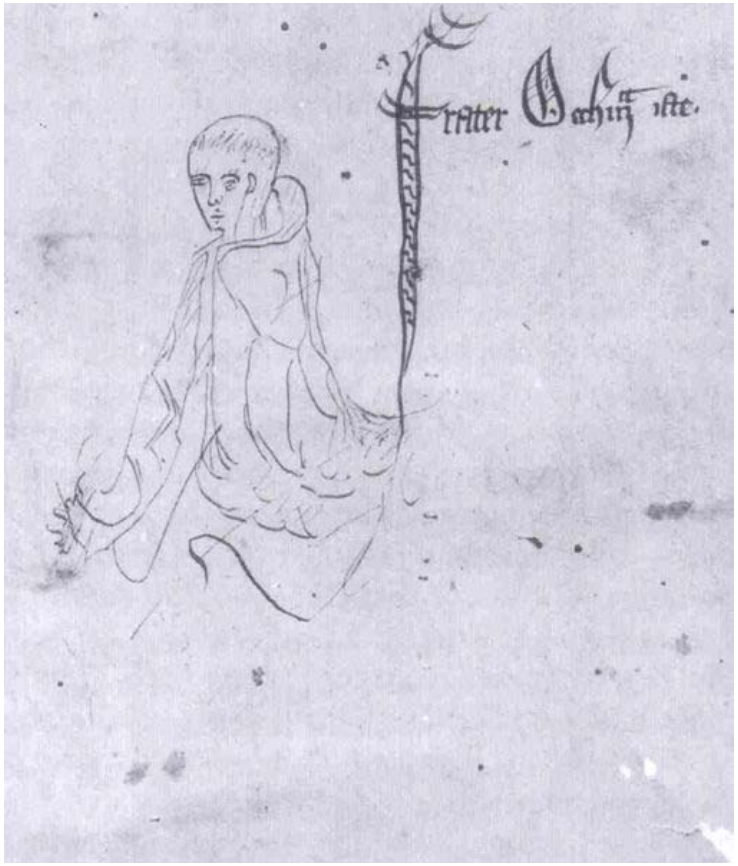
Occam’s Razor

Occam's razor (sometimes spelled Ockham's razor) is a principle attributed to the 14th-century English logician and Franciscan friar William of Ockham.

The principle states that the explanation of any phenomenon should make as few assumptions as possible, eliminating those that make no difference in the observable predictions of the explanatory hypothesis or theory.



Occam’s Razor

This is often paraphrased as "All other things being equal, the simplest solution is the best."

In other words, when multiple competing theories are equal in other respects, the principle recommends selecting the theory that introduces the fewest assumptions and postulates the fewest entities. It is in this sense that Occam's razor is usually understood.

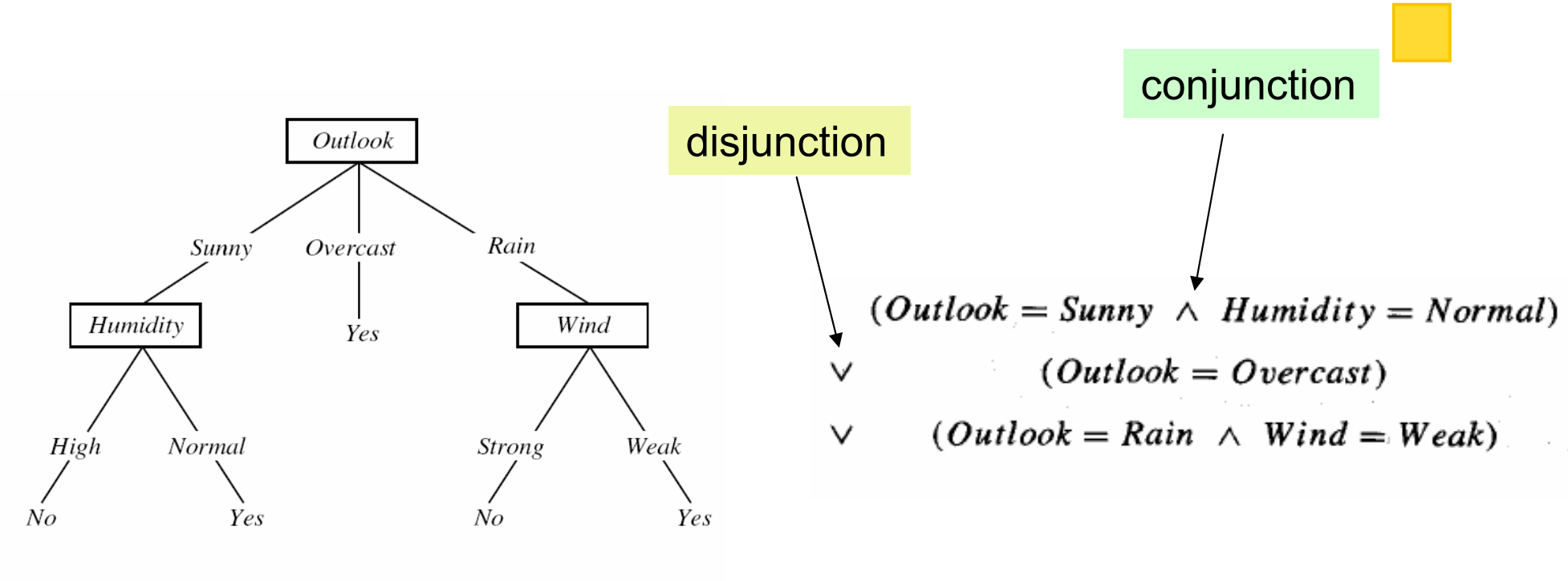
Prefer the simplest hypothesis that fits the data

Why it's called Occam's Razor

Tom M. Mitchell say's…. Occam got this idea during shaving

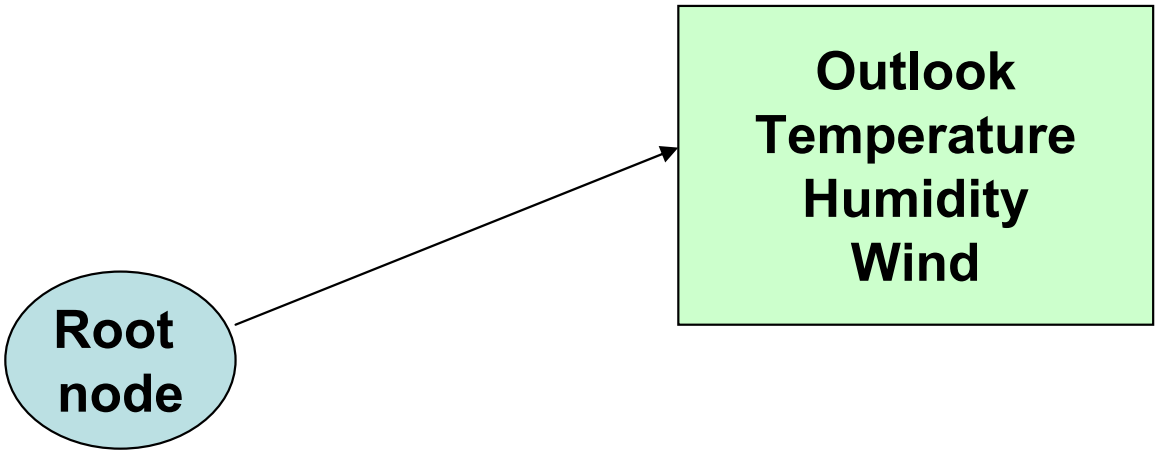
Wikipedia say's….. The term *razor* refers to the act of shaving away unnecessary assumptions to get to the simplest explanation.

Decision Trees as If-then-else rule



- If (Outlook = Sunny AND humidity = Normal) then PlayTennis = Yes
- If (Outlook = Overcast) then PlayTennis = Yes
- If (Outlook = Rain AND Wind = Weak) then PlayTennis = Yes

Building Decision Tree



Information Gain Measure

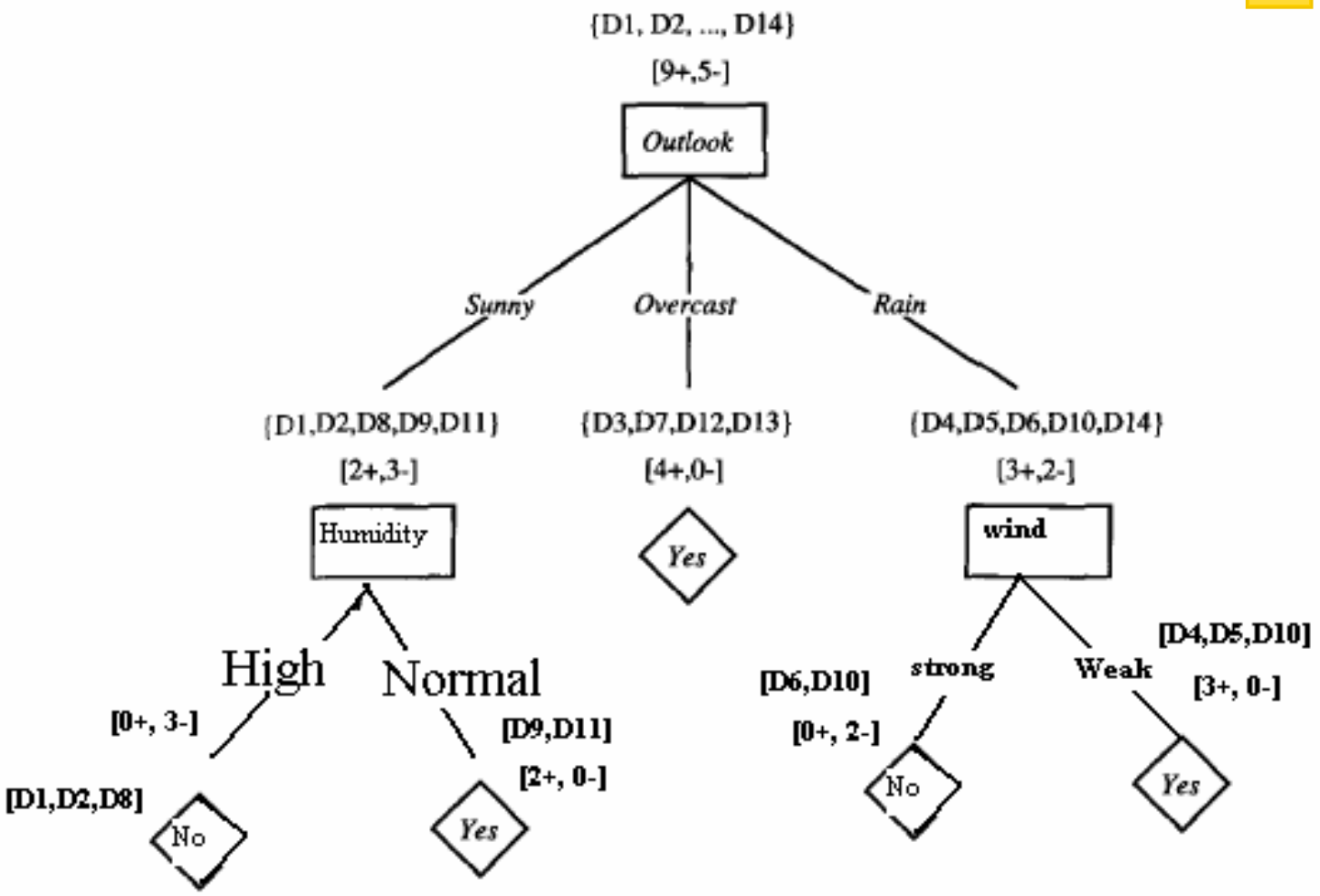
$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Entropy of S — **Entropy of S after partition**

Gain(S, A) is the expected reduction in entropy caused by knowing the value of attribute A.

Gain(S, A) is the information provided about the ***target &action value,*** given the value of some other attribute ***A.*** The value of ***Gain(S, A)*** is the number of bits saved when encoding the target value of an arbitrary member of ***S,*** by knowing the value of attribute ***A.***

Final Decision Tree



ID3 Strategy for Selecting Hypothesis

- Selects trees that place the attributes with highest information gain closest to the root.
- Selects in favor of shorter trees over longer ones.

Why a shorter, more simpler tree is preferred over larger one?

- The heart of Machine Learning.
- Learning is the process of generalization. A simple tree is likely to be more general than a complex tree. A tree that is general is probably going to be more accurate on the population at large than a tree that is not general. This is essentially what Occam's razor says. When two hypotheses equally explain the training set, pick the more general of the two.

Later formulations

- Isaac Newton:"We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances. Therefore, to the same natural effects we must, as far as possible, assign the same causes.”
- Bertrand Russell:"Whenever possible, substitute constructions out of known entities for inferences to unknown entities."

Other Applications

- **Medicine**
  - When discussing Occam's razor in contemporary medicine, doctors and philosophers of medicine speak of diagnostic parsimony. Diagnostic parsimony advocates that when diagnosing a given injury, ailment, illness, or disease a doctor should strive to look for the fewest possible causes that account for all the symptoms
- **Business management**
  - Occam razor law in business management can be further evolved into simple and complex rule: things become complex and very simple, simple things become very complicated. The law of demand, we in dealing with the matter, to grasp the essence of things mainly, grasp the mainstream, the fundamental problems, especially the need to comply with nature, don't make things artificially complicated, so that it can handle things.