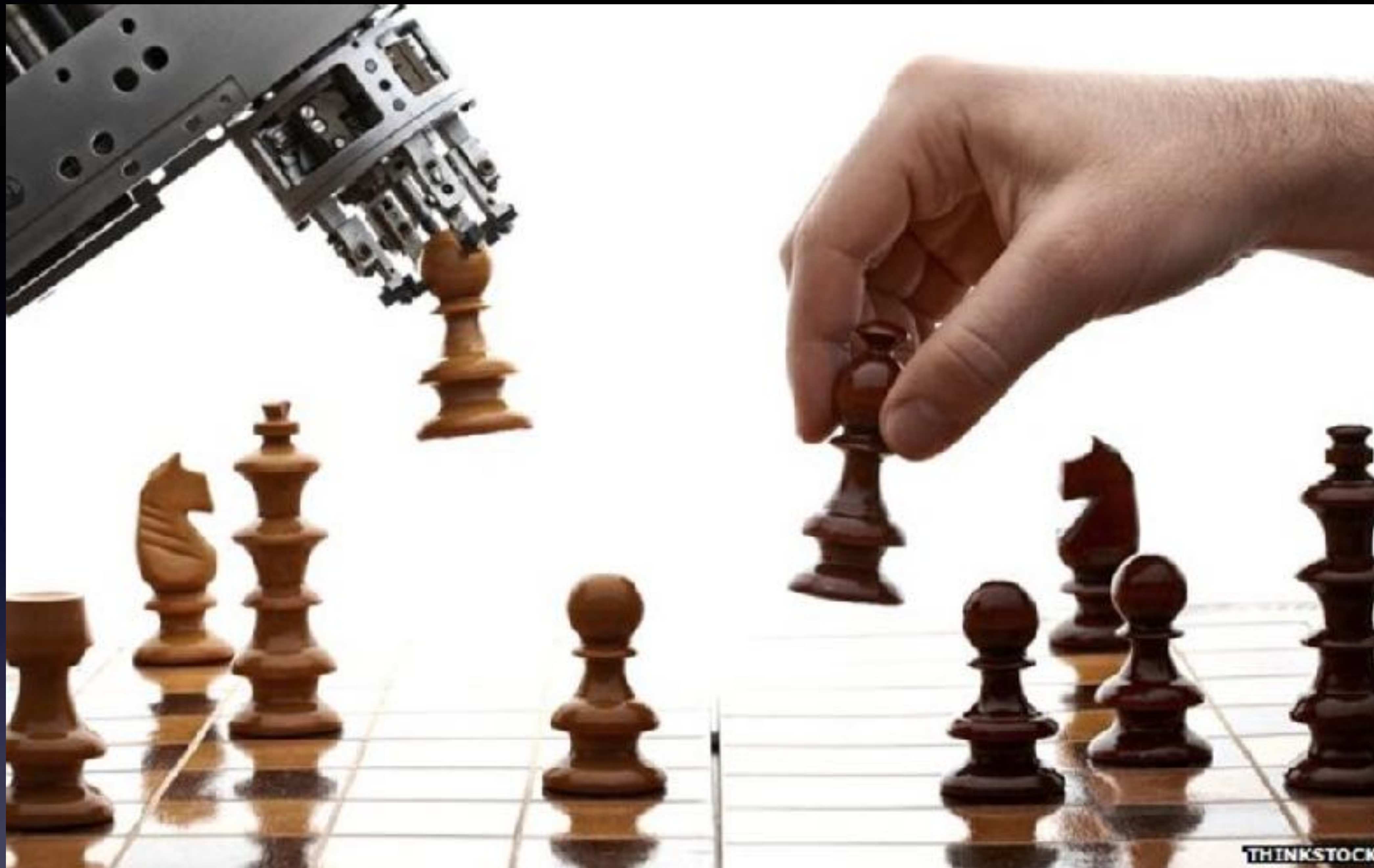


Introduction to Artificial Intelligence

COSC 4550 / COSC 5550

Professor Nick Cheney

9/20/17



Introduction to Artificial Intelligence

COSC 4550 / COSC 5550

Guest lecture by Joost Huizinga

9/20/17

Today: uncertainty and probability

- Goals:
  - Get familiar with probability notation
  - Review probability theory
  - Get everybody on the same page

Ch. 13: Uncertainty

- Uncertainty is pervasive in the world
  - e.g. diagnosing an illness
- Goal: maximize expected utility/value
- Probability Theory is our best tool
  - Lots of help with basic equations & notation in the book

Bayesian Statistics

- Very important in AI
- Allow you to have prior knowledge about the world
  - e.g. phones don't have cameras
- update your knowledge of the world
  - e.g. most phones now have cameras

Bayesian Statistics

- Priors
  - belief before seeing evidence
  - - e.g. most phones don't have cameras (belief in 2000)
  - aka "unconditional probabilities" or "prior probabilities"
- Posterior
  - aka "conditional probabilities" or "posterior probabilities"

Bayesian Statistics

- Prior
  - P(two dice sum to 12) = ??
    - Read: the probability that two dice sum to 12
- Posterior
  - P(two dice sum to 12 | Die1=6) = ??
    - Read: the probability that two dice sum to 12, given that one die landed on a six

Conditional probabilities

Mathematically speaking, conditional probabilities are defined in terms of unconditional probabilities as follows: for any propositions a and b, we have

$$P(a \mid b) = \frac{P(a \land b)}{P(b)}, \quad (13.3)$$

^ means "and"

which holds whenever $P(b) > 0$. For example,

$$P(doubles \mid Die_1 = 5) = \frac{P(doubles \land Die_1 = 5)}{P(Die_1 = 5)}.$$

The definition of conditional probability, Equation (13.3), can be written in a different form called the **product rule:**

$$P(a \land b) = P(a \mid b)P(b),$$

The product rule is perhaps easier to remember: it comes from the fact that, for a and b to be true, we need b to be true, and we also need a to be true given b.

Notation

- Capital letters indicate the Variable
  - e.g. Weather
- Lowercase letters indicate a value/instance of that variable
  - e.g. Weather = sunny (sometimes abbreviated as just sunny)
  - A = true abbreviated as a
  - A = false as ~a or $\neg a$
- What does this mean? $P(cavity \,|\, \neg toothache \wedge teen) = 0.1$

More notation

- Bolded letter is a vector $\mathbf{P}(Weather) = \langle 0.6, 0.1, 0.29, 0.01 \rangle$
- Bolded P over two variables indicates all possible combinations of values over those variables

$\mathbf{P}(Weather, Cavity)$

  - 4 x 2 table if there are 4 possible weather conditions and cavity is a binary random variable

- $P(sunny, cavity)$ means the same as $P(sunny \land cavity)$

Joint Distribution

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

- Three Boolean random variables
- Values sum to 1
- P(~toothache)?

Marginalization

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

- P(~toothache)?

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z} \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}, \mathbf{z}) \ ,$$

  - (called the “unconditional probability” or “marginal probability”)

Marginalization

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

- P(~toothache)?

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z} \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}, \mathbf{z}) \, ,$$

  - (called the “unconditional probability” or “marginal probability”)

- P(~toothache) = 0.072+0.008+0.144+0.576= .8

Marginalization

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z} \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}, \mathbf{z}) \,,$$

- P(~toothache ^ catch) ?

Marginalization

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z} \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}, \mathbf{z}) \,,$$

- P(~toothache ^ catch) = 0.072+0.144= 0.216

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)} \,,$$

Bayesian Statistics

$$P(cavity \mid toothache) =$$

(Alone or with your neighbor)

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)} \,,$$

Bayesian Statistics

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)} ,$$

Bayesian Statistics

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$
$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6 .$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)} ,$$

Bayesian Statistics

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$
$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6 .$$

Just to check, we can also compute the probability that there is no cavity, given a toothache:

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$
$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4 .$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)},$$

Bayesian Statistics

Why divide by *P(toothache)*?

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6 .$$

Just to check, we can also compute the probability that there is no cavity, given a toothache:

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4 .$$

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

Bayesian Statistics

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$

Why divide by *P(toothache)*?

- In the world where we have a toothache, probabilities still need to sum to 1.
  - P(cavity ^ toothache) = 0.108 + 0.012 = 0.12
  - P(~cavity ^ toothache) = 0.016 + 0.064 = 0.08
- 0.12 and 0.08 are the right relative proportions, but they don't sum to 1.
- Thus, 1/*P(toothache)* functions as a normalization constant

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)} ,$$

Bayesian Statistics

- Such normalization constants will be denoted by α
- Note: you can calculate **P**(Cavity | toothache) without separately calculating P(toothache)

$$\begin{aligned}
\mathbf{P}(Cavity \mid toothache) &= \alpha\, \mathbf{P}(Cavity, toothache) \\
&= \alpha\, [\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)] \\
&= \alpha\, [\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle] = \alpha\, \langle 0.12, 0.08 \rangle = \langle 0.6, 0.4 \rangle .
\end{aligned}$$

Conditioning

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y} \mid \mathbf{z}) P(\mathbf{z}) \,. \qquad (13.8)$$

This rule is called **conditioning.** Marginalization and conditioning turn out to be useful rules for all kinds of derivations involving probability expressions.

- P(cavity), where z = <catch, !catch>?

(Alone or with your neighbor)

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

Conditioning

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y} \mid \mathbf{z}) P(\mathbf{z}) \,.$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

- P(cavity) = P(cavity|catch)*P(catch) + P(cavity|!catch)*P(!catch)
- P(cavity|catch) =

Conditioning

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y} \mid \mathbf{z}) P(\mathbf{z}) .$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

- P(cavity) = P(cavity|catch)*P(catch) + P(cavity|!catch)*P(!catch)
- P(cavity|catch) = (.108+.072)/(.108+.016+.072+.144) = .53
- P(cavity|!catch) =

Conditioning

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y} \mid \mathbf{z}) P(\mathbf{z}) \, .$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

- P(cavity) = P(cavity|catch)*P(catch) + P(cavity|!catch)*P(!catch)
- P(cavity|catch) = (.108+.072)/(.108+.016+.072+.144) = .53
- P(cavity|!catch) = (.012+.008)/(.012+.008+.064+.576) = .03
- P(catch) =

Conditioning

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y} \mid \mathbf{z}) P(\mathbf{z}) .$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

- P(cavity) = P(cavity|catch)*P(catch) + P(cavity|!catch)*P(!catch)
- P(cavity|catch) = (.108+.072)/(.108+.016+.072+.144) = .53
- P(cavity|!catch) = (.012+.008)/(.012+.008+.064+.576) = .03
- P(catch) = .108+.016+.072+.144 = .34
- P(!catch) =

Conditioning

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y} \mid \mathbf{z}) P(\mathbf{z}) .$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

- P(cavity) = P(cavity|catch)*P(catch) + P(cavity|!catch)*P(!catch)
- P(cavity|catch) = (.108+.072)/(.108+.016+.072+.144) = .53
- P(cavity|!catch) = (.012+.008)/(.012+.008+.064+.576) = .03
- P(catch) = .108+.016+.072+.144 = .34
- P(!catch) = .012+.064+.008+.576 = .66
- P(cavity) =

Conditioning

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y} \mid \mathbf{z}) P(\mathbf{z}) .$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

- P(cavity) = P(cavity|catch)*P(catch) + P(cavity|!catch)*P(!catch)
- P(cavity|catch) = (.108+.072)/(.108+.016+.072+.144) = .53
- P(cavity|!catch) = (.012+.008)/(.012+.008+.064+.576) = .03
- P(catch) = .108+.016+.072+.144 = .34
- P(!catch) = .012+.064+.008+.576 = .66
- P(cavity) = .53*.34+.03*.66 = .2

Conditioning

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y} \mid \mathbf{z}) P(\mathbf{z}) .$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

- P(cavity) = P(cavity|catch)*P(catch) + P(cavity|!catch)*P(!catch)
- P(cavity|catch) = (.108+.072)/(.108+.016+.072+.144) = .53
- P(cavity|!catch) = (.012+.008)/(.012+.008+.064+.576) = .03
- P(catch) = .108+.016+.072+.144 = .34
- P(!catch) = .012+.064+.008+.576 = .66
- P(cavity) = .53*.34+.03*.66 = .2 = .108+.012+.072+.008

Full Joint Distributions

- Entry for every combination of random variables
- What is the size of the table for N boolean variables?

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

Full Joint Distributions

- Entry for every combination of random variables
- Does not scale
  - (2^N) for boolean variables, much worse for non!

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

Independence

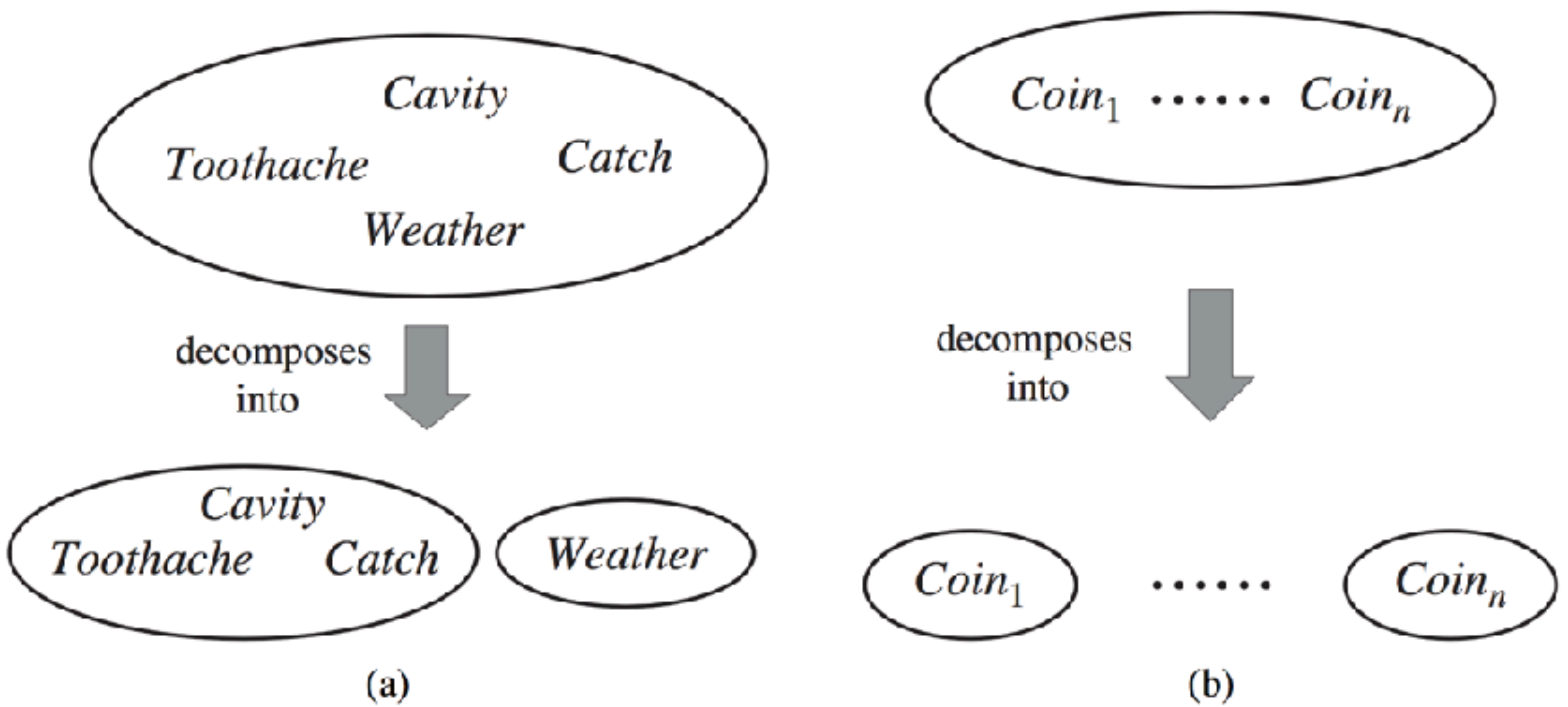


Figure 13.4 Two examples of factoring a large joint distribution into smaller distributions, using absolute independence. (a) Weather and dental problems are independent. (b) Coin flips are independent.