Hidden Markov Models

Garrett M. Dancik, PhD

Some definitions

- The **probability** of an event *A*, denoted by $P(A)$, represents the proportion of times that *A* occurs over the long run
- For example, if we flip a coin once, $P(H) = 0.50$, indicating that we expect to get heads 50% of the time
- If all outcomes are equally likely, $P(A)$ is the proportion of outcomes where A occurs.
- The **conditional probability** of *A*, given that *B* has occurred, is denoted by $P(A|B)$and has the formula

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

Conditional probability

Class status	M	F	Total
Soph	4	2	**6**
Junior	3	2	**5**
Senior	3	1	**4**
Total	**10**	**5**	**15**

From the conditional probability formula, it is true that

$$P(A \text{ and } B) = P(A|B)P(B)$$

Example:

$$P(Female \text{ and } Soph) = P(Female|Soph)P(Soph)$$
$$= 2/6 \times 6/15$$
$$= 2/15$$

This should make sense, as there are 2 female sophomores and 15 total students

Bayes' Theorem

- $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

Class status	M	F	Total
Soph	4	2	**6**
Junior	3	2	**5**
Senior	3	1	**4**
Total	**10**	**5**	**15**

$$P(Soph|Female) = \frac{P(Female|Soph)P(Soph)}{P(Female)}$$

$$= \frac{\frac{2}{6} \times \frac{6}{15}}{\frac{5}{15}}$$

$$= \frac{2}{5}$$

This should make sense, as there are 2 female sophomores and 5 total females

(given the person is female – and there are 5 females – there are 2 sophomores)

Bayes' Theorem

- For some calculations, we do not need to calculate the denominator, $P(B)$.
- Instead, we can use the fact that

$$P(A|B) \propto P(B|A)P(A),$$

where $\propto$ means "is proportional to".

Bayes' Theorem

- A female student is selected. Is the student more likely to be a sophomore or a senior?

Class status	M	F	Total
Soph	4	2	**6**
Junior	3	2	**5**
Senior	3	1	**4**
Total	**10**	**5**	**15**

$$P(Soph|Female) \propto P(Female|Soph)P(Soph) = \frac{2}{6} \times \frac{6}{15} = \frac{2}{15}$$

$$P(Senior|Female) \propto P(Female|Senior)P(Senior) = \frac{1}{4} \times \frac{4}{15} = \frac{1}{15}$$

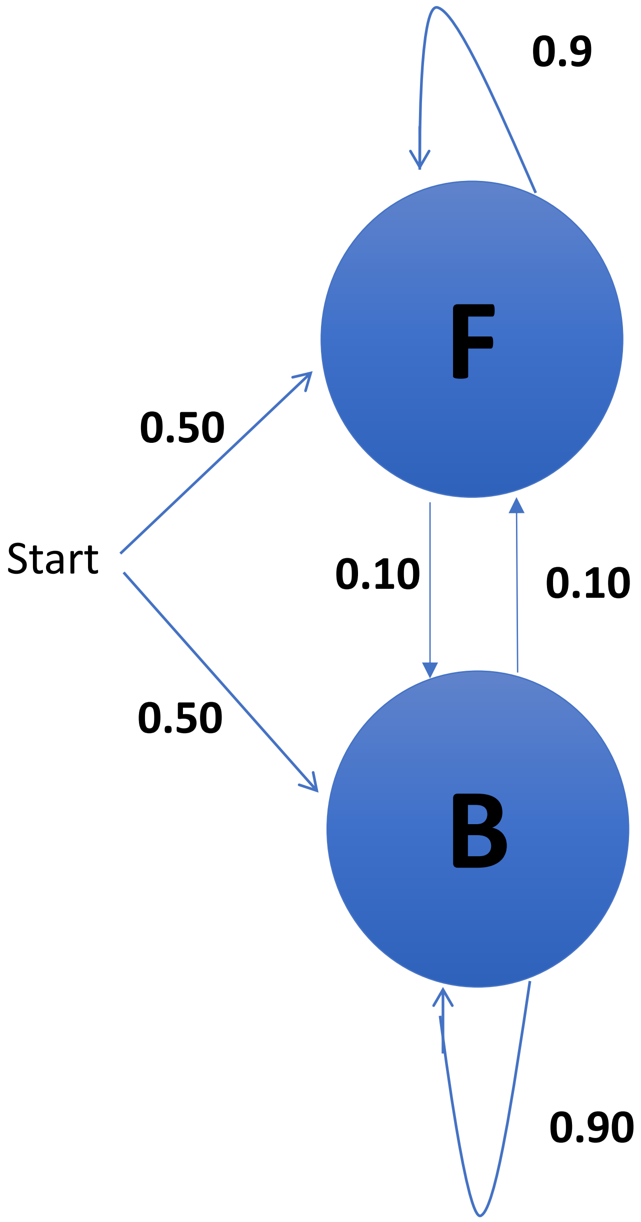
Since $P(Soph|Female) \propto \frac{2}{15}$ and $P(Senior|Female) \propto \frac{1}{15}$ the selected individual is

$\frac{P(Soph|Female)}{P(Senior|Female)} = \frac{2/15}{1/15} =$ 2 times as likely to be a sophomore than a senior

Markov chains

- A **Markov chain** is a sequence of random variables (or states) X_1, X_2, X_3, … with the property that the next state X_{n+1} depends on the m previous states (including the current one).
- Usually, m is taken to be 1 in which case the next state depends only on the current one, and the Markov chain is said to have the Markov property and is a first order Markov model.
- Formally, first order Markov models have the property

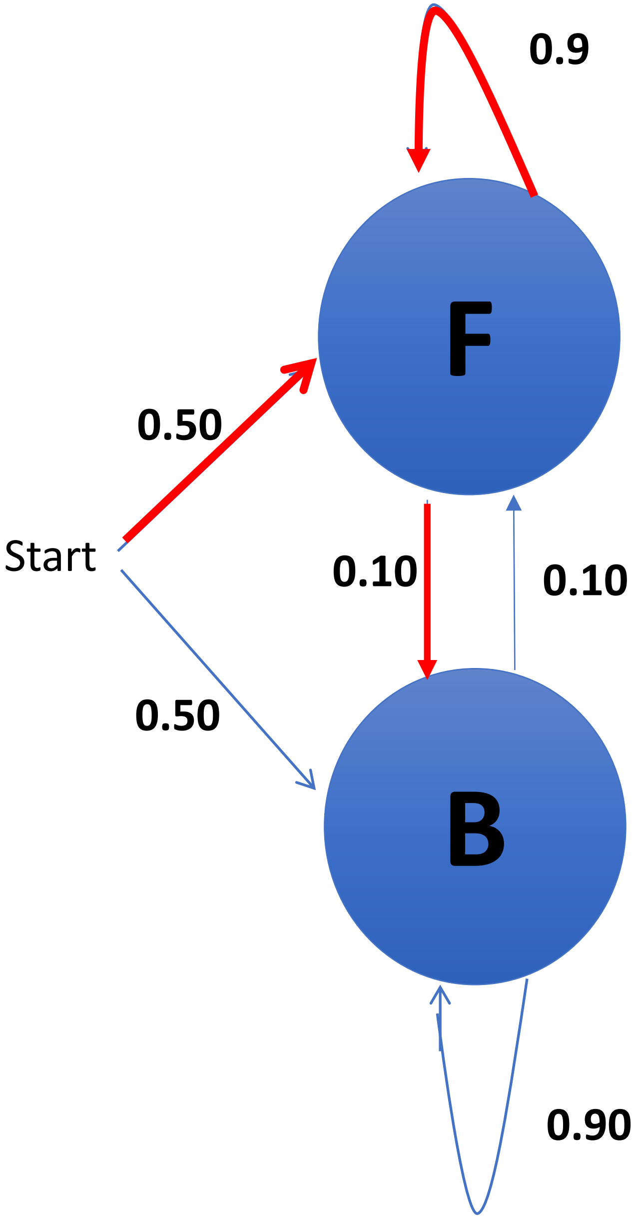
$$P(X_{n+1}|X_1, X_2, \dots, X_n) = P(X_{n+1}|\ X_n)$$



Example:

- An individual has two coins, a fair (F) coin and a biased (B)coin.
- Before each coin toss, there is a 10% chance that the individual will switch coins.
- Initially, there is a 50% chance the individual selects the fair (or biased) coin.
- Find the probability that the selected coins are FFB.
- Note that this is a 1st order Markov Chain. The subscript *i* will be used for the *i*th selected coin

$$
\begin{aligned}
P(F_1 F_2 B_3) &= P(F_1) P(F_2|F_1) P(B_3|F_2, F_1) = \\
&= P(F_1)\, P(F_2|F_1)\, P(B_3|F_2) \quad \text{(Markov assumption)} \\
&= 0.50 \times 0.90 \quad \times\ 0.10 \\
&= 0.045
\end{aligned}
$$



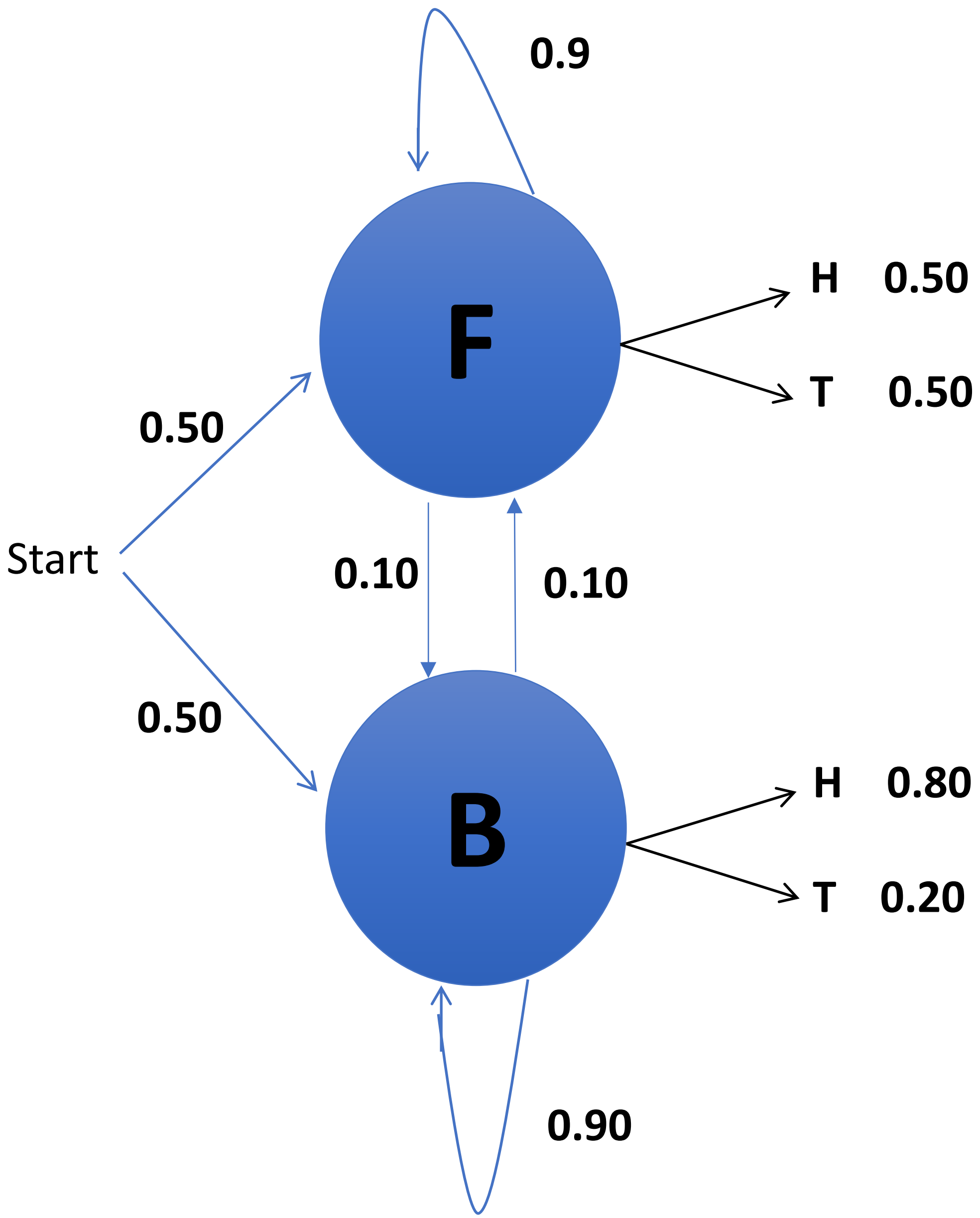
- Find the probability that the selected coins are FFB.
- We don't need the fancy notation. Just follow the arrows through the Markov Model and write the corresponding probabilities

State	Start	→ F	→ F	→B
Prob	-	0.50	0.90	0.10

- Now multiply the probabilities together to get the probability of the state (e.g., FFB)

$$
\begin{aligned}
\mathrm{P(FFB)} &= 0.50 \times 0.90 \quad \times\ 0.10 \\
&= 0.045
\end{aligned}
$$

Hidden Markov models



- A **Hidden Markov model** (HMM) is a Markov chain where the states are hidden (unobserved)
- Hidden states emit observed values with certain probabilities.
- The hidden states can then be deduced based on the observed values.

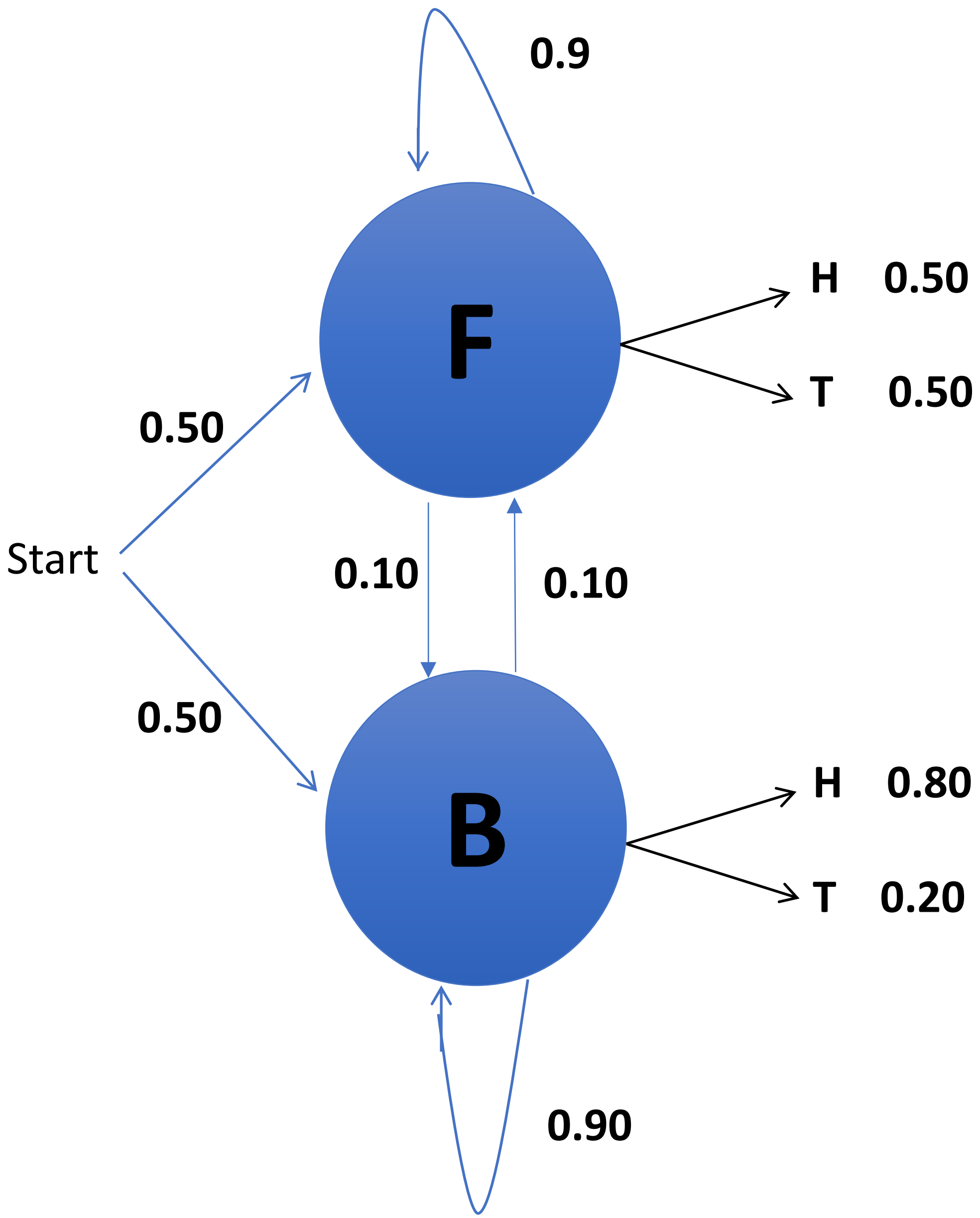
- General HMM notation:

Observed states: $O_1, O_2, \dots O_n$

Hidden states: $H_1, H_2, \dots H_n$

Initial state probabilities:	$P(H_1)$
Transition state probabilities:	$P(H_{i+1}\|H_i)$ for all i
Emission probabilities:	$P(O_i\|H_i)$ for all i

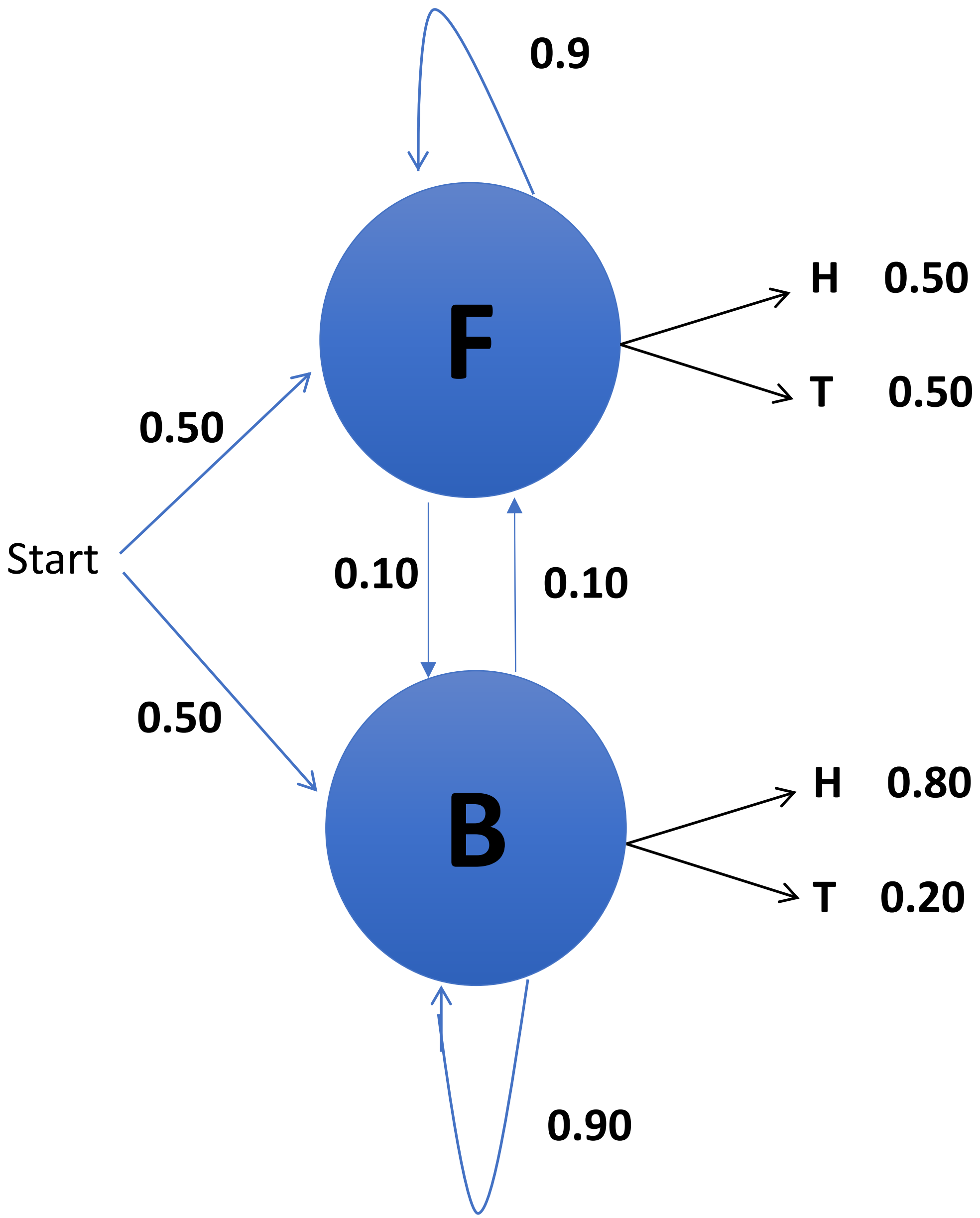
Hidden Markov models



- We can specify a HMM graphically or by specifying the relevant probabilities:

Initial state probabilities:	$P(F_1) = P(B_1) = 0.50$
Transition state probabilities:	$P(F_{i+1}\|F_i) = 0.90, i > 1$ $P(B_{i+1}\|F_i) = 0.10, i > 1$ $P(F_{i+1}\|B_i) = 0.10, i > 1$ $P(B_{i+1}\|B_i) = 0.90, i > 1$
Emission probabilities:	$P(H\|F_i) = 0.50$ for all i $P(T\|F_i) = 0.50$ for all i $P(H\|B_i) = 0.80$ for all i $P(T\|B_i) = 0.20$ for all i

Hidden Markov models



- Three coins are flipped and HTH is observed. Which is more likely, that the coins were FFF or BFB?
- How likely is it that the coins were FFF, given that we observe HTH?
- We start by calculating $P(F_1 F_2\ F_3|H_1\ T_1\ H_1)$

$$P(F_1 F_2\ F_3|H_1\ T_1\ H_1) \propto$$

$$P(H_1|\ F_1)P(F_1)$$

$$(0.50 \times 0.50)$$

Start with fair coin, flip it, get heads

$$\times \quad P(T_2|\ F_2)P(F_2|F_1)$$

$$(0.50 \times 0.90)$$

Keep fair coin, flip it, get tails

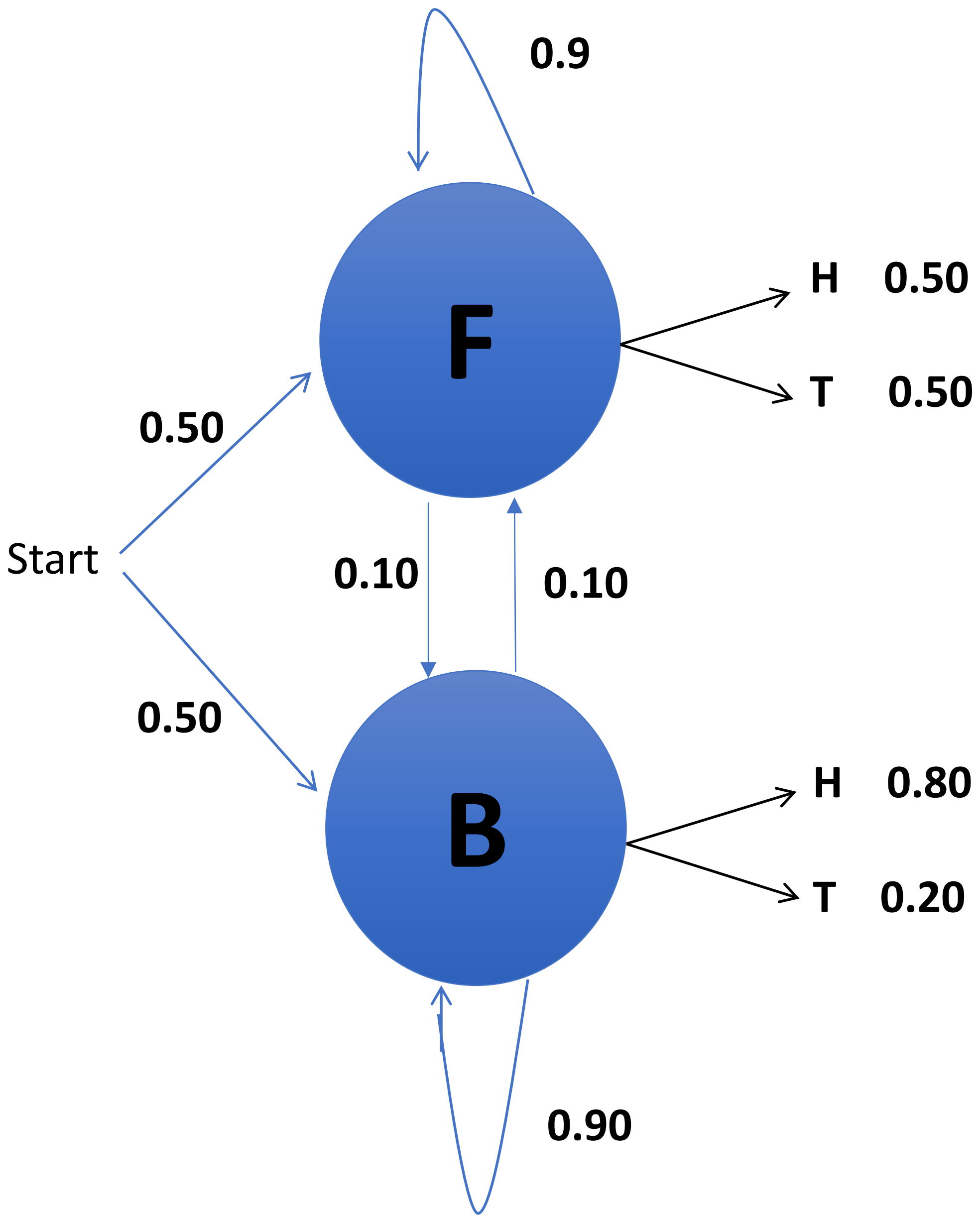
$$\times \quad P\ (H_3|\ F_3)P(F_3|F_2)$$

$$(0.50 \times 0.90)$$

Keep fair coin, flip it, get heads

$$= 0.050625$$

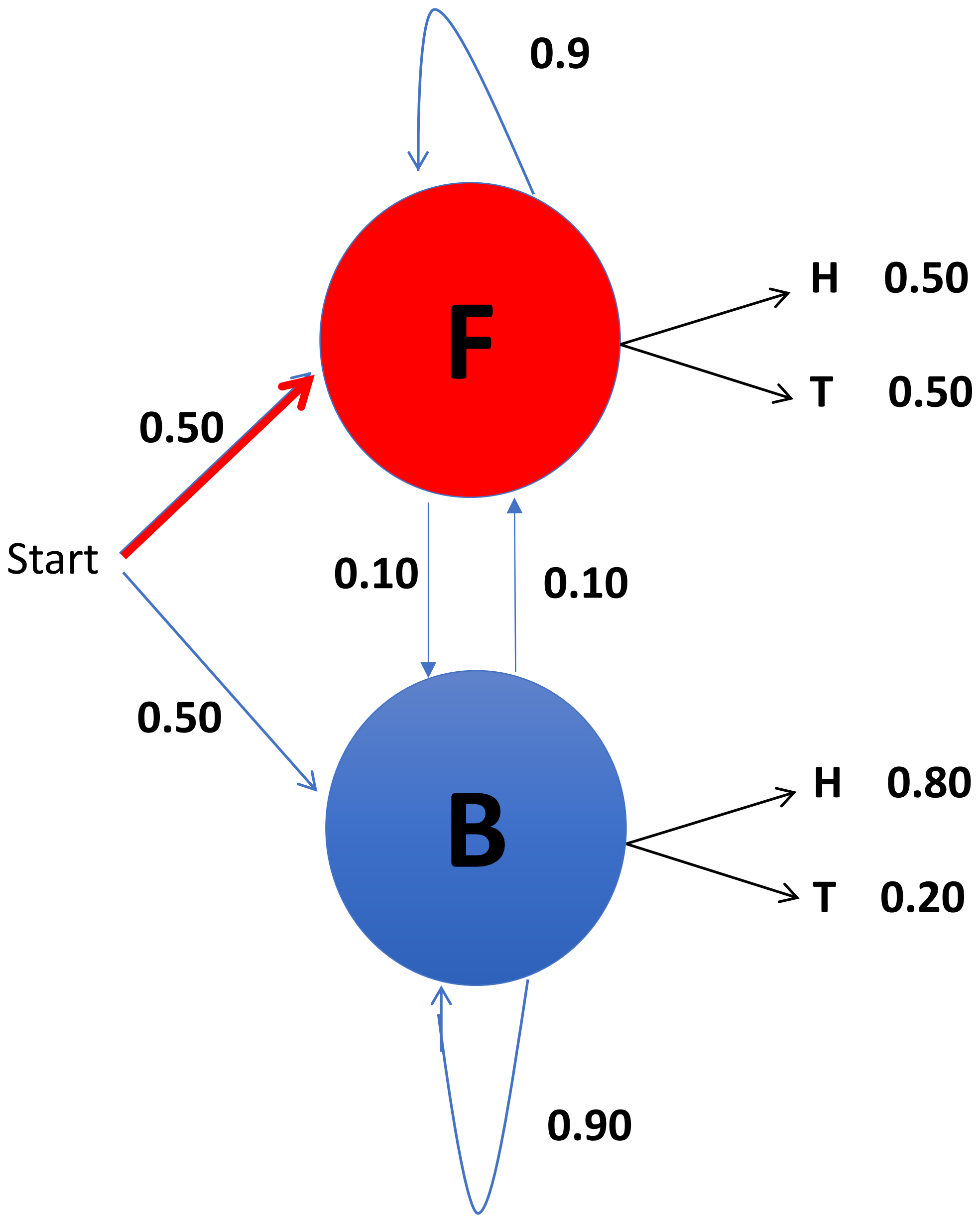
Hidden Markov models – P(FFF | HTH)



- How likely is it that the coins were FFF, given that we observe HTH?

Prob (state)				
State	Start	→ F	→ F	→F
Emission (observation)		H	T	H
Prob (observed)	-			

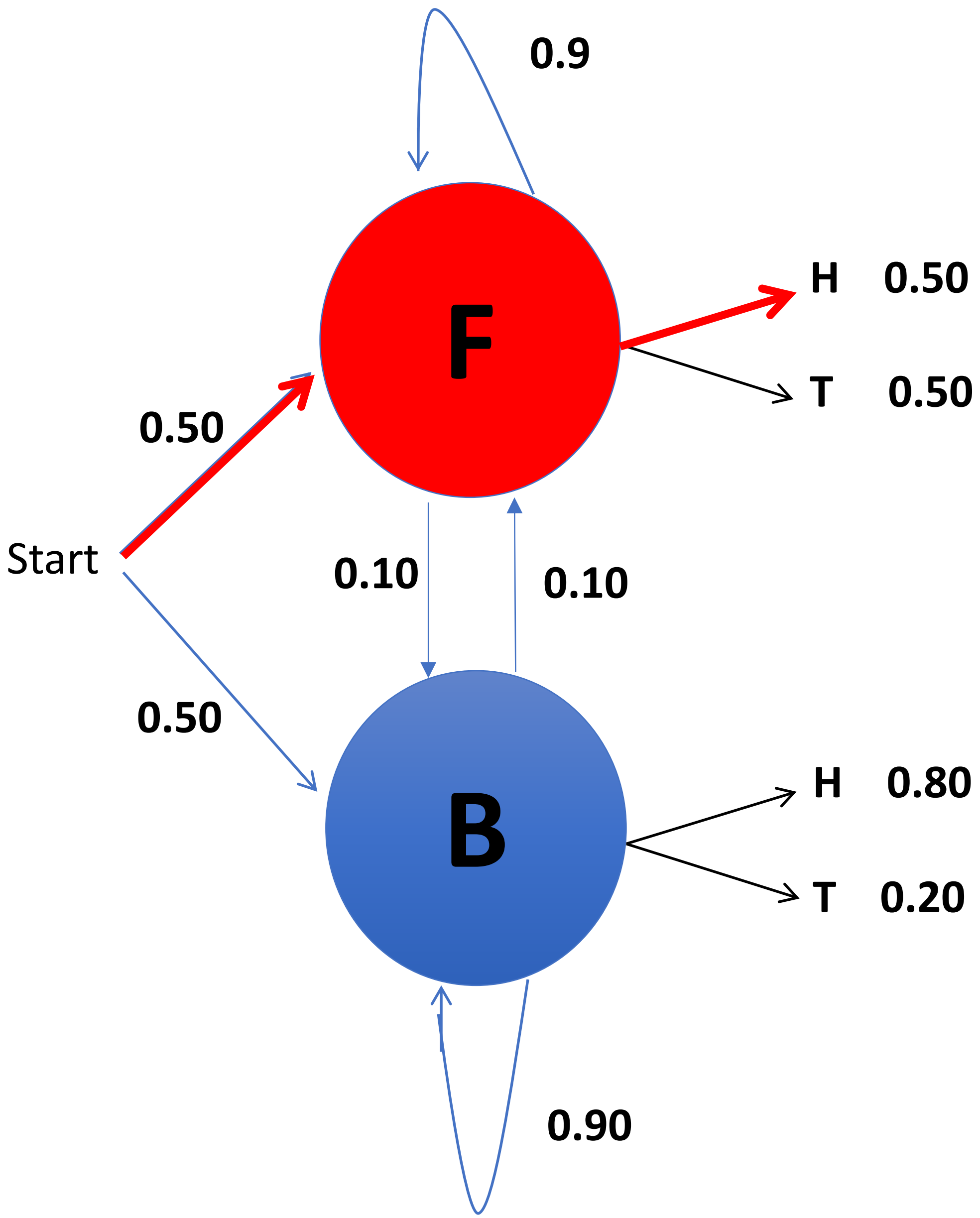
Hidden Markov models – P(FFF | HTH)



- How likely is it that the coins were FFF, given that we observe HTH?

Prob (state)		0.50		
State	Start	→ F	→ F	→F
Emission (observation)		H	T	H
Prob (observed)	-			

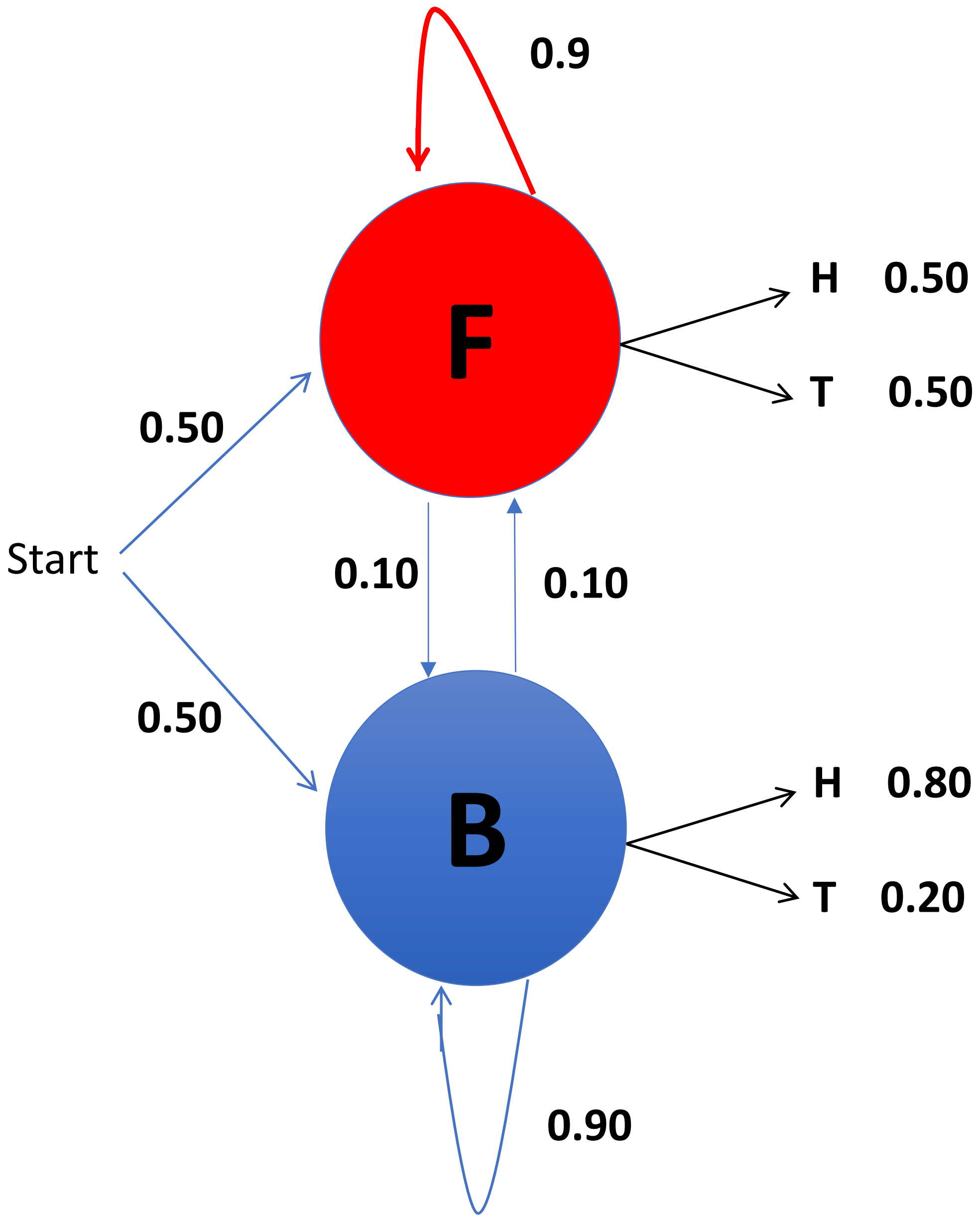
Hidden Markov models – P(FFF | HTH)



- How likely is it that the coins were FFF, given that we observe HTH?

Prob (state)		0.50		
State	Start	→ F	→ F	→F
Emission (observation)		H	T	H
Prob (observed)	-	0.50		

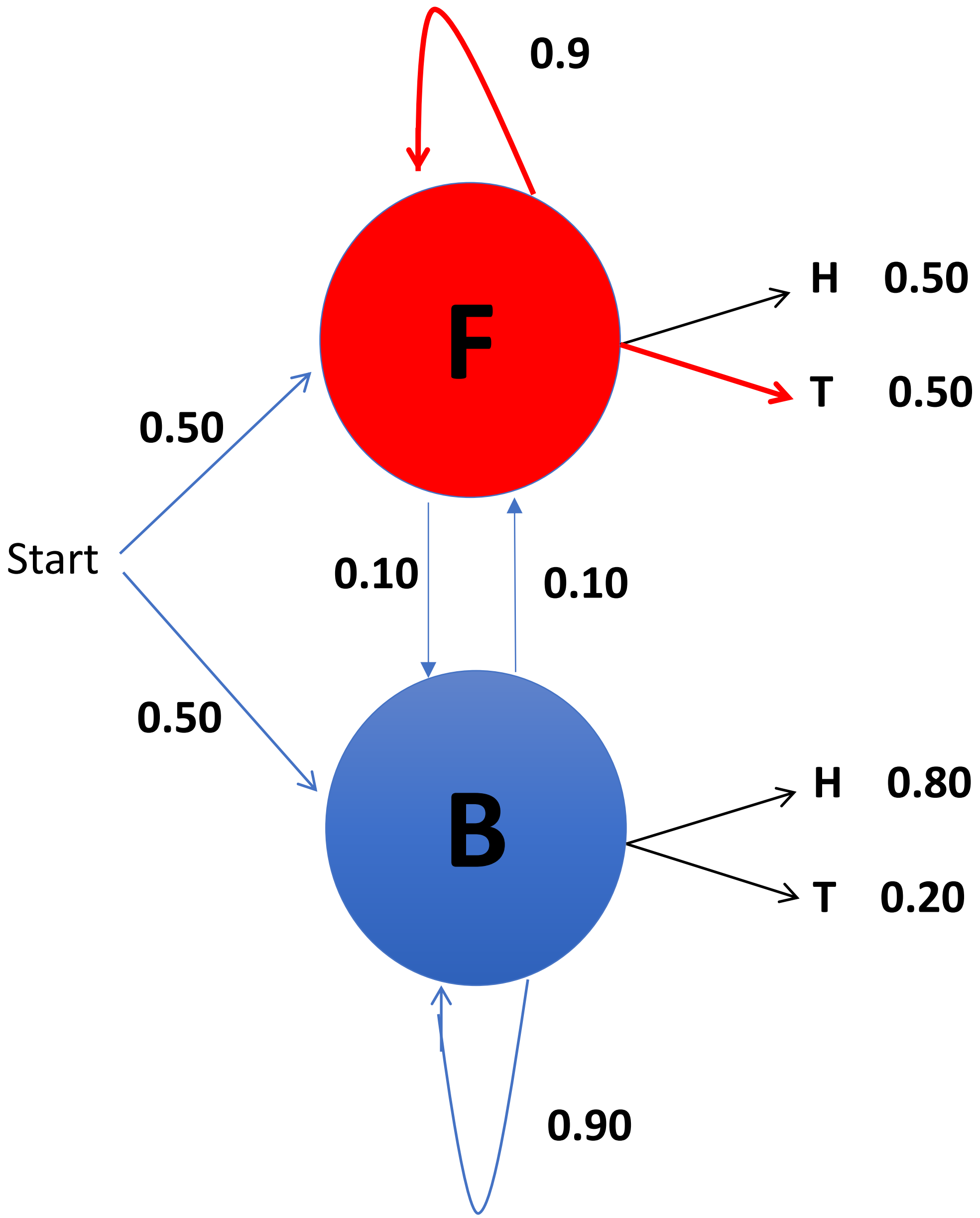
Hidden Markov models – P(FFF | HTH)



- How likely is it that the coins were FFF, given that we observe HTH?

Prob (state)		0.50	0.90	
State	Start	→ F	→ F	→F
Emission (observation)		H	T	H
Prob (observed)	-	0.50		

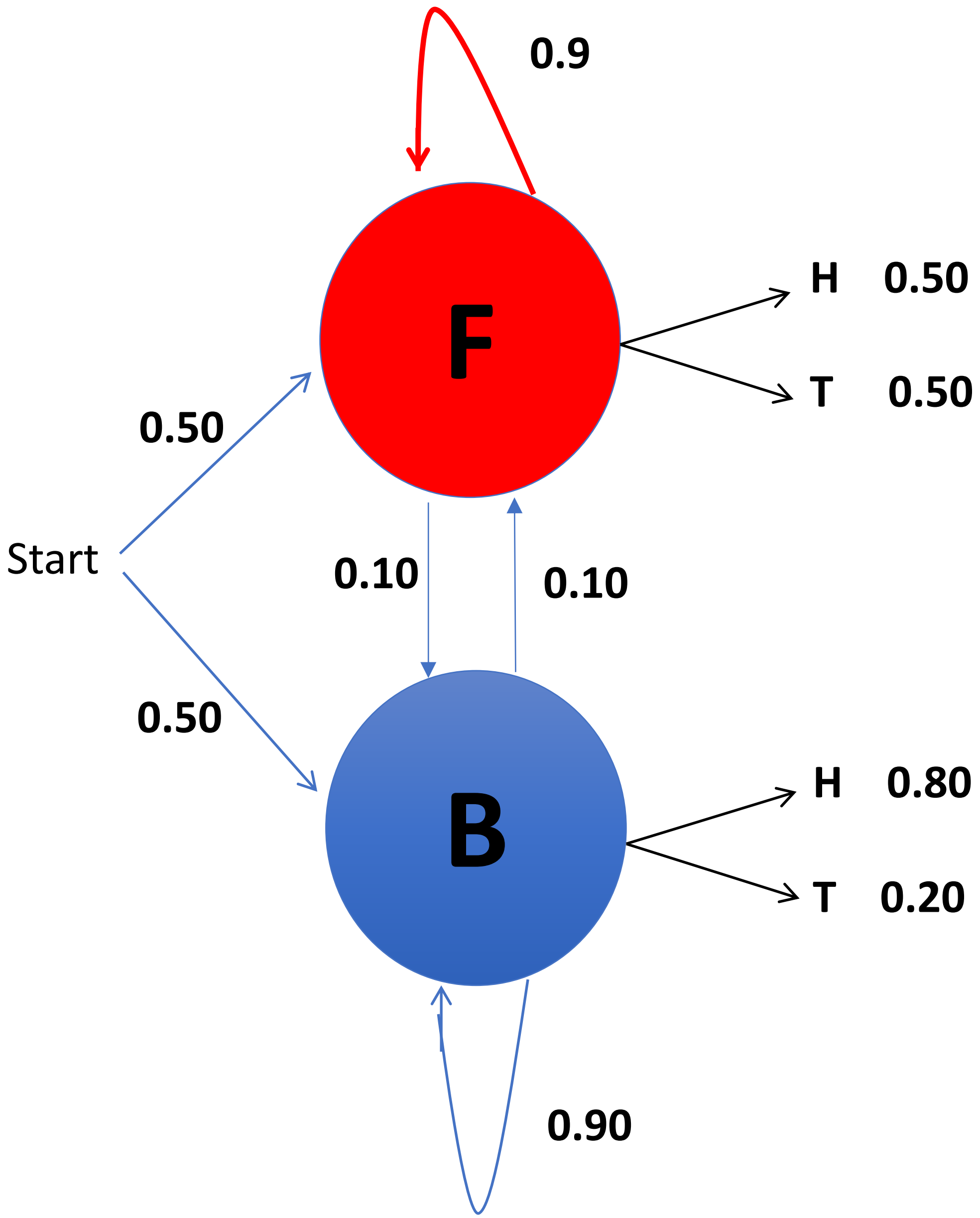
Hidden Markov models – P(FFF | HTH)



- How likely is it that the coins were FFF, given that we observe HTH?

Prob (state)		0.50	0.90	
State	Start	→ F	→ F	→F
Emission (observation)		H	T	H
Prob (observed)	-	0.50	0.50	

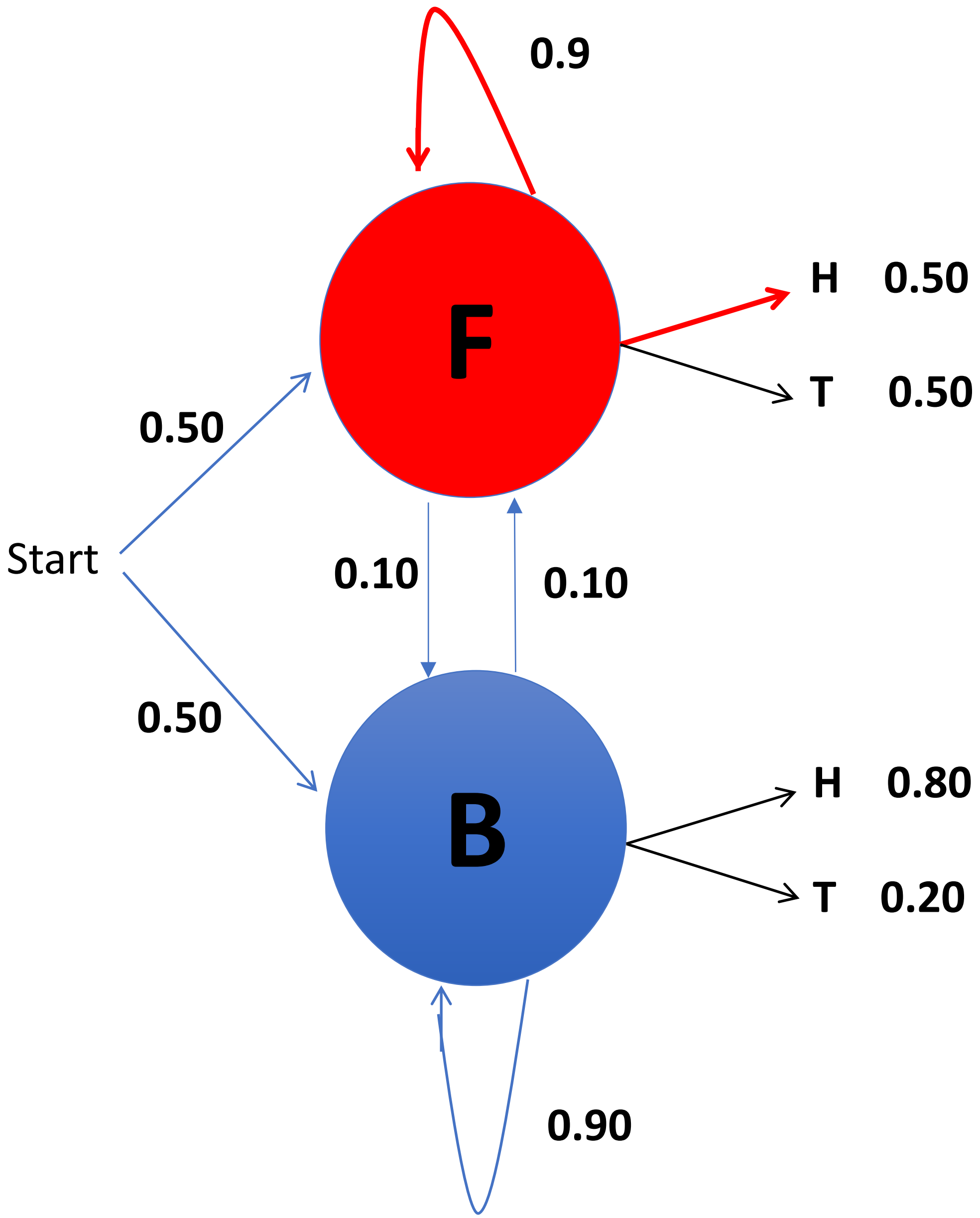
Hidden Markov models – P(FFF | HTH)



- How likely is it that the coins were FFF, given that we observe HTH?

Prob (state)		**0.50**	**0.90**	**0.90**
State	Start	→ F	→ F	→ F
Emission (observation)		H	T	H
Prob (observed)	-	**0.50**	**0.50**	

Hidden Markov models – P(FFF | HTH)



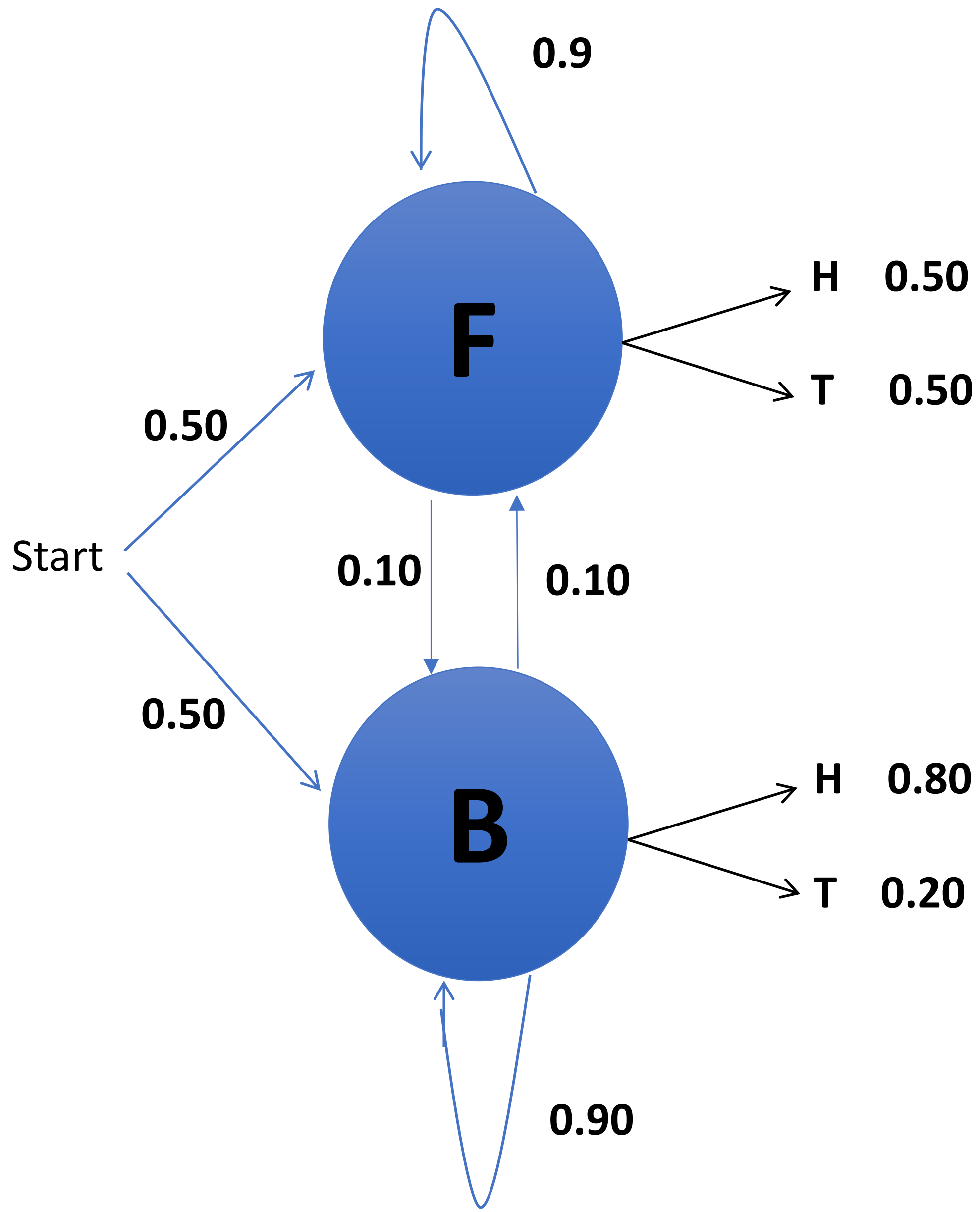
- How likely is it that the coins were FFF, given that we observe HTH?

Prob (state)		**0.50**	**0.90**	**0.90**
State	Start	→ F	→ F	→F
Emission (observation)		H	T	H
Prob (observed)	-	**0.50**	**0.50**	**0.50**

- Now multiply all the probabilities together to get the probability of the state, given the observations:

- $$P(FFF|HTH) \propto 0.50 \times 0.50 \times 0.90 \times 0.50 \times 0.90 \times 0.50$$
$$= 0.50^4 \times 0.90^2$$
$$= 0.050625$$

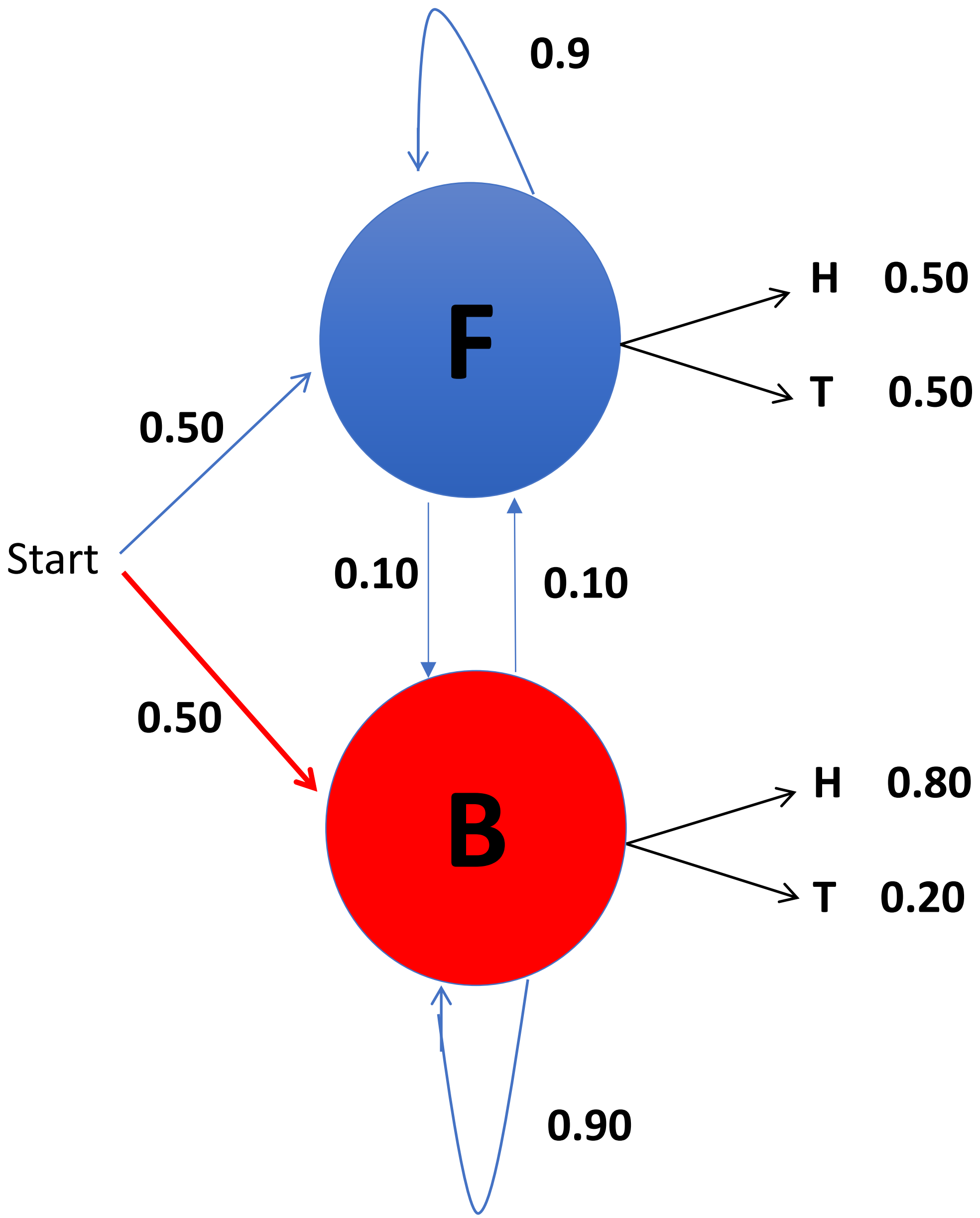
Hidden Markov models – P(BFB | HTH)



- How likely is it that the coins were BFB, given that we observe HTH?

Prob (state)				
State	Start	→ B	→ F	→ B
Emission (observation)		H	T	H
Prob (observed)	-			

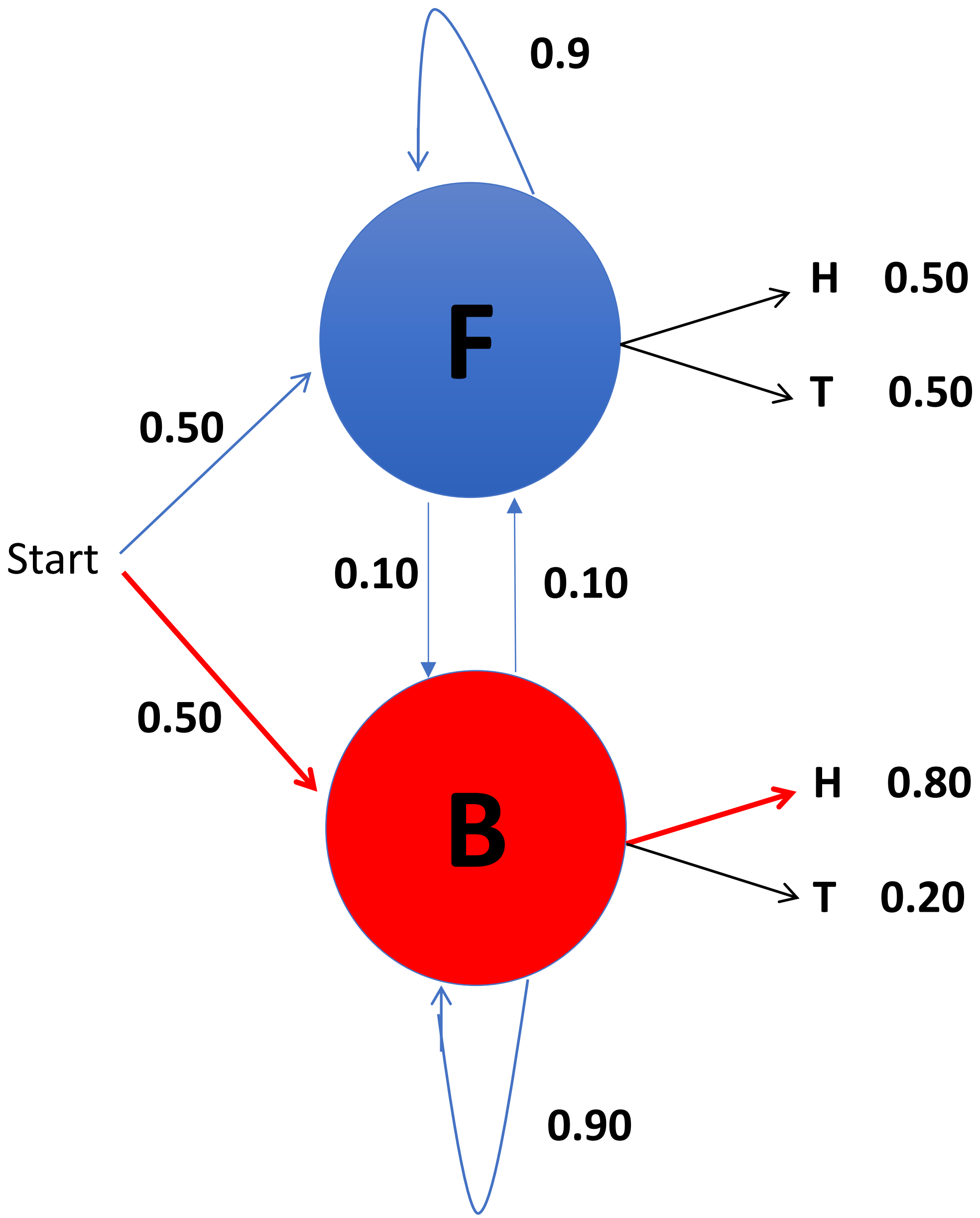
Hidden Markov models – P(BFB | HTH)



- How likely is it that the coins were BFB, given that we observe HTH?

Prob (state)		**0.50**		
State	Start	→ B	→ F	→B
Emission (observation)		H	T	H
Prob (observed)	-			

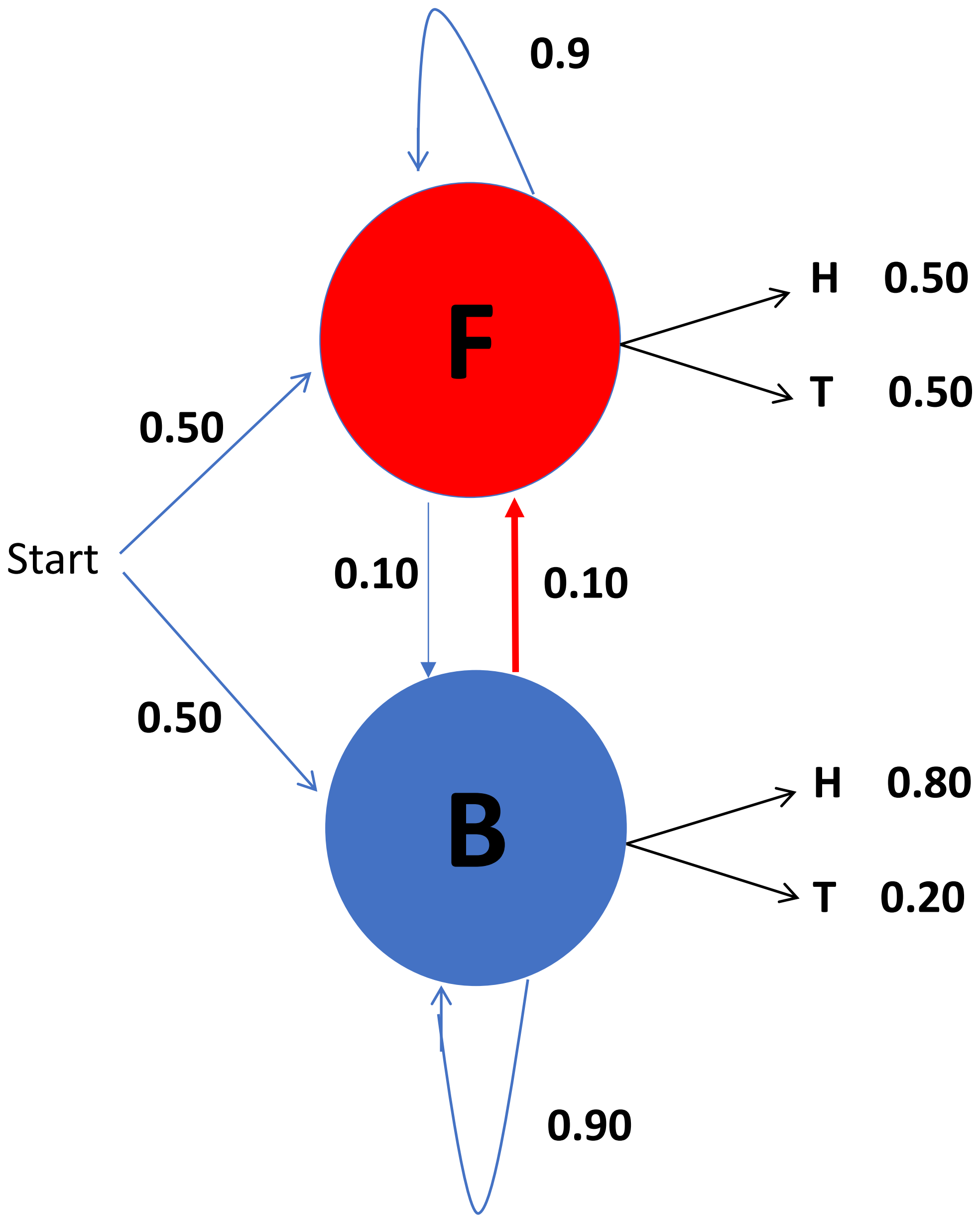
Hidden Markov models – P(BFB | HTH)



- How likely is it that the coins were BFB, given that we observe HTH?

Prob (state)		0.50		
State	Start	→ B	→ F	→B
Emission (observation)		H	T	H
Prob (observed)	-	0.80		

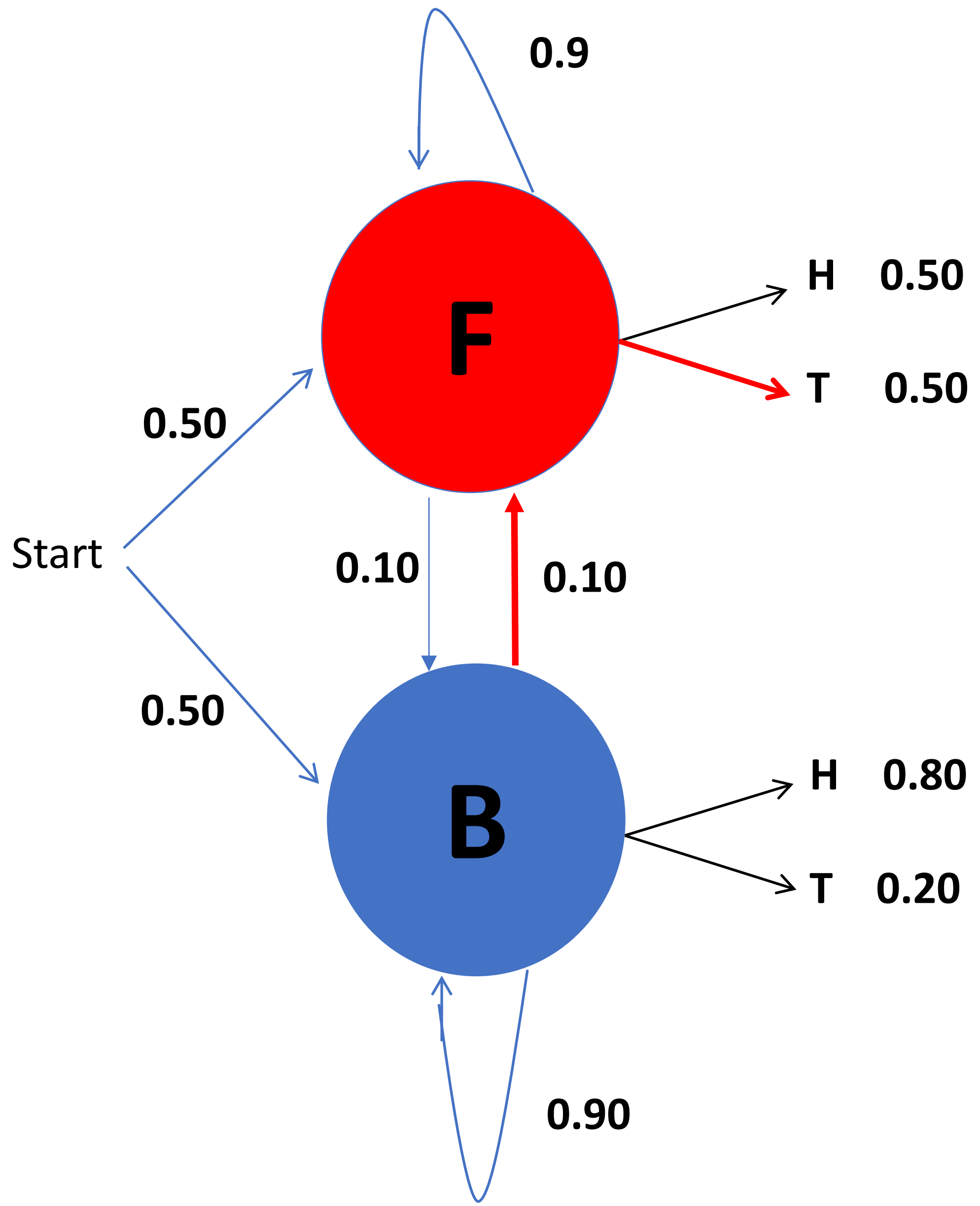
Hidden Markov models – P(BFB | HTH)



- How likely is it that the coins were BFB, given that we observe HTH?

Prob (state)		0.50	0.10	
State	Start	→ B	→ F	→B
Emission (observation)		H	T	H
Prob (observed)	-	0.80		

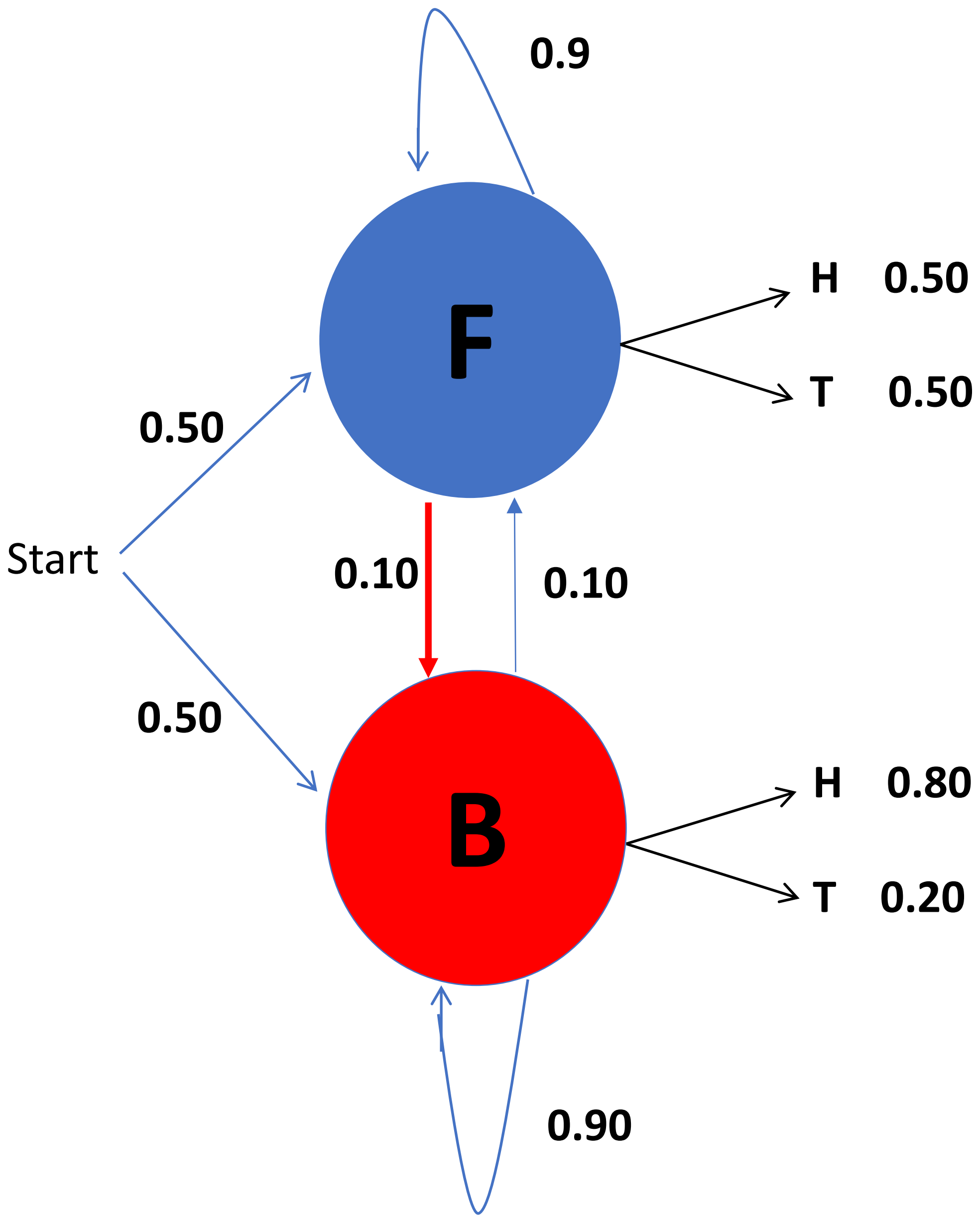
Hidden Markov models – P(BFB | HTH)



- How likely is it that the coins were BFB, given that we observe HTH?

Prob (state)		0.50	0.10	
State	Start	→ B	→ F	→B
Emission (observation)		H	T	H
Prob (observed)	-	0.80	0.50	

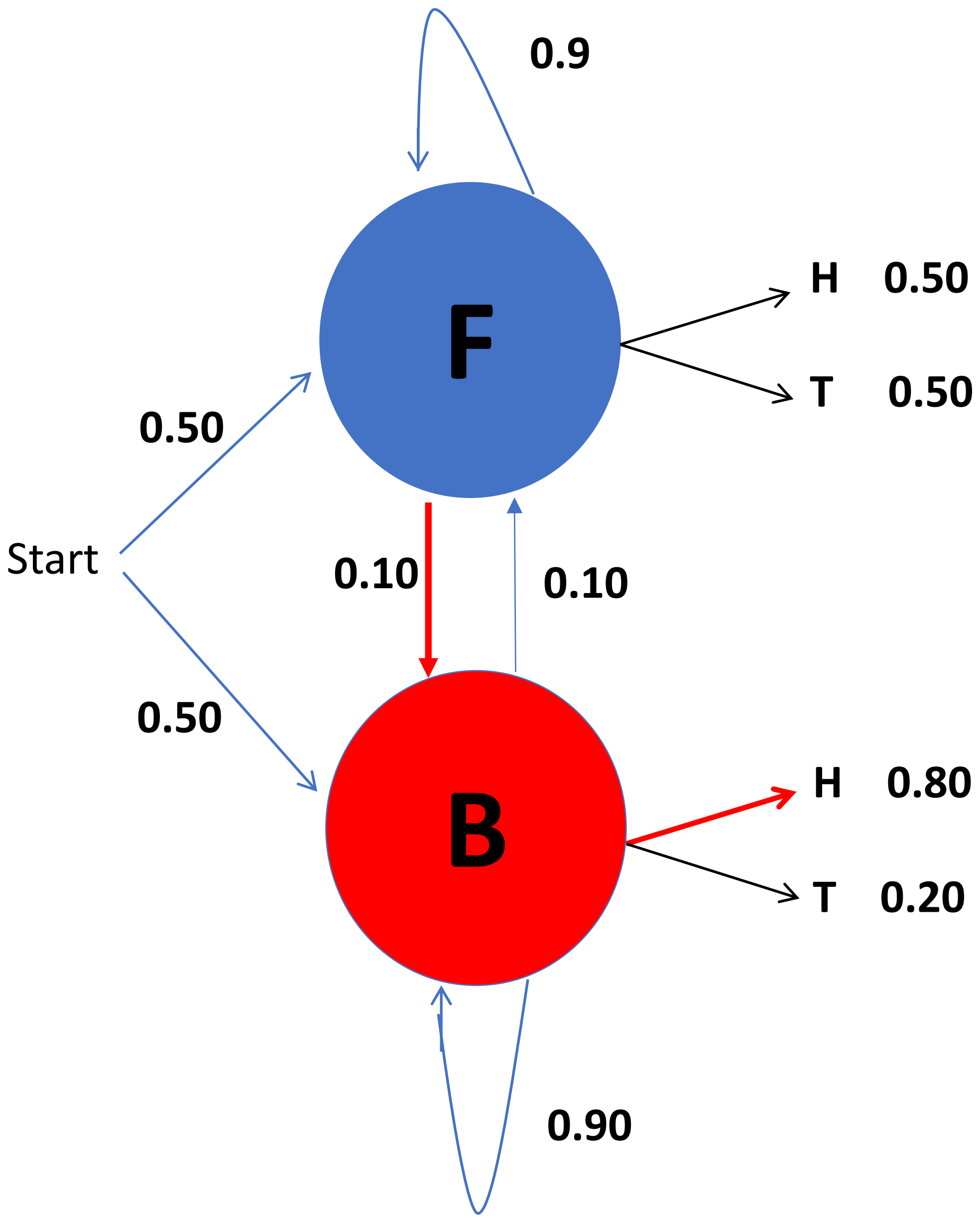
Hidden Markov models – P(BFB | HTH)



- How likely is it that the coins were BFB, given that we observe HTH?

Prob (state)		**0.50**	**0.10**	**0.10**
State	Start	→ B	→ F	→ B
Emission (observation)		H	T	H
Prob (observed)	-	**0.80**	**0.50**	

Hidden Markov models – P(BFB | HTH)



- How likely is it that the coins were BFB, given that we observe HTH?

Prob (state)		0.50	0.10	0.10
State	Start	→ B	→ F	→ B
Emission (observation)		H	T	H
Prob (observed)	-	0.80	0.50	0.80

- Now multiply all the probabilities together to get the probability of the state, given the observations:

- $P(BFB|HTH) \propto 0.50 \times 0.80 \times 0.10 \times 0.50 \times 0.10 \times 0.80$

$$= 0.50^2 \times 0.80^2 \times 0.10^2$$

$$= 0.0016$$

Which is more likely?

- $P(FFF|HTH) \propto 0.050625$
- $P(BFB|HTH) \propto 0.0016$

- $\frac{P(FFF|HTH)}{P(BFB|HTH)} = \frac{0.050625}{0.0016} \approx 31.64$

If we observe *HTH*, we are about 32 times more likely to have flipped only the fair coin (FFF) than the biased, fair, and biased (BFB) coins.

But what about other possible states, such as BBB, BFF, etc?

Hidden Markov Models

- The goal of an HMM is to find the set of hidden states (such as the gene structure), which is unknown.
- We can (almost) never be certain, but the most likely set of hidden states is the state sequence $H_1, H_2, \ldots, H_n$ that maximizes

$$P(H_1, \ldots, H_n \mid O_1, \ldots, O_n) \propto$$
$$P(O_1|H_1)P(H_1) \times P(O_2|H_2)P(H_2|H_2) \times \cdots \times P(O_n|H_n)P(H_n|H_{n-1})$$

- Most of the time, we work with probabilities on the log scale, where the log of a product is equal to the sum of the logs.

Prob (state)		0.50	0.10	0.10
State	Start	→ B	→ F	→ B
Emission (observation)		H	T	H
Prob (observed)	-	0.80	0.50	0.80

- $P(BFB|HTH) \propto 0.50 \times 0.80 \times 0.10 \times 0.50 \times 0.10 \times 0.80$
  $= 0.50^2 \times 0.80^2 \times 0.10^2$
  $= 0.0016$

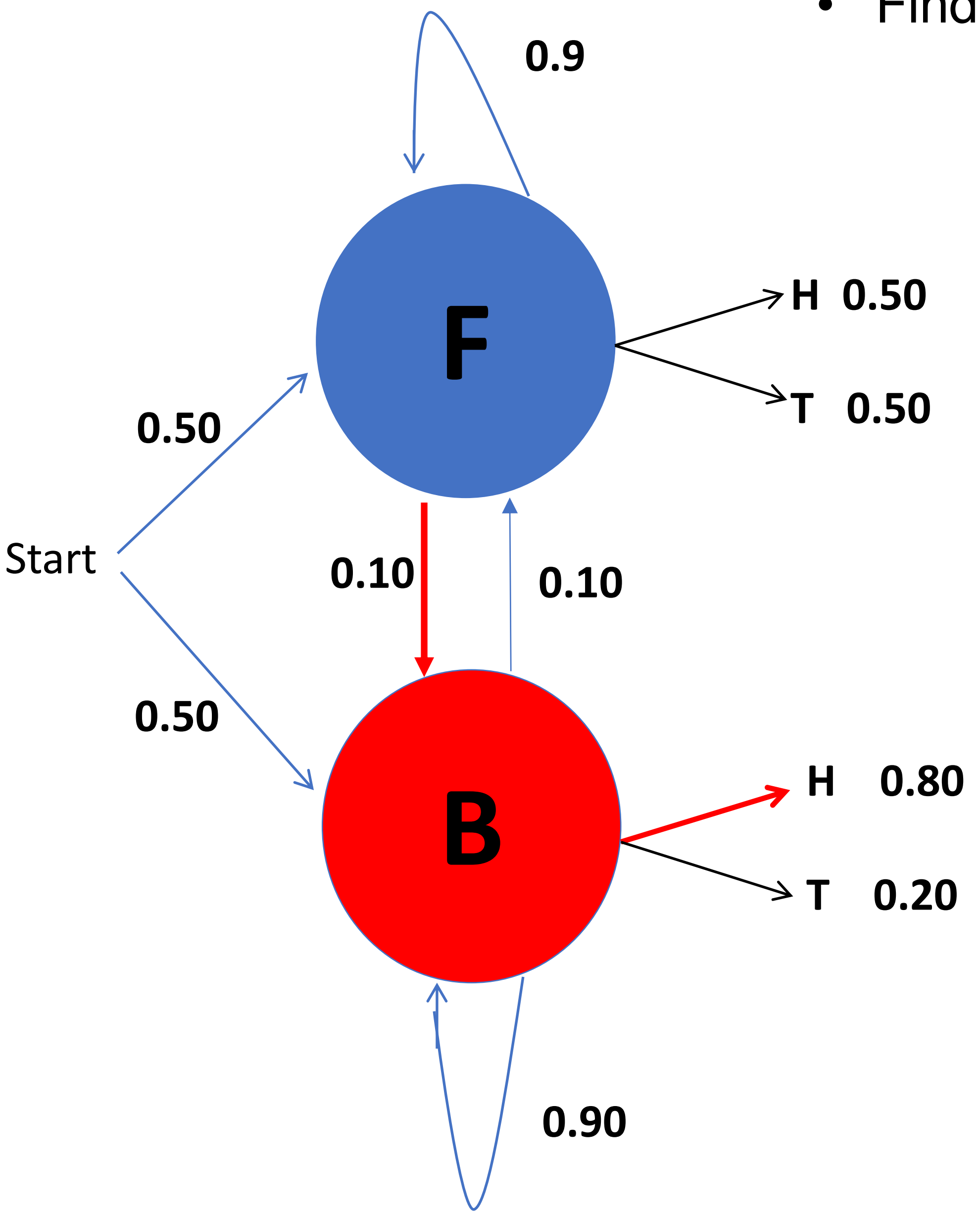
Prob (state) (log scale)		log(0.50)	log(0.10)	log(0.10)
State	Start	→ B	→ F	→ B
Emission (observation)		H	T	H
Prob (observed) (log scale)	-	log(0.80)	log(0.50)	log(0.80)

- $P(BFB|HTH) \propto \log(0.50) + \log(0.80) + \log(0.10) + \log(0.50) + \log(0.10) + \log(0.80)$
  $= 2\times \log(0.50) + 2 \times \log(0.80) + 2\times \log(0.10)$
  $= —$-2.79588 (note that $10^{-2.79588} = 0.0016$)

Viterbi algorithm

- How do we determine the optimal sequence of hidden states?
- Let's continue with our coin tossing example, where the hidden state sequence ends with either *F* or *B*.
- Suppose we know the optimal hidden states for the first two observations, ending with *F* or *B*. Then there are 4 possibilities for the next hidden state:
  - F → F
  - B → F
  - F → B
  - B → B
- This lends itself to a dynamic programming solution, known as the **Viterbi algorithm**.

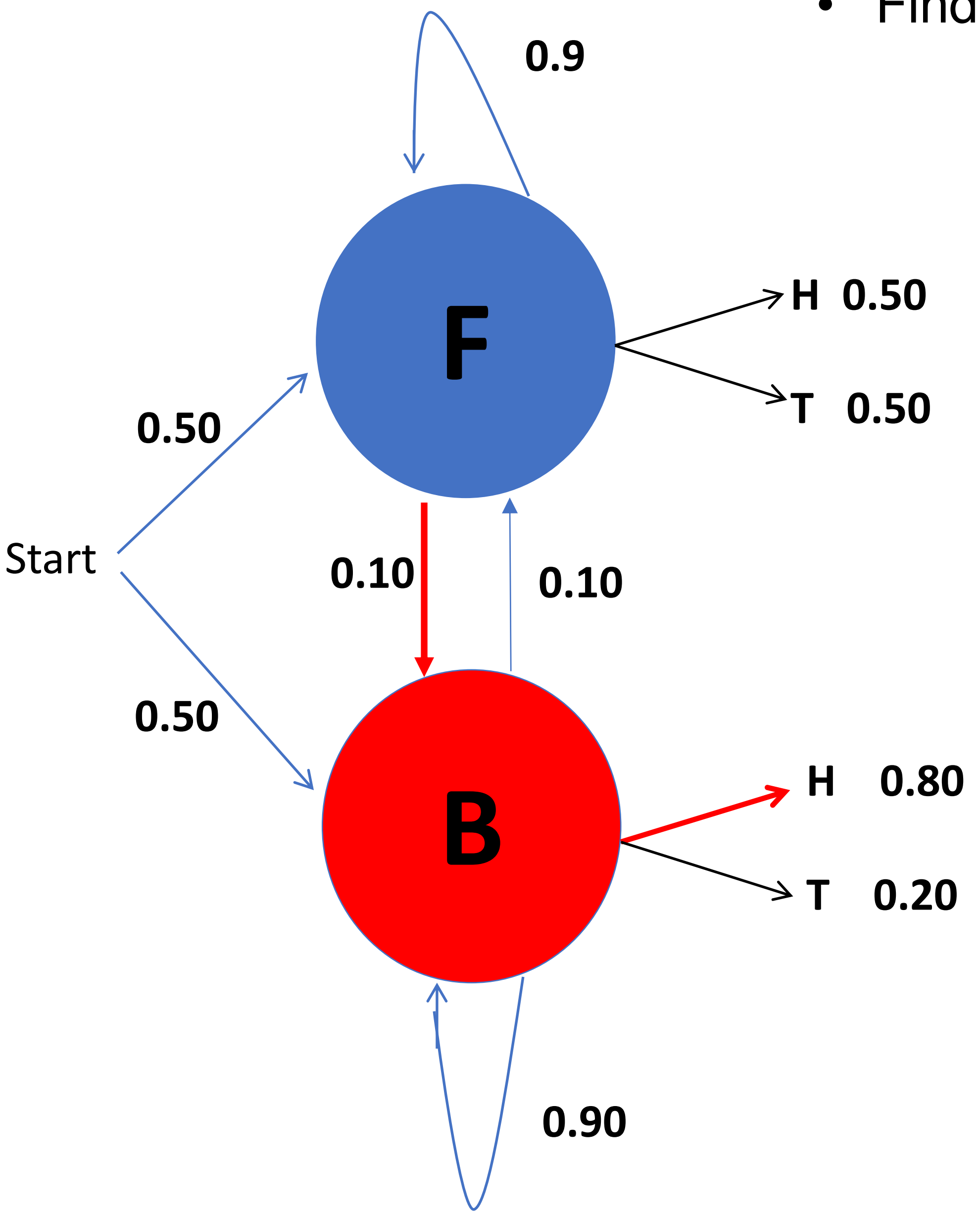
Hidden Markov models



- Find the optimal state that would generate HTH.

	H	T	H
F	0.50(0.50)=0.25		
B	0.50(0.80)=0.40		

Hidden Markov models



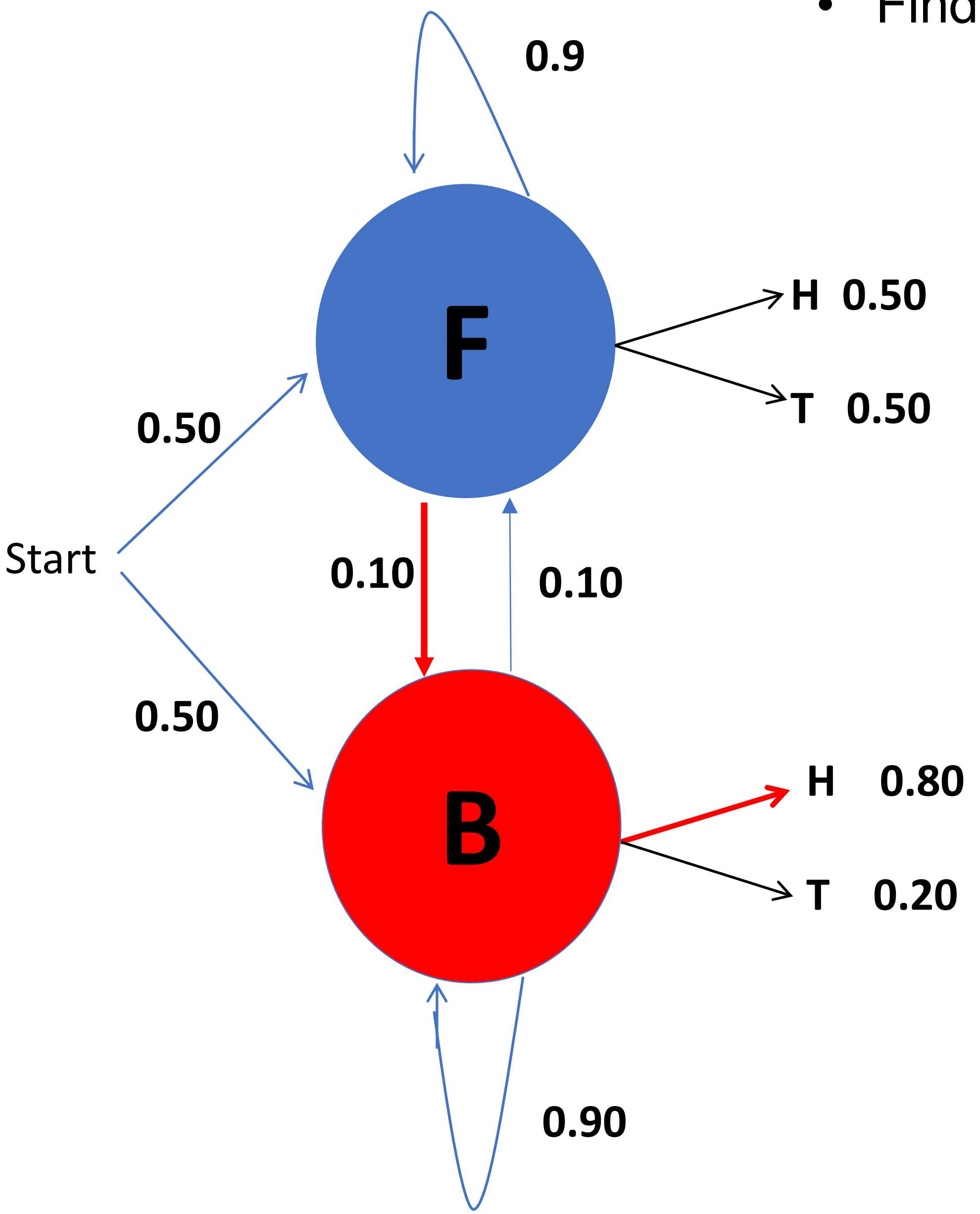
- Find the optimal state that would generate HTH.

	H	T	H
F	0.50(0.50)=0.25 →	0.1125*	
B	0.50(0.80)=0.40 →	0.072*	

$F \rightarrow F$: $0.25 \times 0.90(0.50) = 0.1125^*$
$B \rightarrow F$: $0.40 \times 0.10(0.50) = 0.02$

$F \rightarrow B$: $0.25 \times 0.10(0.20) = 0.005$
$B \rightarrow B$: $0.40 \times 0.90(0.20) = 0.072^*$

Hidden Markov models



- Find the optimal state that would generate HTH.

	H	T	H
F	0.50(0.50)=0.25 →	0.1125 →	0.050625
B	0.50(0.80)=0.40 →	0.072 →	0.05184

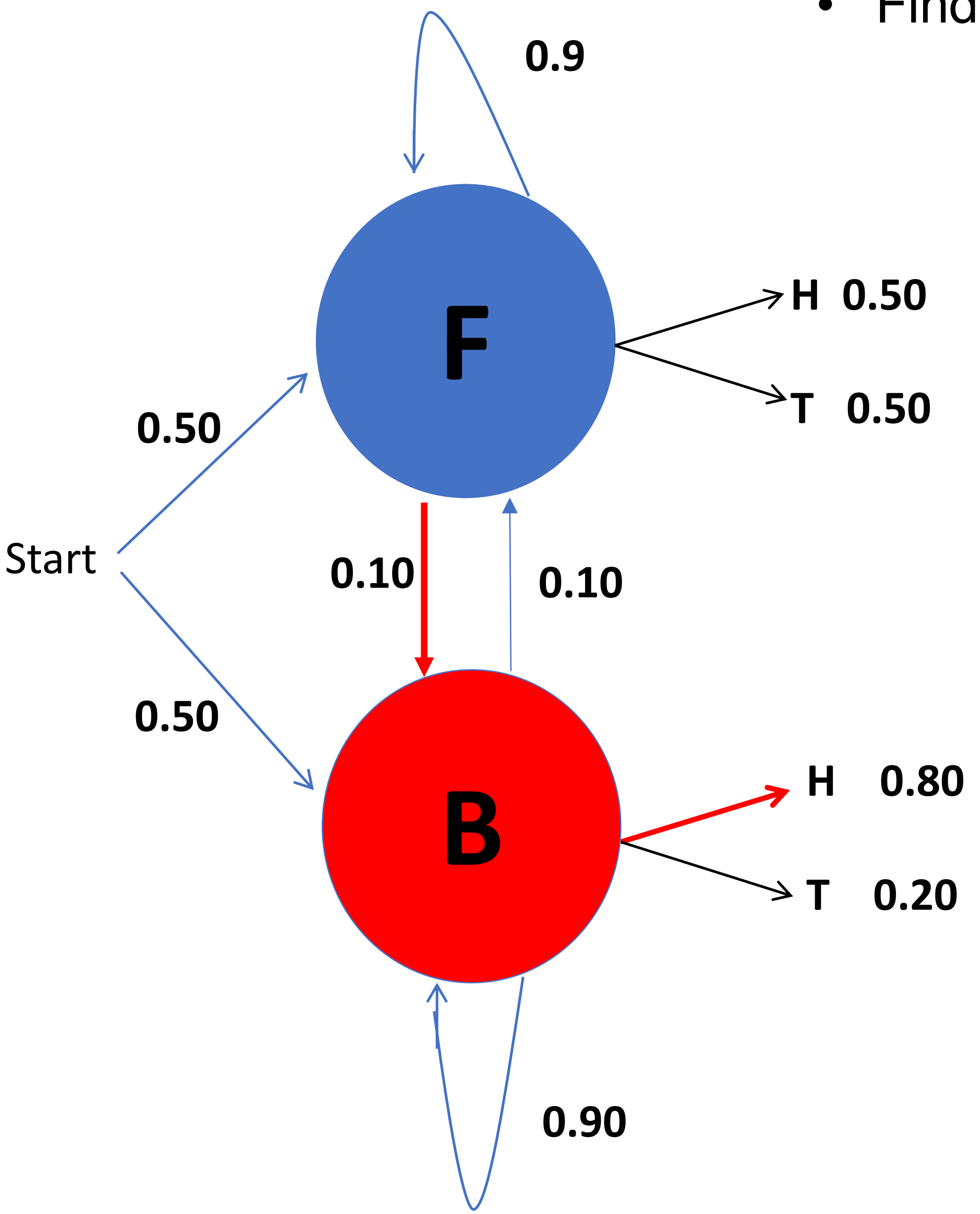
$F \rightarrow F$: $0.1125 \times 0.90(0.50) = 0.050625$*

$B \rightarrow F$: $0.072 \times 0.10(0.50) = 0.0324$

$F \rightarrow B$: $0.1125 \times 0.10(0.80) = 0.009$

$B \rightarrow B$: $0.072 \times 0.90(0.80) = 0.05184$*

Hidden Markov models



- Find the optimal state that would generate HTH.

	H	T	H
F	0.50(0.50)=0.25 →	0.1125 →	0.050625
B	0.50(0.80)=0.40 →	0.072 →	0.05184

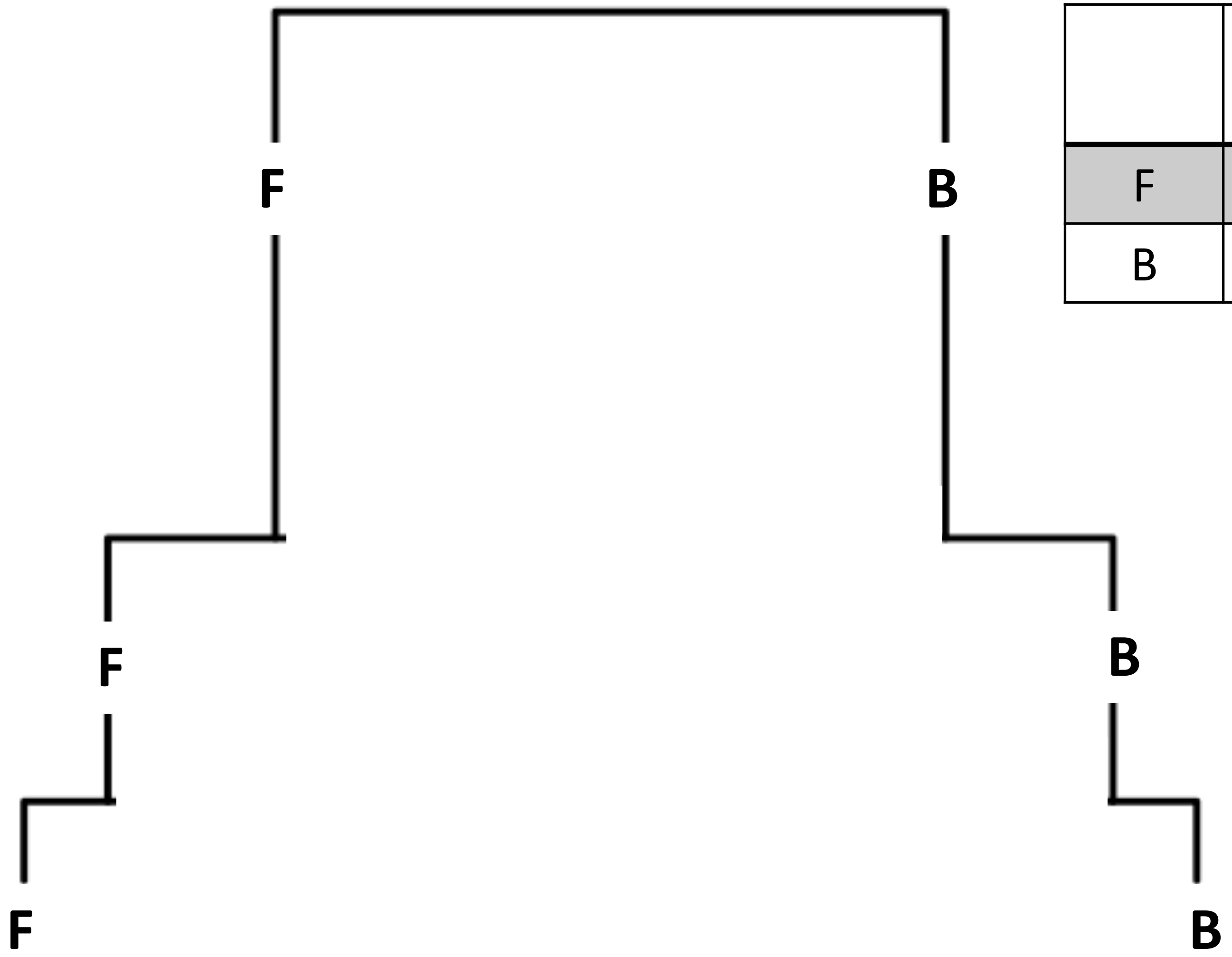
The optimal final state ends in B, since 0.05184 > 0.050625.

We then use *traceback* to find the optimal path, in this case yielding B → B → B.

The optimal state sequence is BBB

Hidden Markov models

Viterbi algorithm eliminates non-optimal paths

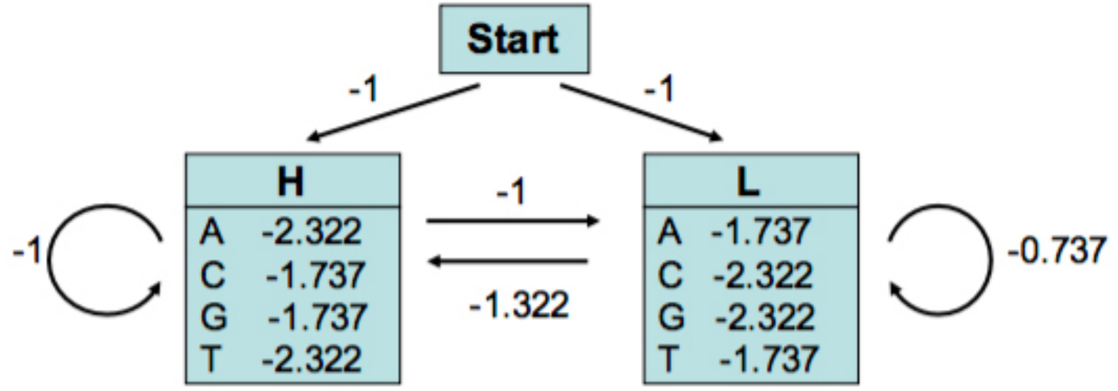


	H	T	H
F	0.25 →	0.1125 →	0.050625
B	0.40 →	0.072 →	0.05184

Suppose 100 coin tosses, then

- Brute force: 2^{100} possibilities (31 digits. Not feasible …
- Viterbi: (roughly) $100 \times 4 = 400$ calculations (no problem)

HMM : Viterbi algorithm - a toy example



GGCACTGAA

back-tracking

(= finding the path which corresponds to the highest probability, -24.49)

	G	G	C	A	C	T	G	A	A
H	**-2.73**	**-5.47**	**-8.21**	-11.53	-14.01	...			-25.65
L	-3.32	-6.06	-8.79	**-10.94**	**-14.01**	...			**-24.49**

The most probable path is: **HHHLLLLLL**

Its probability is $2^{-24.49} = 4.25E\text{-}8$ (remember that we used $\log_2(p)$)

"Simple" model: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-5-59

Augustus Model: https://academic.oup.com/bioinformatics/article/19/suppl_2/ii215/180603

Note: probabilities are on the log2 scale.

Source (no longer available), http://homepages.ulb.ac.be/%7Edgonze/TEACHING/viterbi.pdf