Introducción al Procesamiento de Lenguaje Natural (NLP)

Clase II: Sistemas de Reconocimiento Automático de Voz (ASR) III. HMM

Alexander Caicedo

Universidad del Rosario | Escuela de Ingeniería, Ciencia y Tecnología | MACC Matemáticas Aplicadas y Ciencias de la Computación

Contenido

- HMM.
- HMM para ASR systems.

Propiedad de Markov

Restricción de los sistemas dinámicos donde el estado actual depende solamente del estado anterior, y no del historial de estados.

Hidden Markov Models I

HMM Clásico con observaciones deterministas

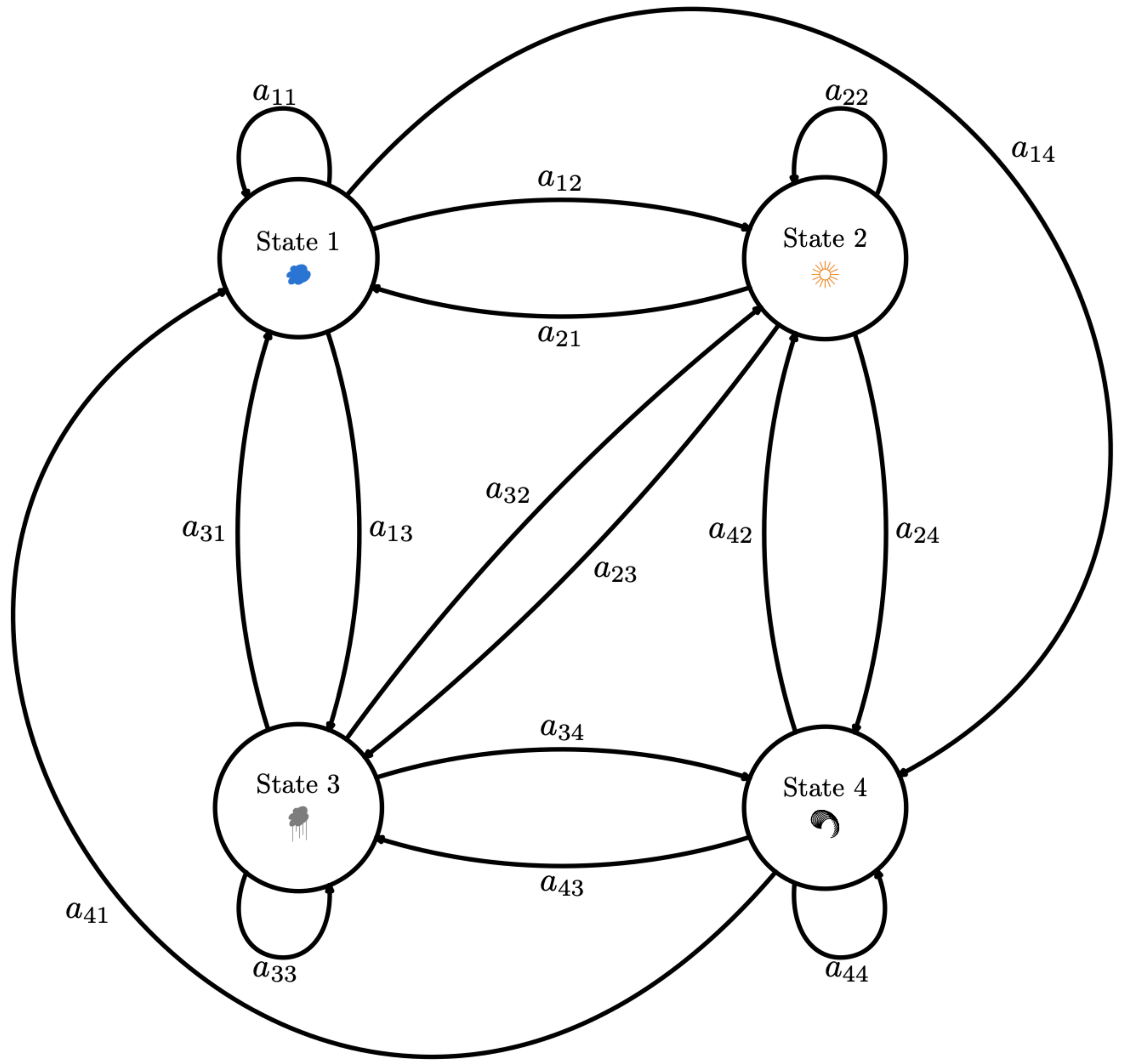


Figure 4.1: Markov model of the weather

- **State 1:** cloudy
- **State 2:** sunny
- **State 3:** rainy
- **State 4:** windy

Matrix de Transición de Probabilidades

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{bmatrix}_{N\times N}$$

Probabilidades de Transición

$$a_{ij} \geq 0 \qquad \forall j, i$$
$$\sum_{j=1}^{N} a_{ij} = 1 \quad \forall i$$

Distribución Inicial de los estados

$$\boldsymbol{\pi} = \begin{bmatrix} \pi_1 = P(q_1 = 1) \\ \pi_2 = P(q_1 = 2) \\ \vdots \\ \pi_N = P(q_1 = N) \end{bmatrix}_{N\times 1} \qquad \sum_{i=1}^{N} \pi_i = 1$$

Source: https://www.diva-portal.org/smash/get/diva2:831263/FULLTEXT01.pdf

Hidden Markov Models II

HMM Clásico con observaciones deterministas

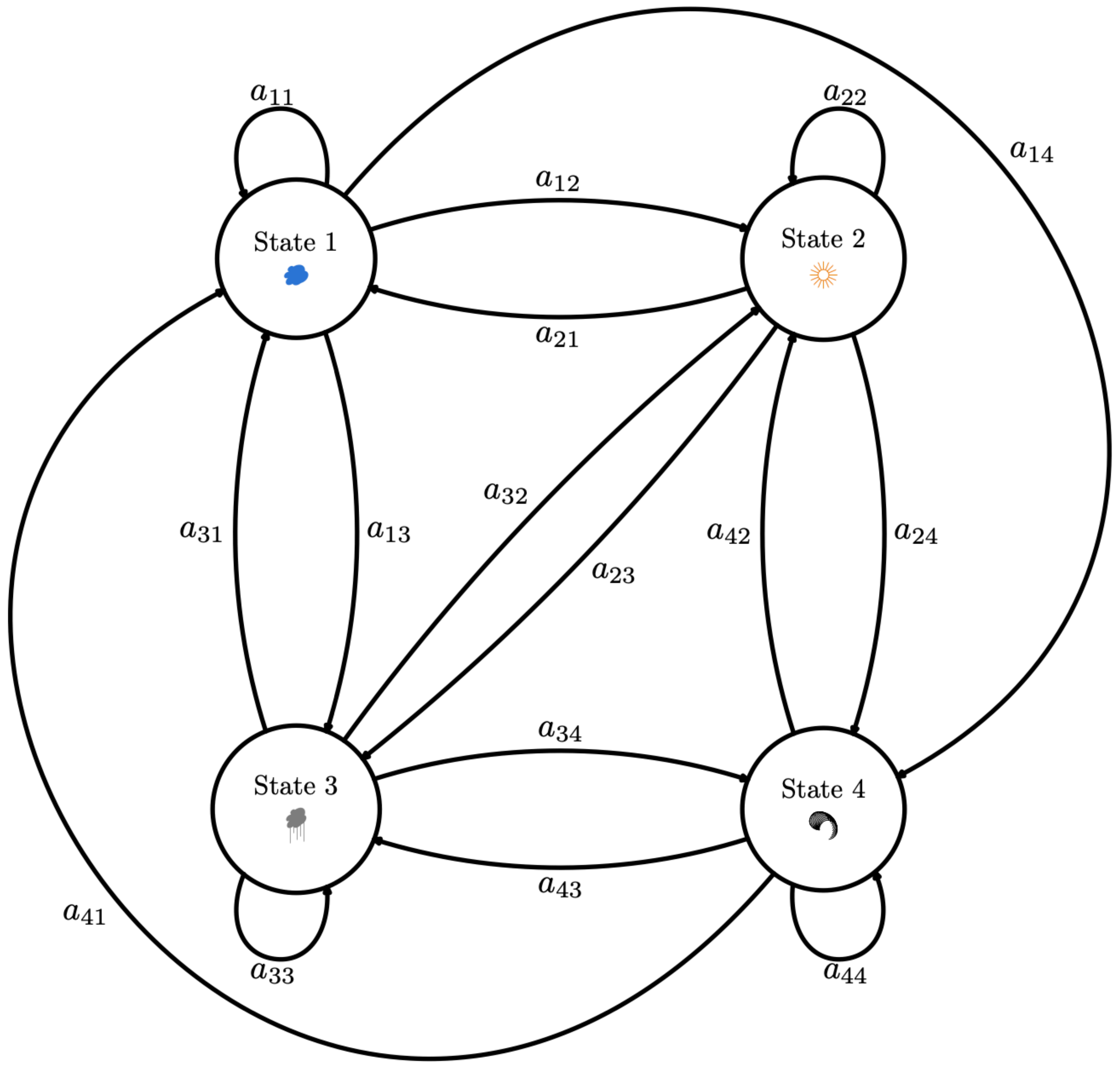


Figure 4.1: Markov model of the weather

En el modelo HMM propuesto, ¿cuál es la probabilidad de que la secuencia de eventos **O** ocurra?

$$\mathbf{O} = (\text{sunny}, \text{rainy}, \text{sunny}, \text{windy}, \text{cloudy}, \text{cloudy}) = (2,\ 3,\ 2,\ 4,\ 1,\ 1)$$

- **State 1:** cloudy
- **State 2:** sunny
- **State 3:** rainy
- **State 4:** windy

$$\begin{aligned} P(\mathbf{O}|\mathbf{A}, \boldsymbol{\pi}) &= P(2,3,2,4,1,1|\mathbf{A}, \boldsymbol{\pi}) \\ &= P(2)P(3|2)P(2|3)P(4|2)P(1|4)P(1|1) \\ &= \pi_2 \cdot a_{23} \cdot a_{32} \cdot a_{24} \cdot a_{41} \cdot a_{11} \end{aligned}$$

Source: https://www.diva-portal.org/smash/get/diva2:831263/FULLTEXT01.pdf

Hidden Markov Models III

¿Qué pasa si los estados ahora no son deterministicos sino probabilisticos?

Caracterización de la distribución de probabilidad de las observaciones

$$b_j(\mathbf{o}_t) = P(\mathbf{o}_t | q_t = j)$$

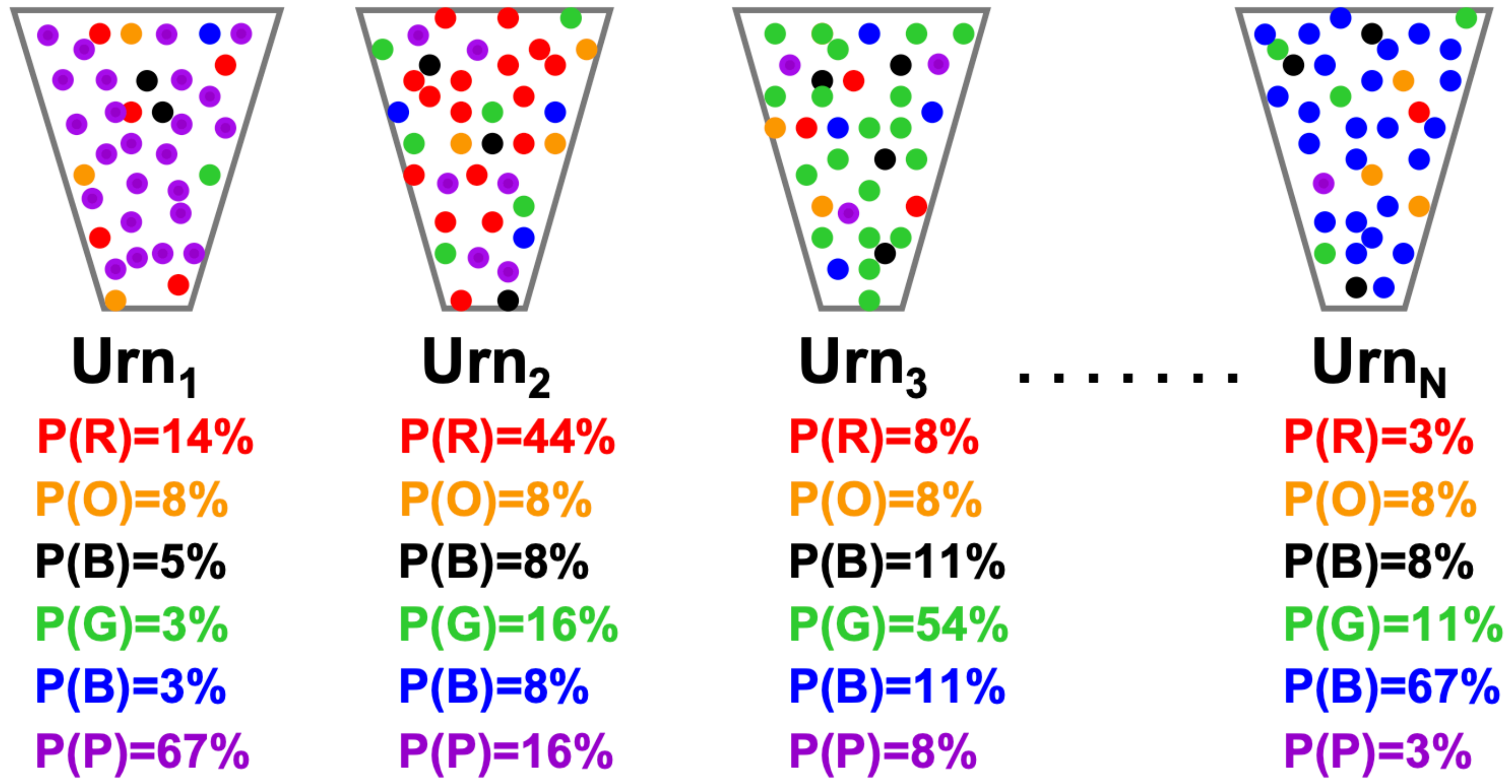


Figure 4.2: Urn and Ball example

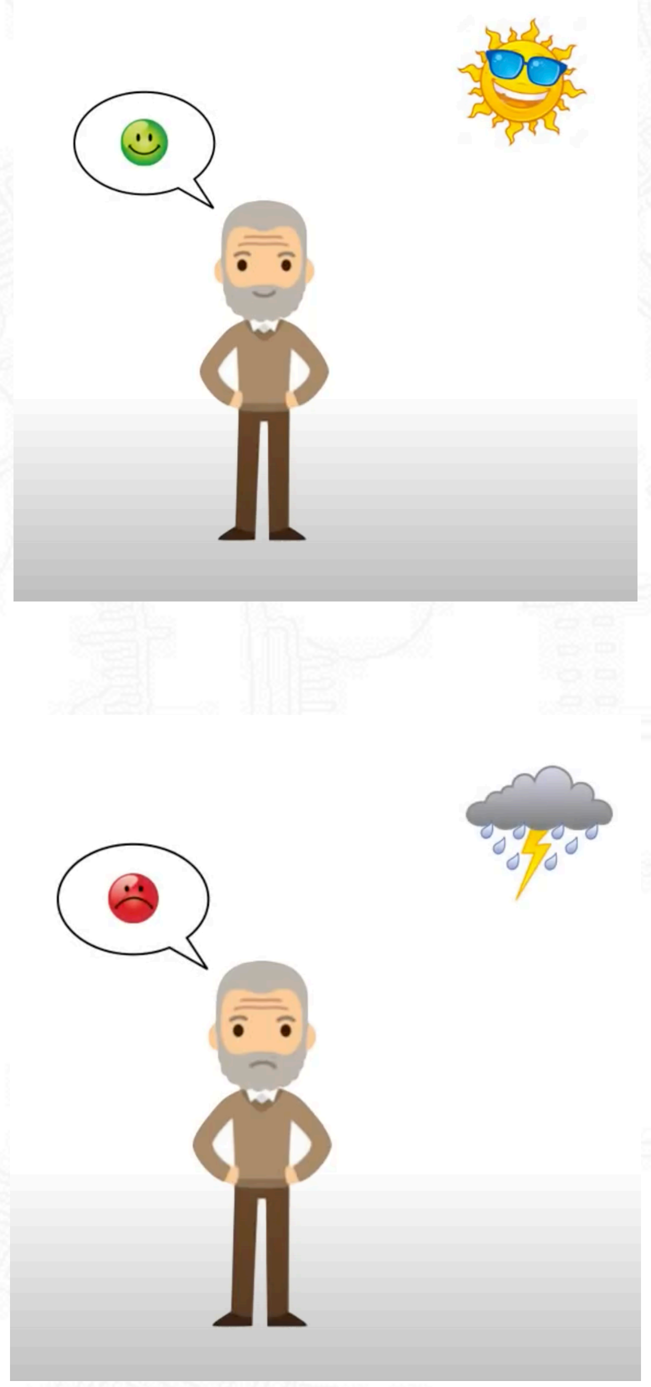
¿Cómo el HMM genera observaciones?

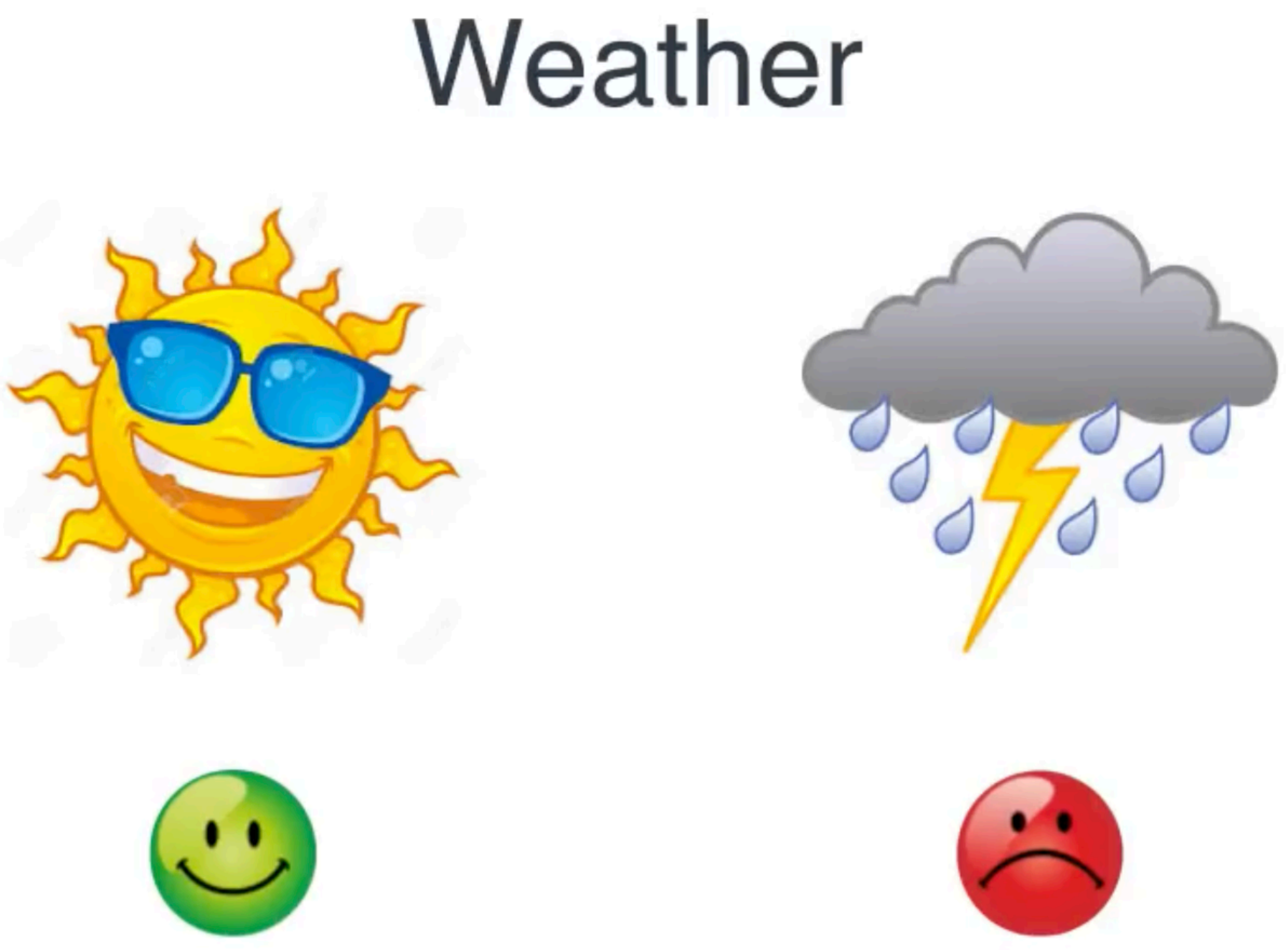
1. Escojo j un estado de acuerdo a la distribución a priori.
2. Defino $t = 1; 1,2,3,...,T$
3. A partir de este estado escojo una observación a partir de su $b_j(o_t)$
4. Me desplazo a otro estado , según la probabilidad de transición.
5. Defino $t = t + 1$
6. Repito desde el paso 3, hasta terminar.

Source: https://www.diva-portal.org/smash/get/diva2:831263/FULLTEXT01.pdf

Ejemplo

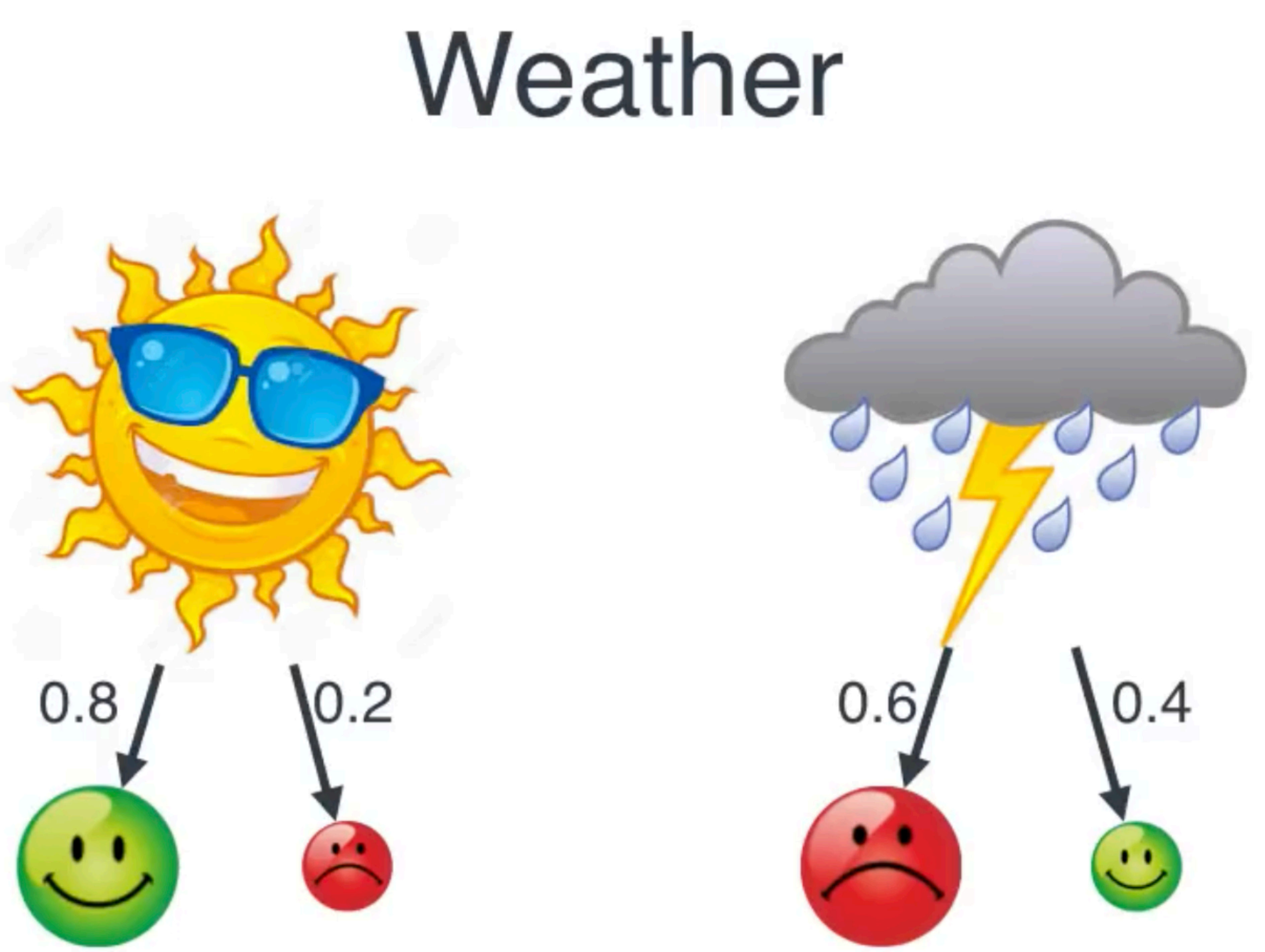
Alice quiere saber el clima, y ella sabe que el estado de animo de Bob se afecta pro el clima

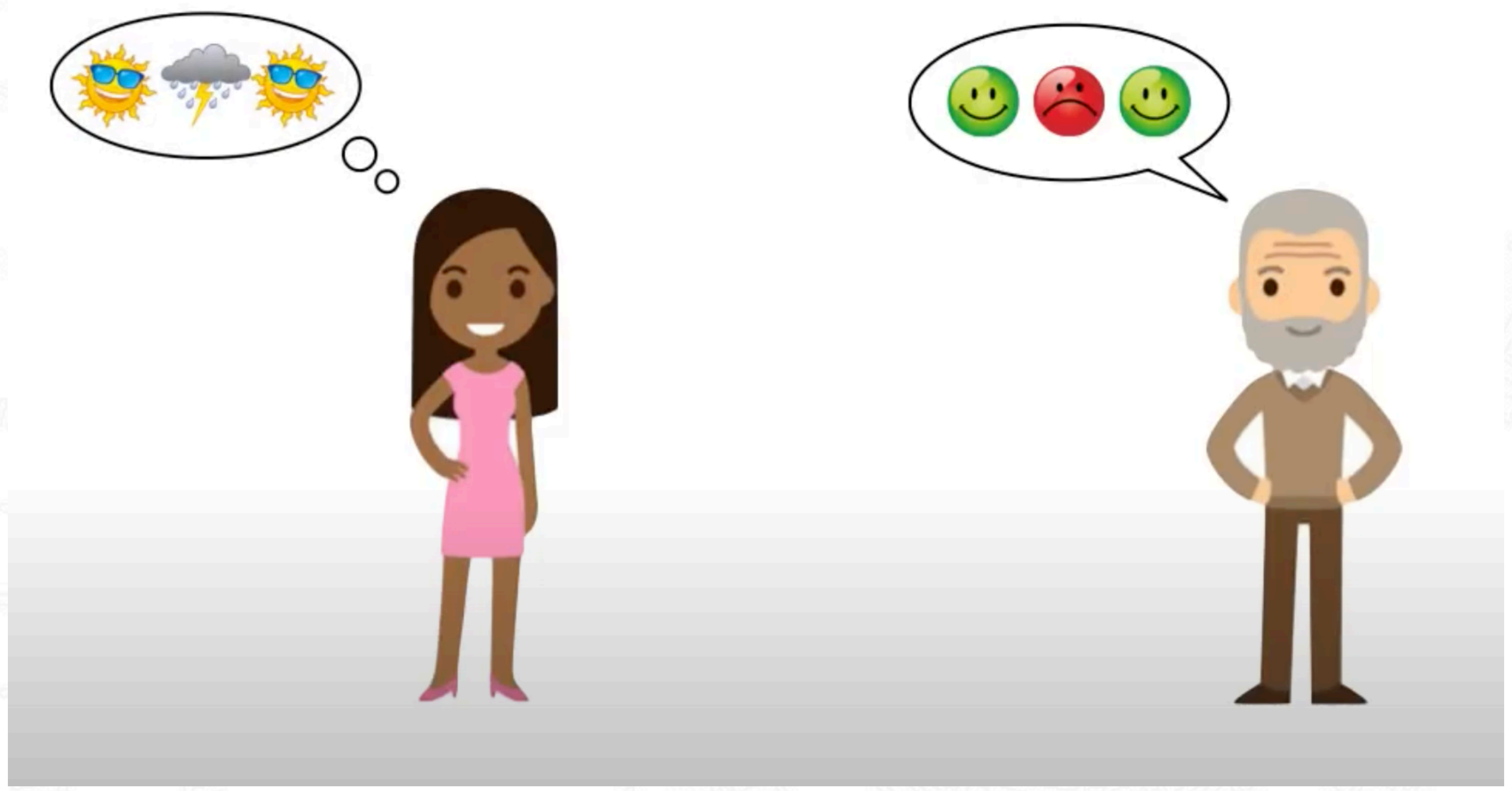




Source: https://www.youtube.com/watch?v=kqSzLo9fenk

Ejemplo

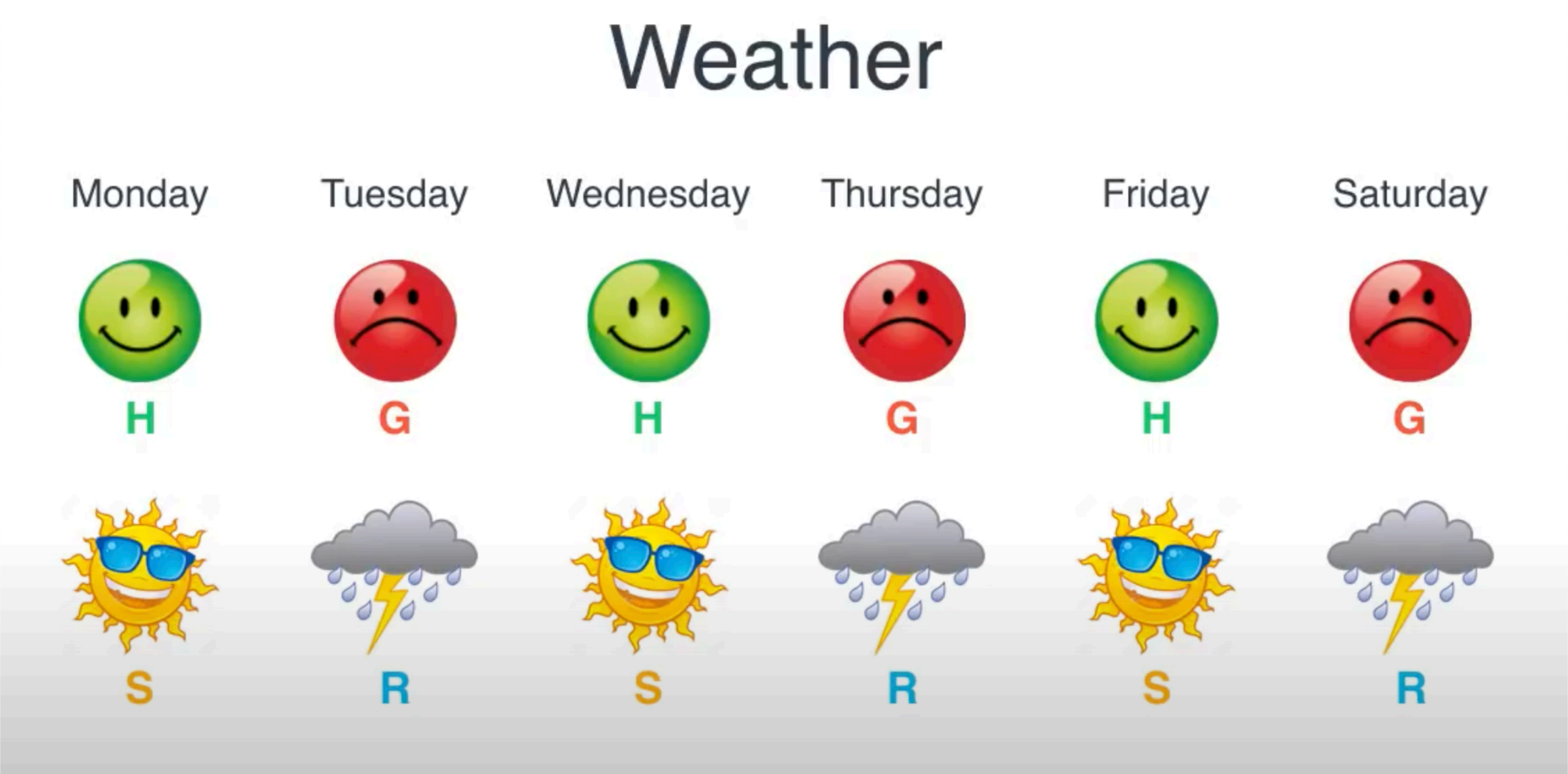




¿Qué sucede si el estado de Bob, aunque es afectado pro el clima, no es deterministico?

Source: https://www.youtube.com/watch?v=kqSzLo9fenk

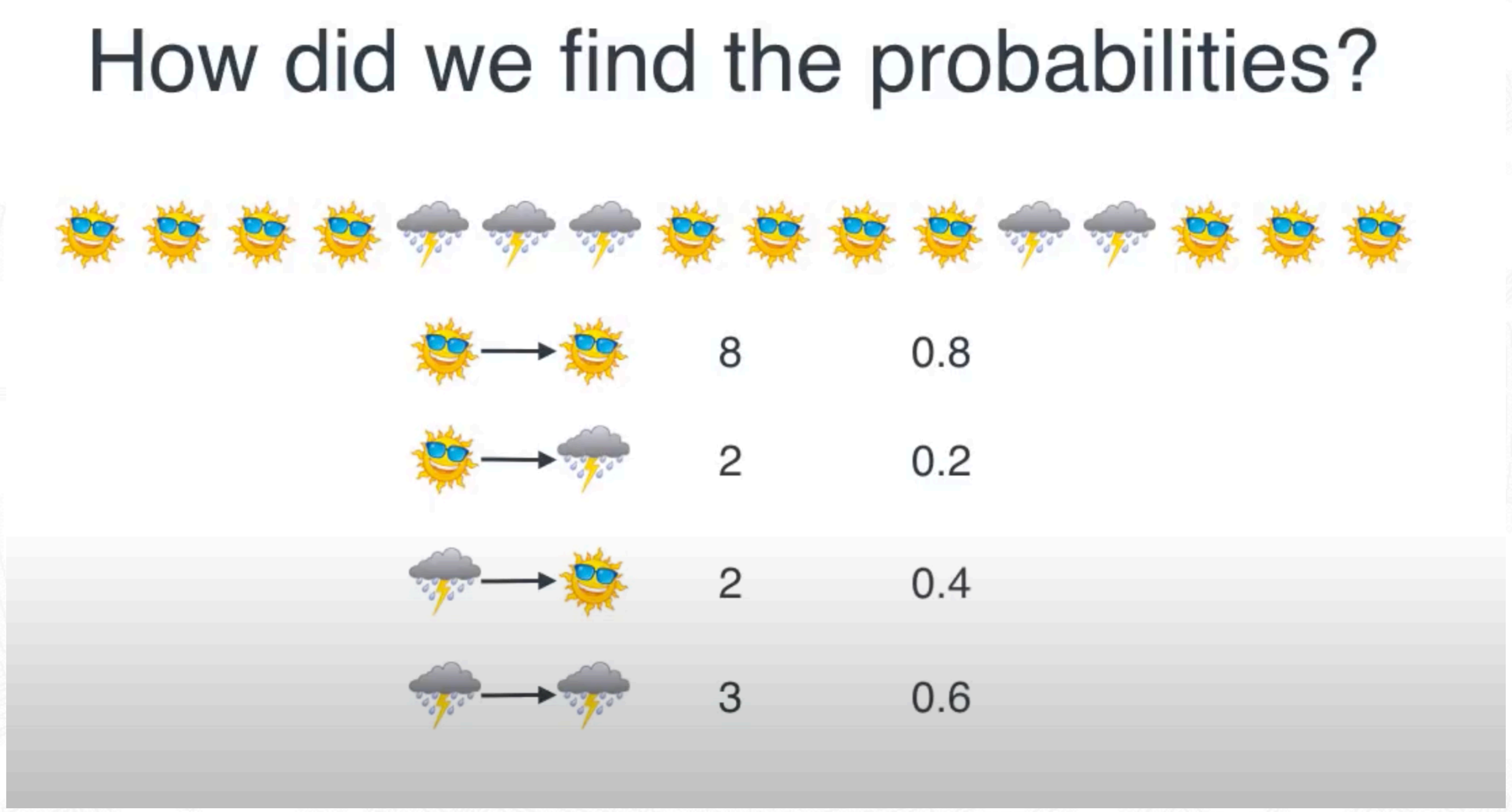
Ejemplo

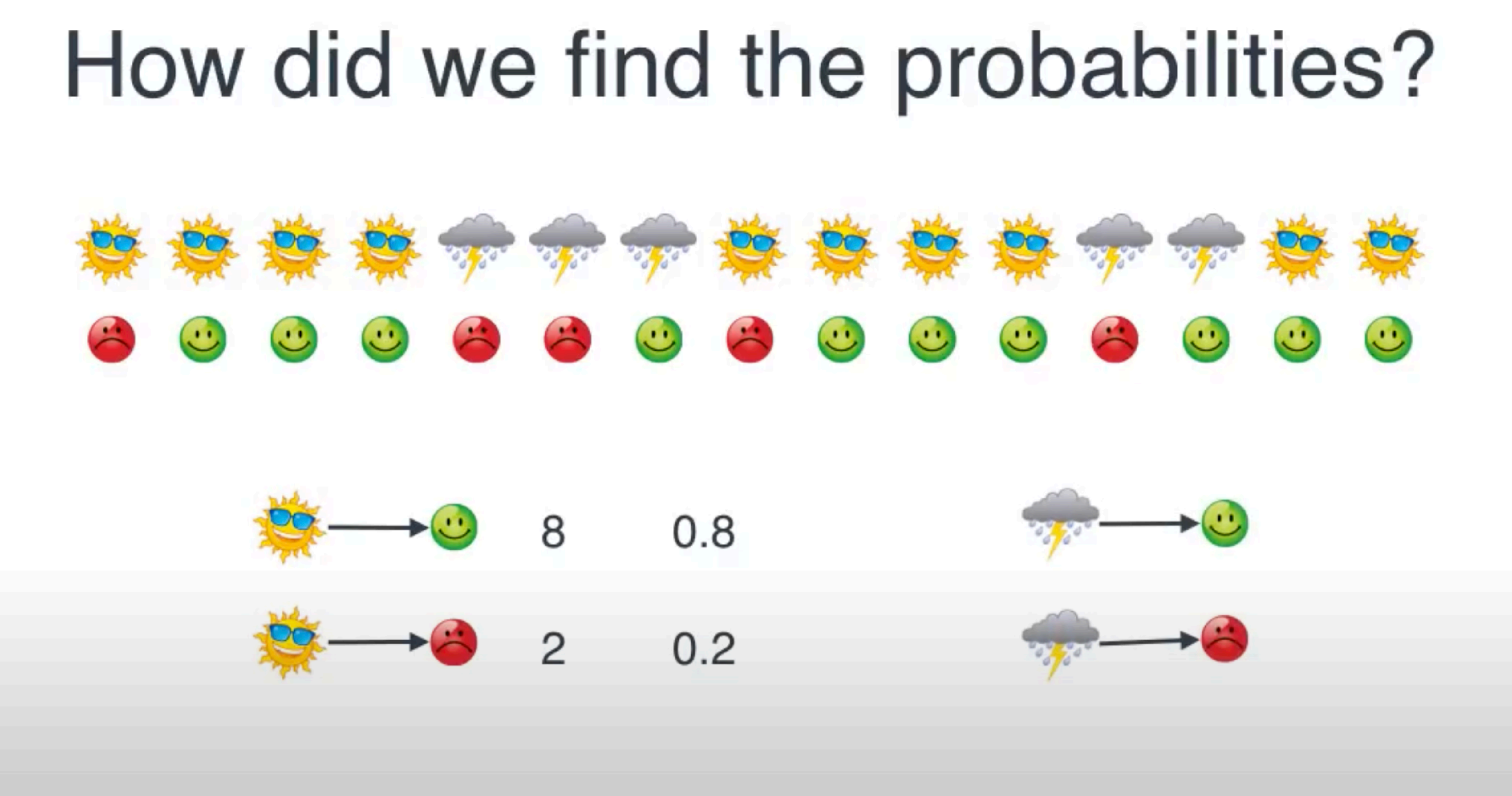


Ahora supongamos que Bob tuvo una semana un poco extraña, según su estado de animo.
¿Es probable que el clima sea tan variable?

Source: https://www.youtube.com/watch?v=kqSzLo9fenk

Ejemplo

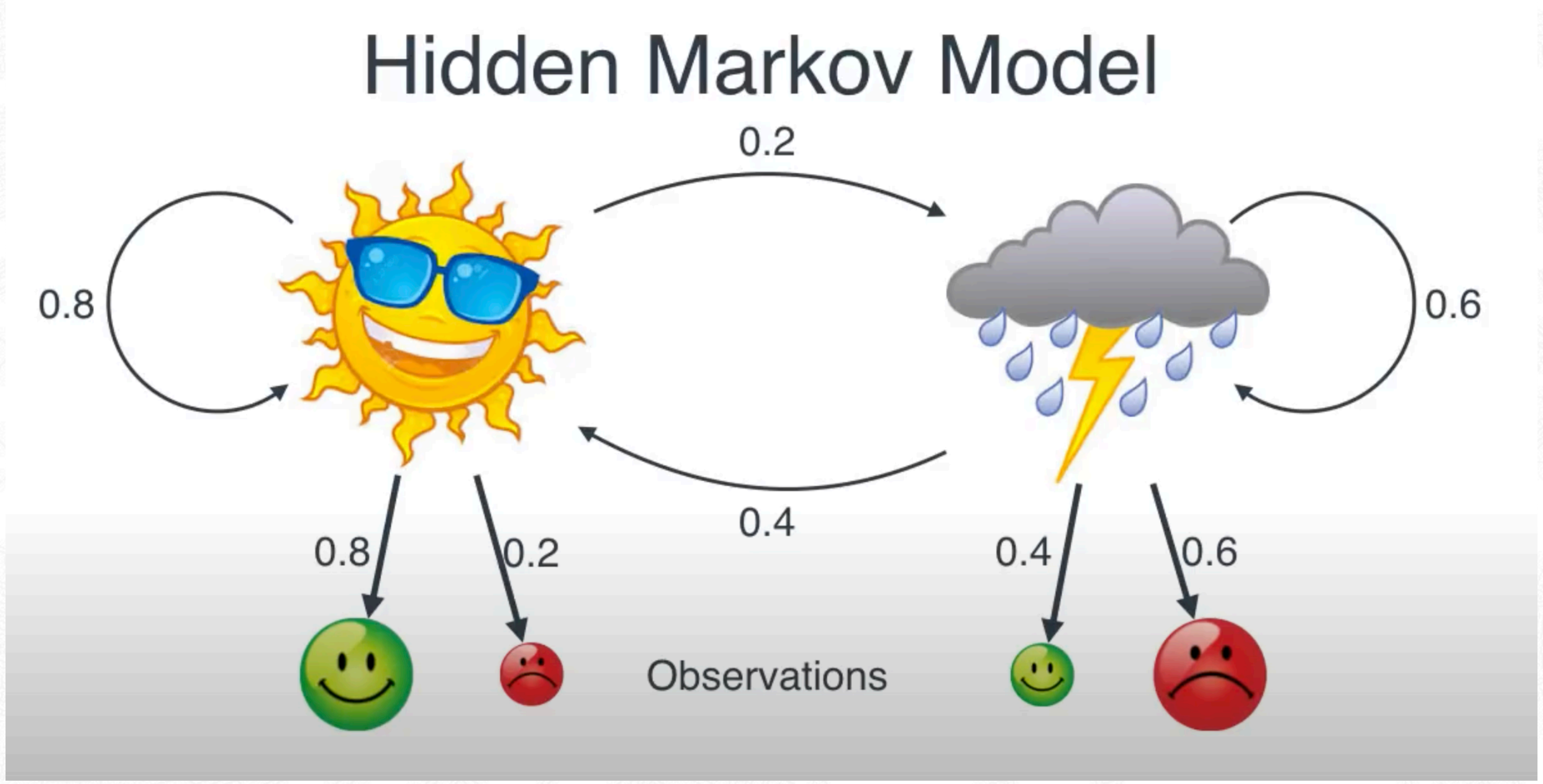




Supongamos que se tienen las siguientes observaciones

Source: https://www.youtube.com/watch?v=kqSzLo9fenk

Ejemplo

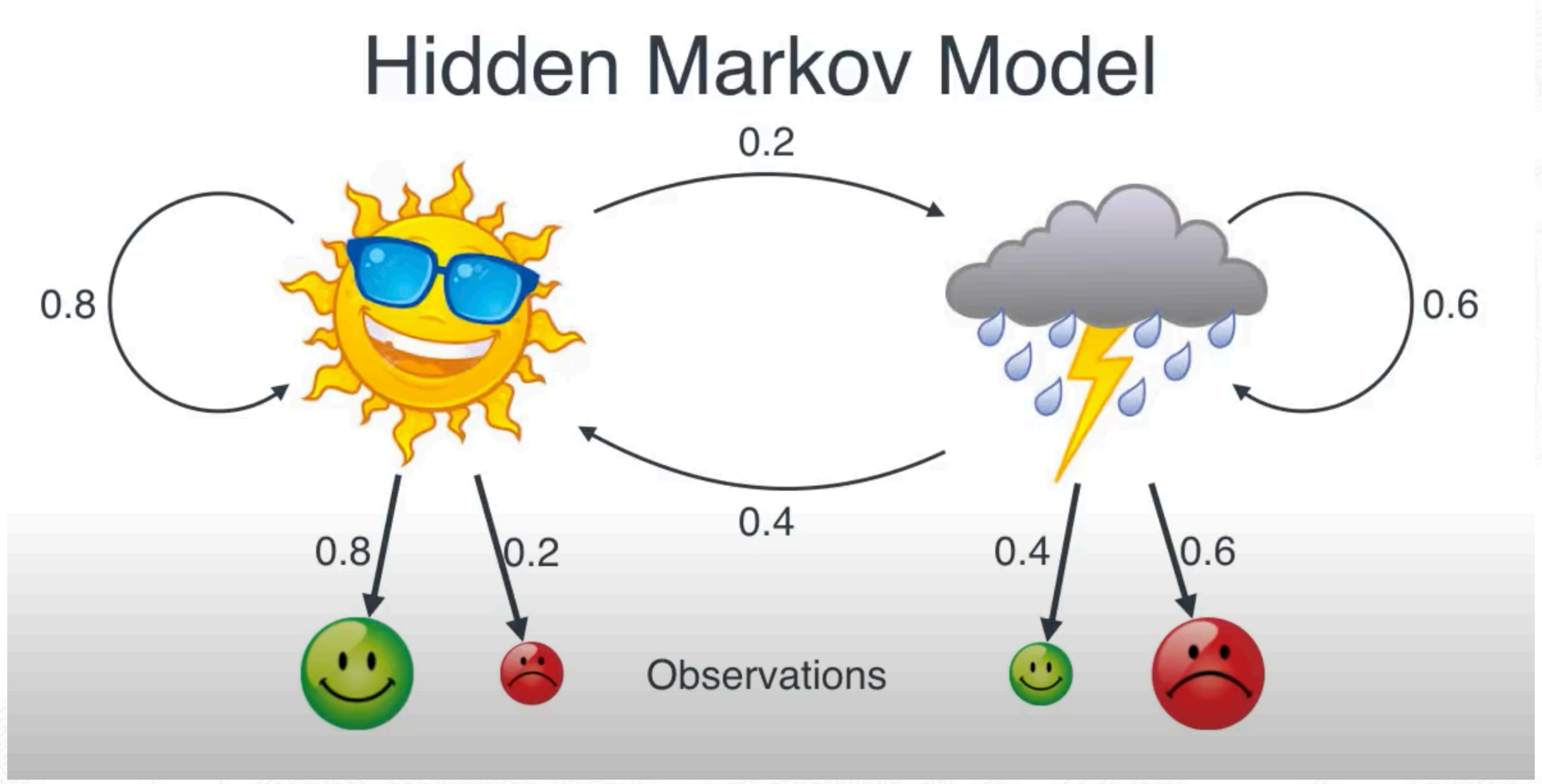


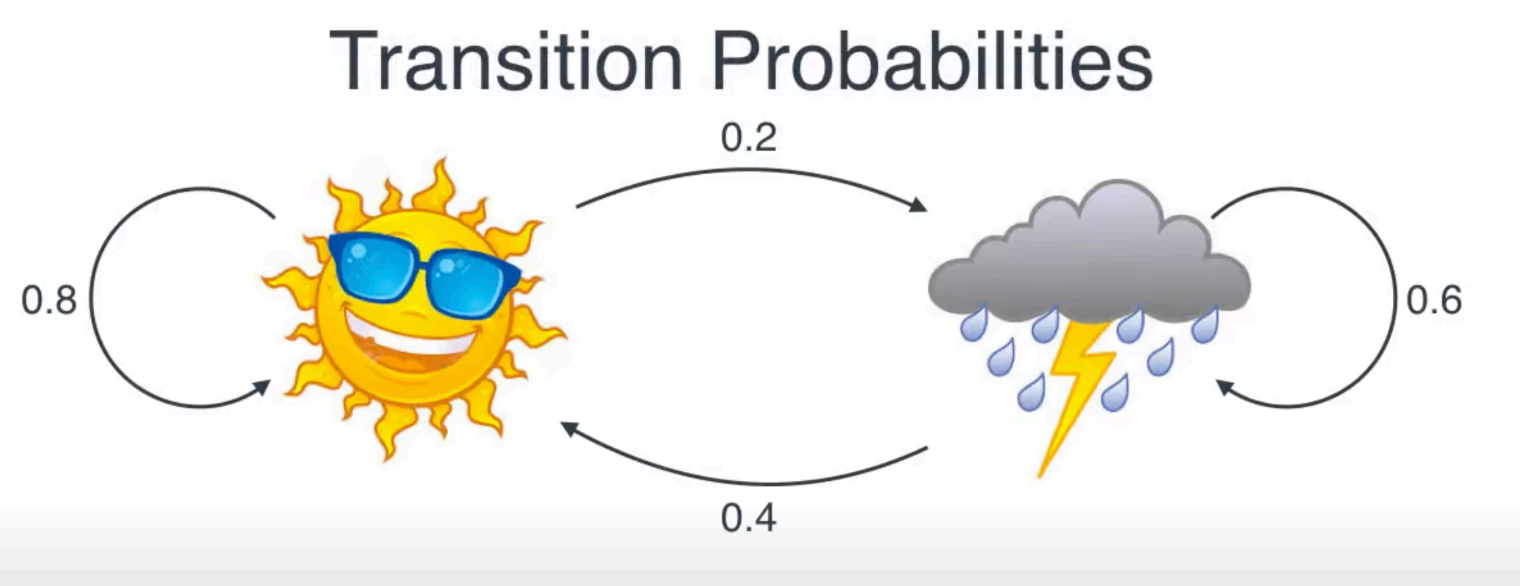
Con las probabilidades calculadas podemos postular el siguiente modelo oculto de Markov.

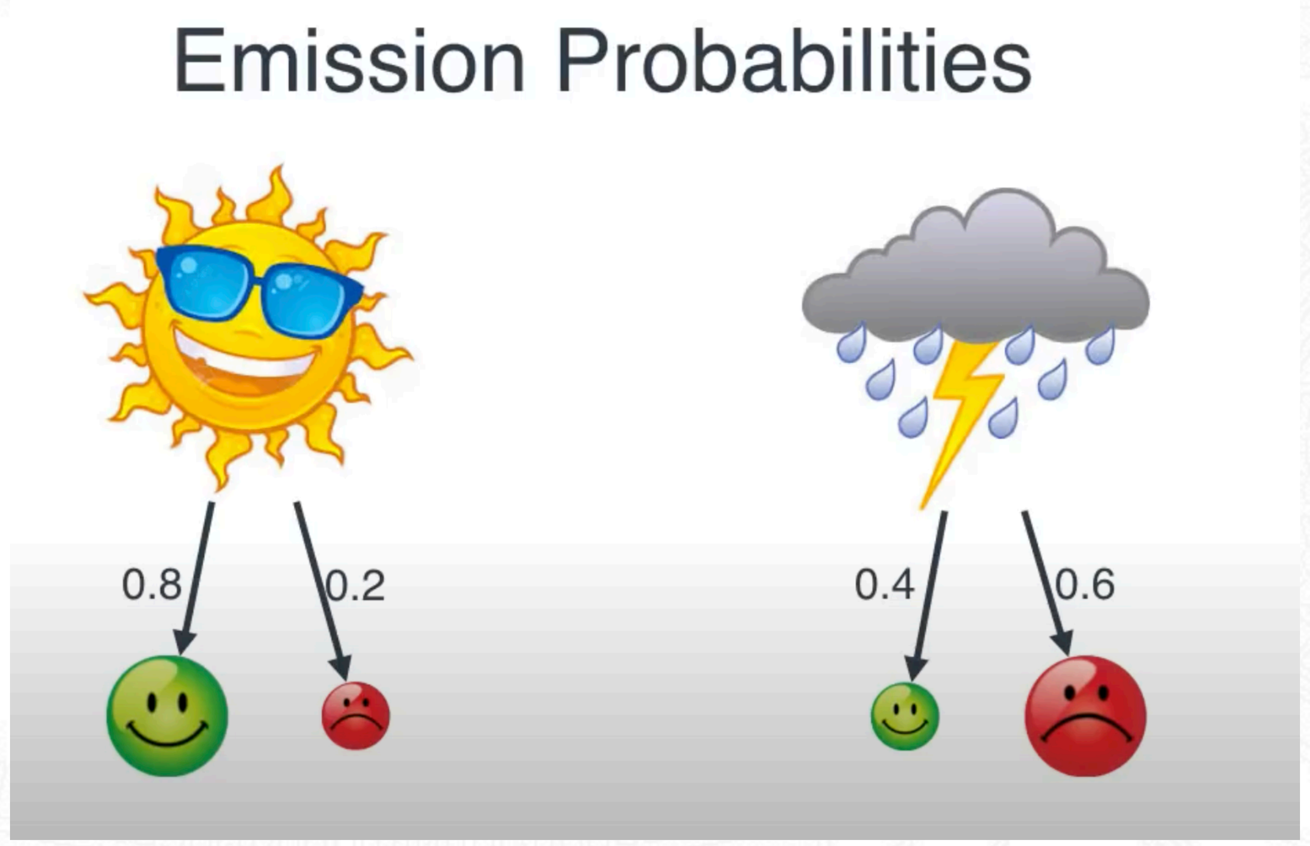
Source: https://www.youtube.com/watch?v=kqSzLo9fenk

Ejemplo

El modelo esta compuesto por las probabilidades de Transición y las de Emisión.



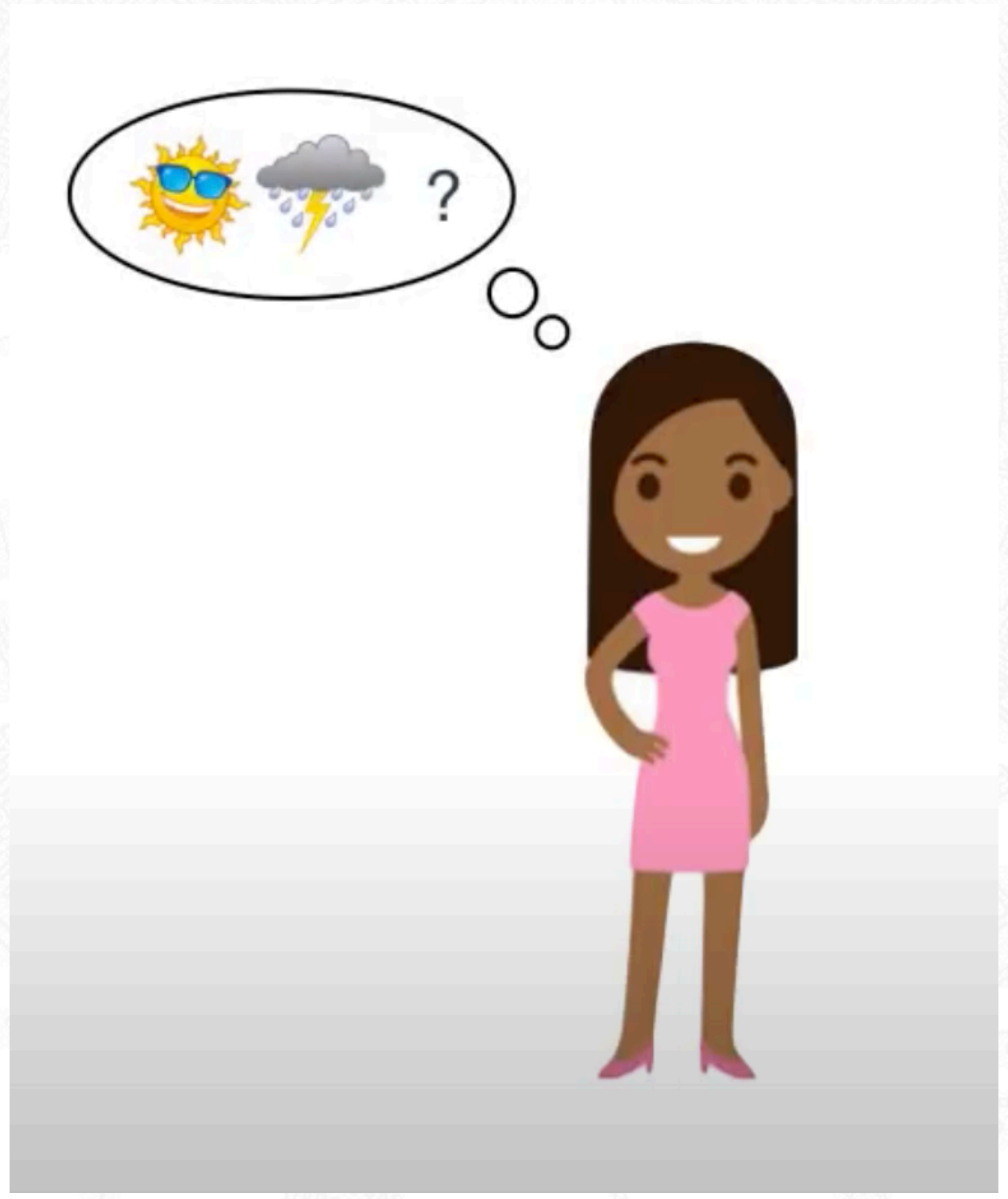


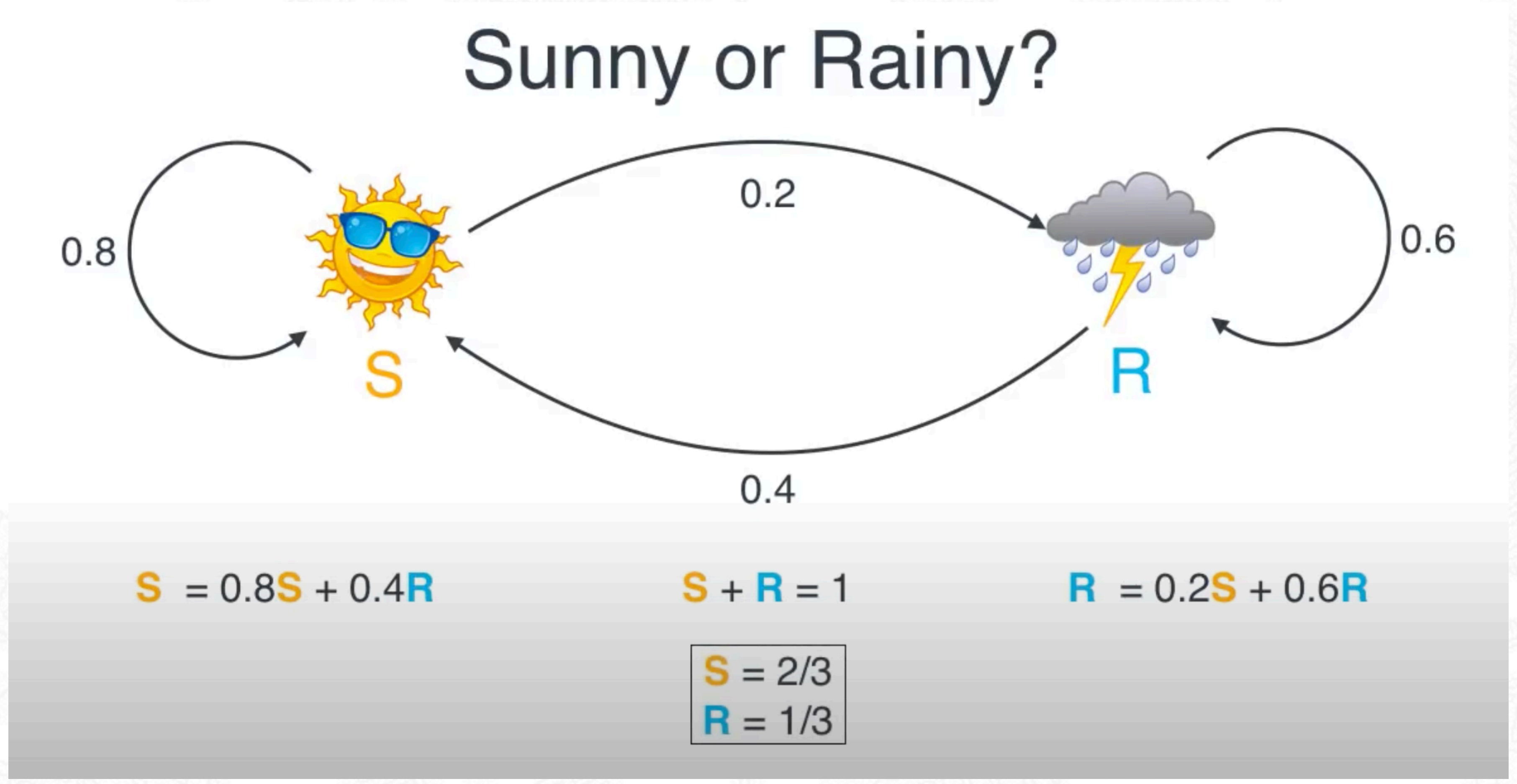


Source: https://www.youtube.com/watch?v=kqSzLo9fenk

Ejemplo

Alice se pregunta, basado en el modelo, ¿cuál es la probabilidad de que sea un día soleado o un día lluvioso?

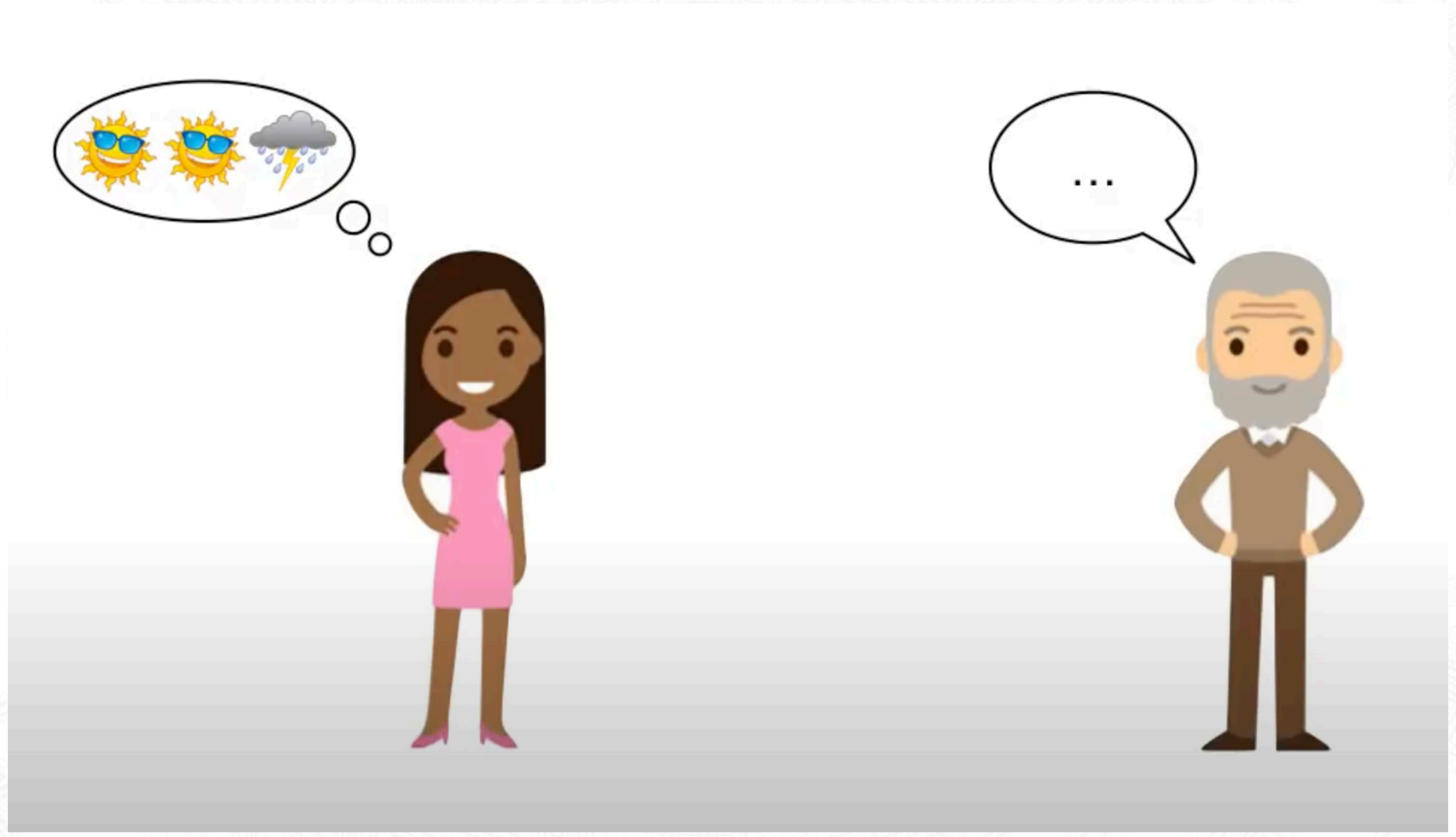


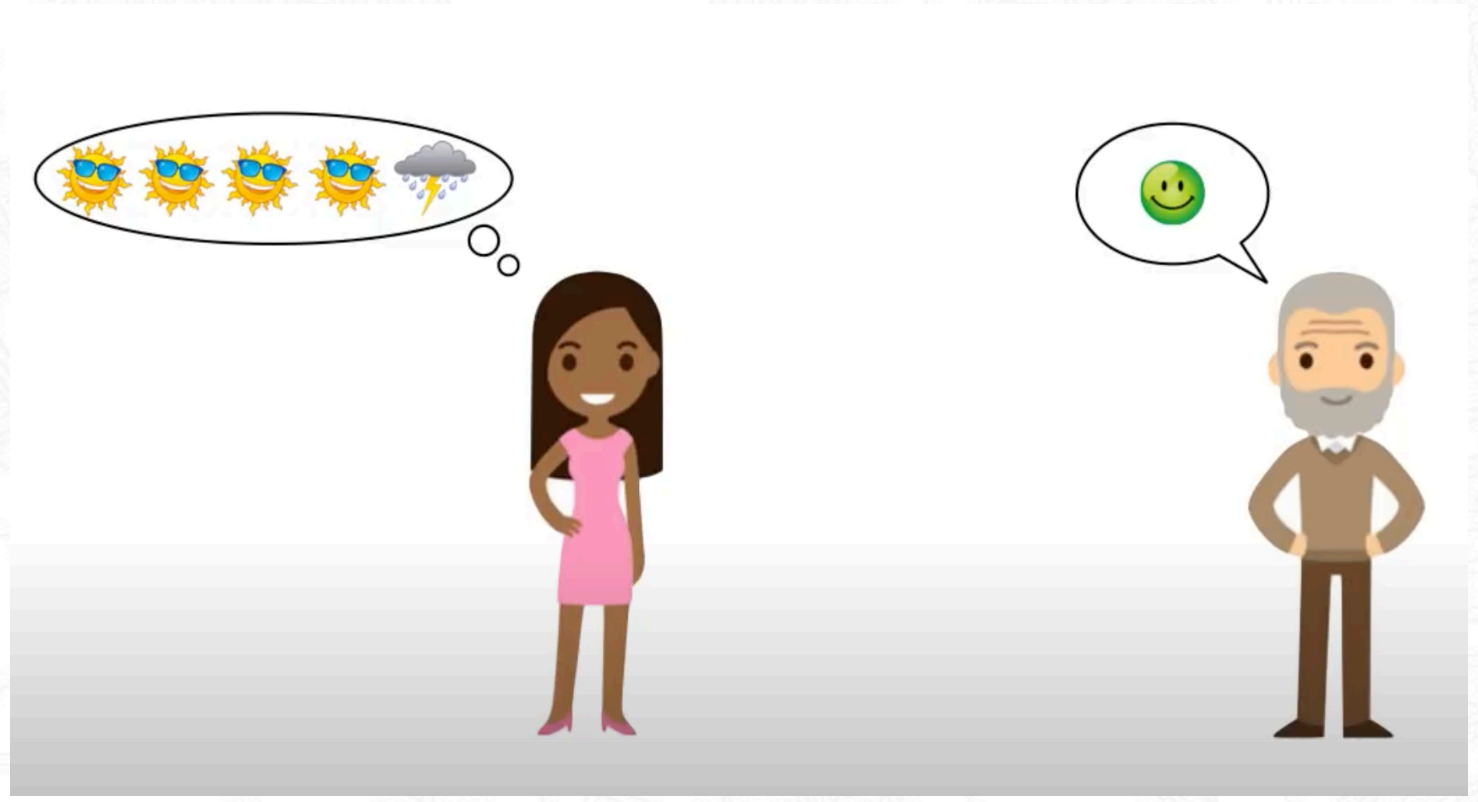


Source: https://www.youtube.com/watch?v=kqSzLo9fenk

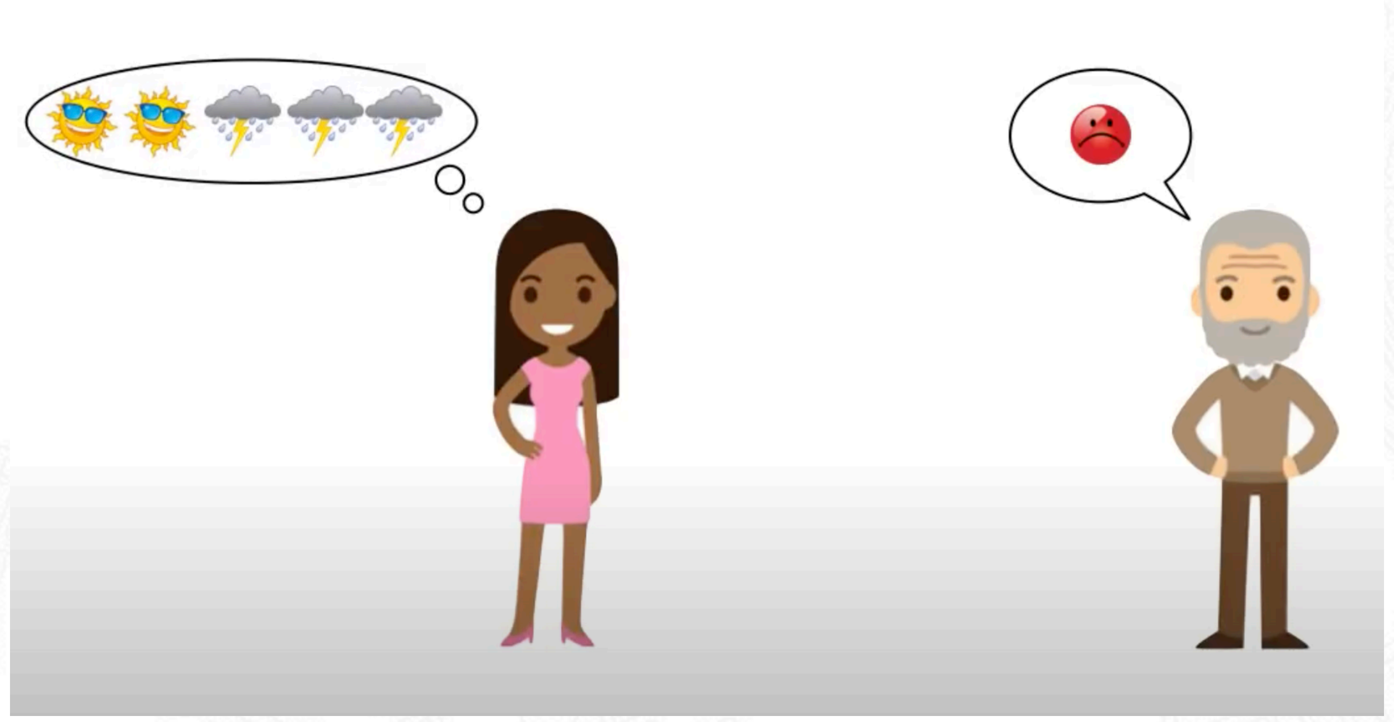
Ejemplo

El prior de un estado es la probabilidad si no tenemos información extra. En este caso el estado de animo de Bob.



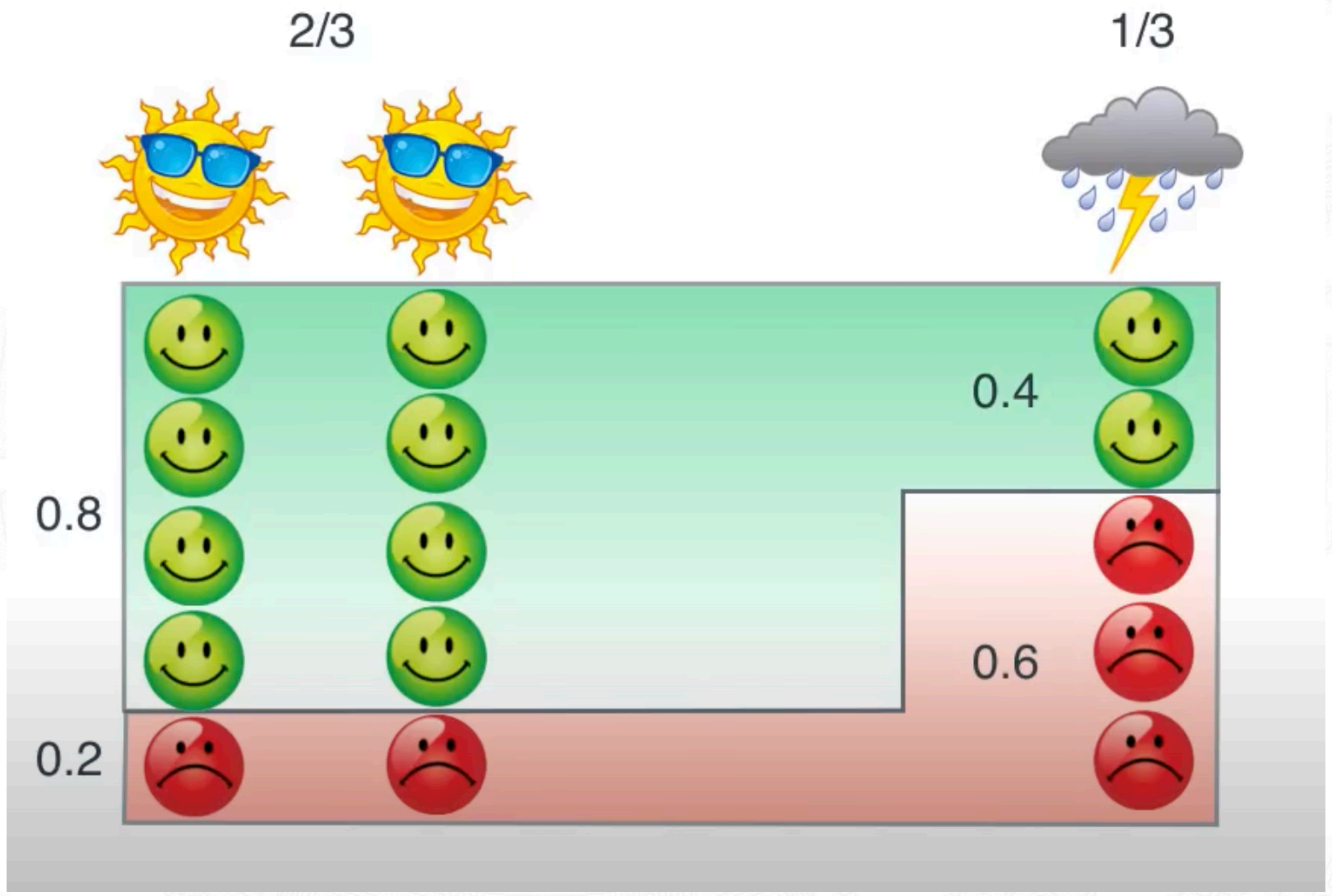


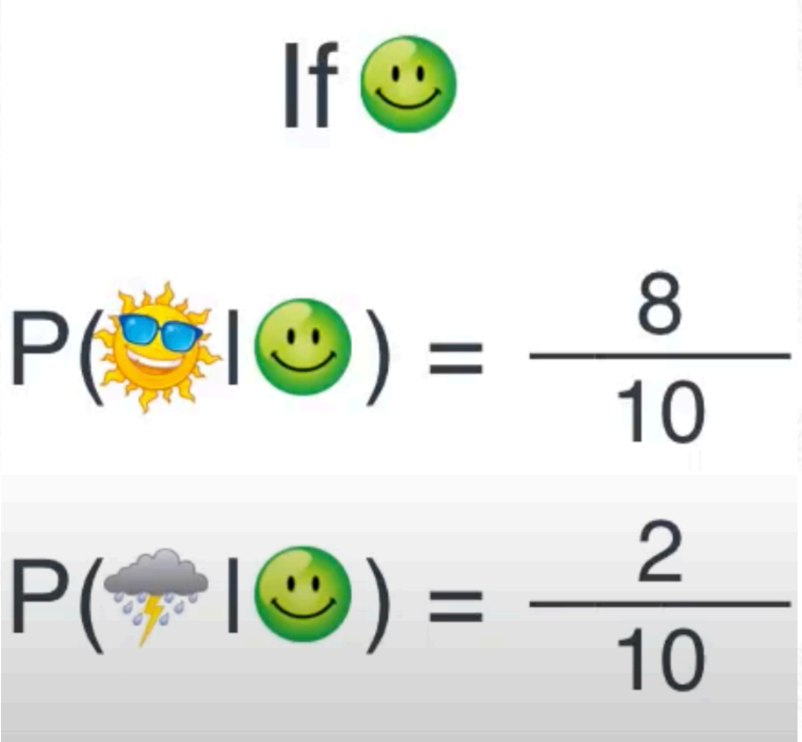
Al tener información sobre el estado de animo de Bob, las probabilidades cambian.

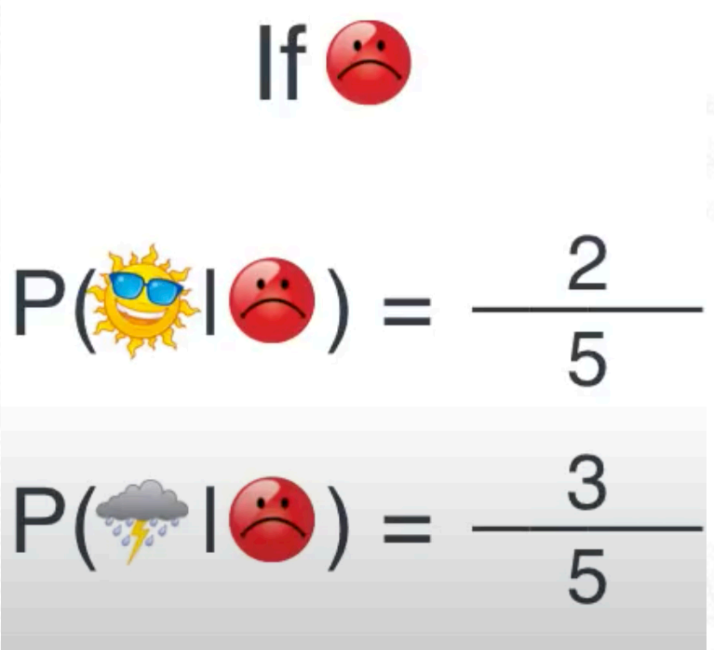


Source: https://www.youtube.com/watch?v=kqSzLo9fenk

Ejemplo





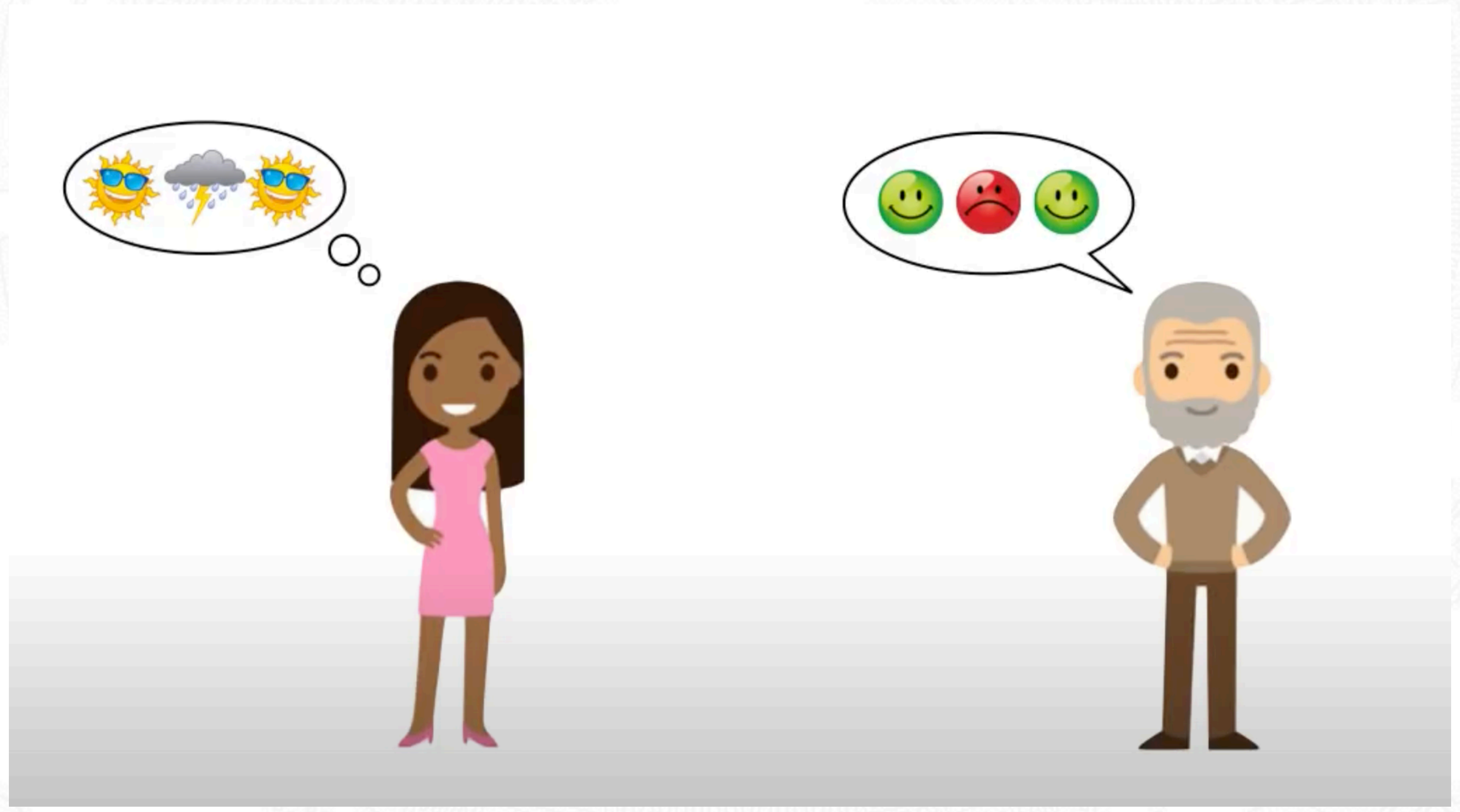


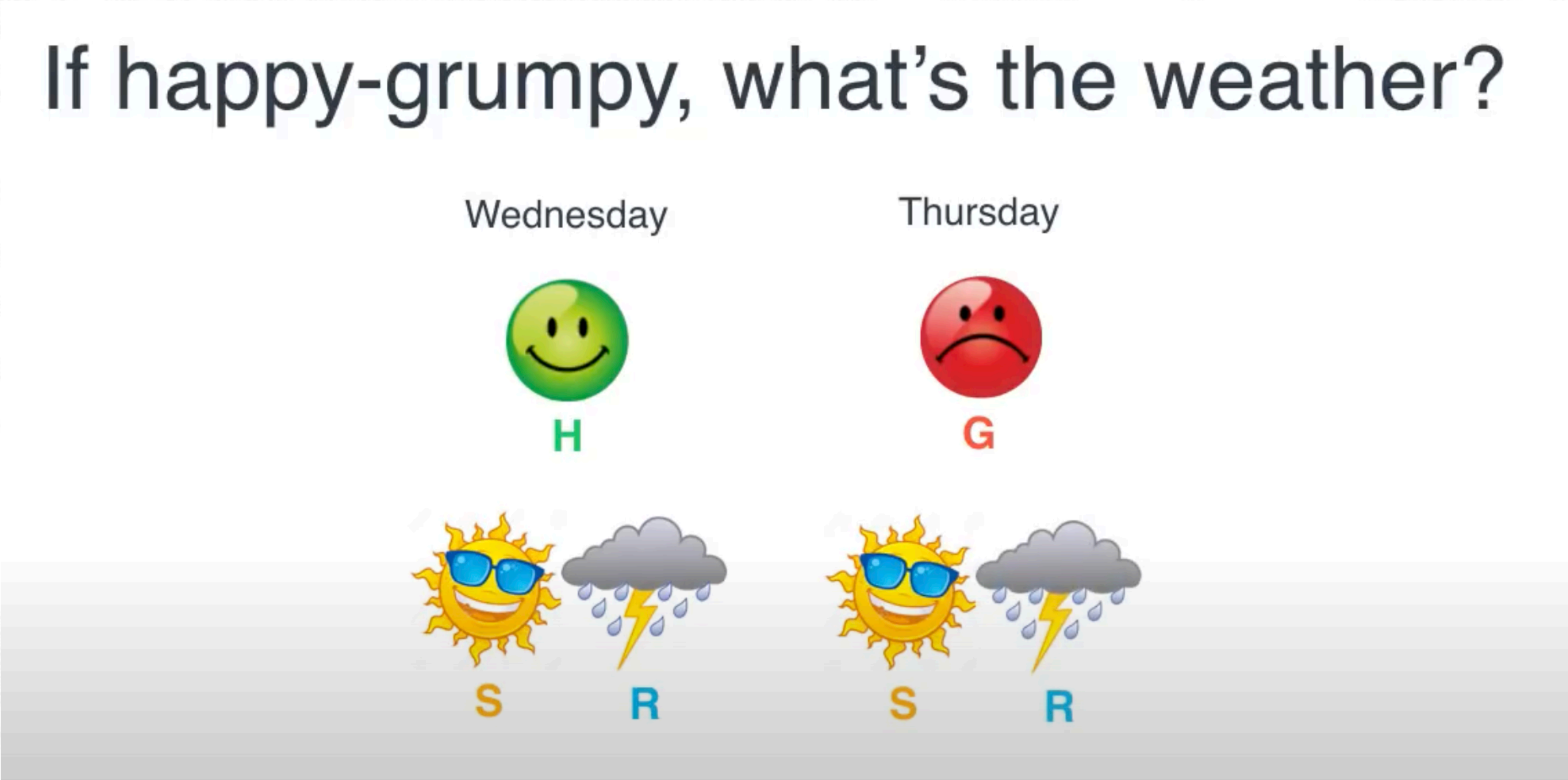
Estas son las nuevas probabilidades condicionadas al estado de animo de Bob (Bayes Theorem)

Source: https://www.youtube.com/watch?v=kqSzLo9fenk

Ejemplo

¿Cuál será la secuencia de estados del clima basados en las observaciones del estado del animo de Bob?

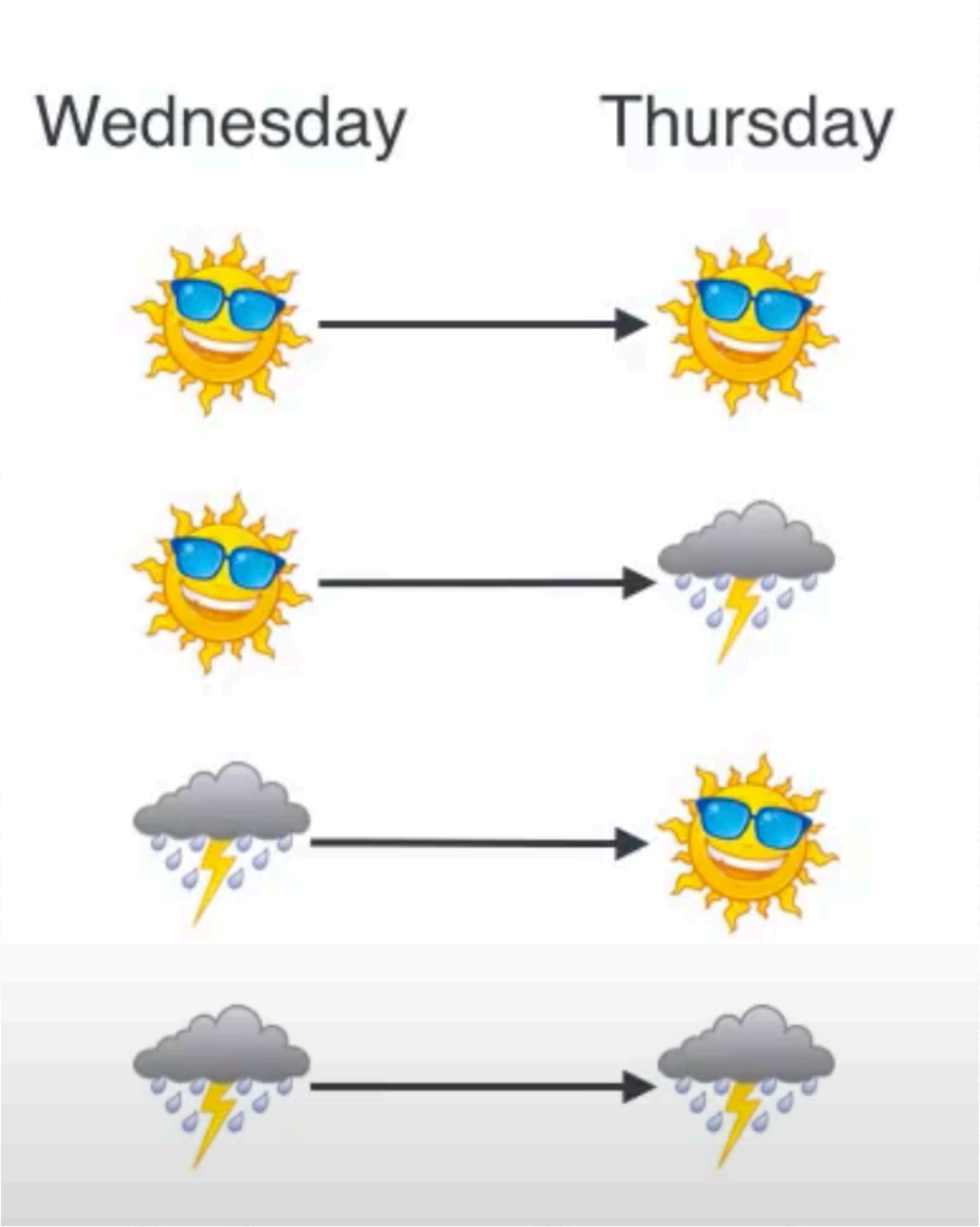


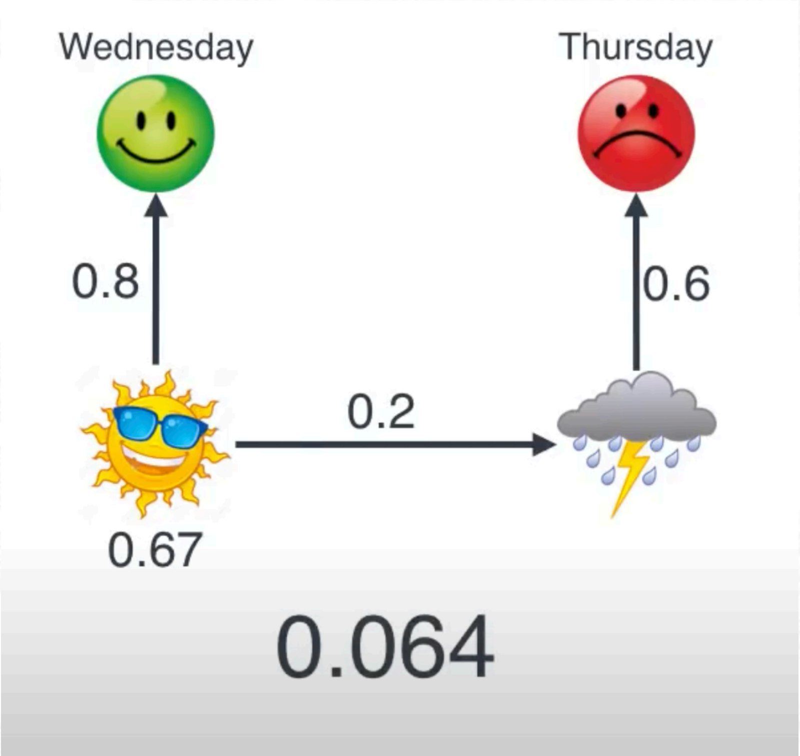


Source: https://www.youtube.com/watch?v=kqSzLo9fenk

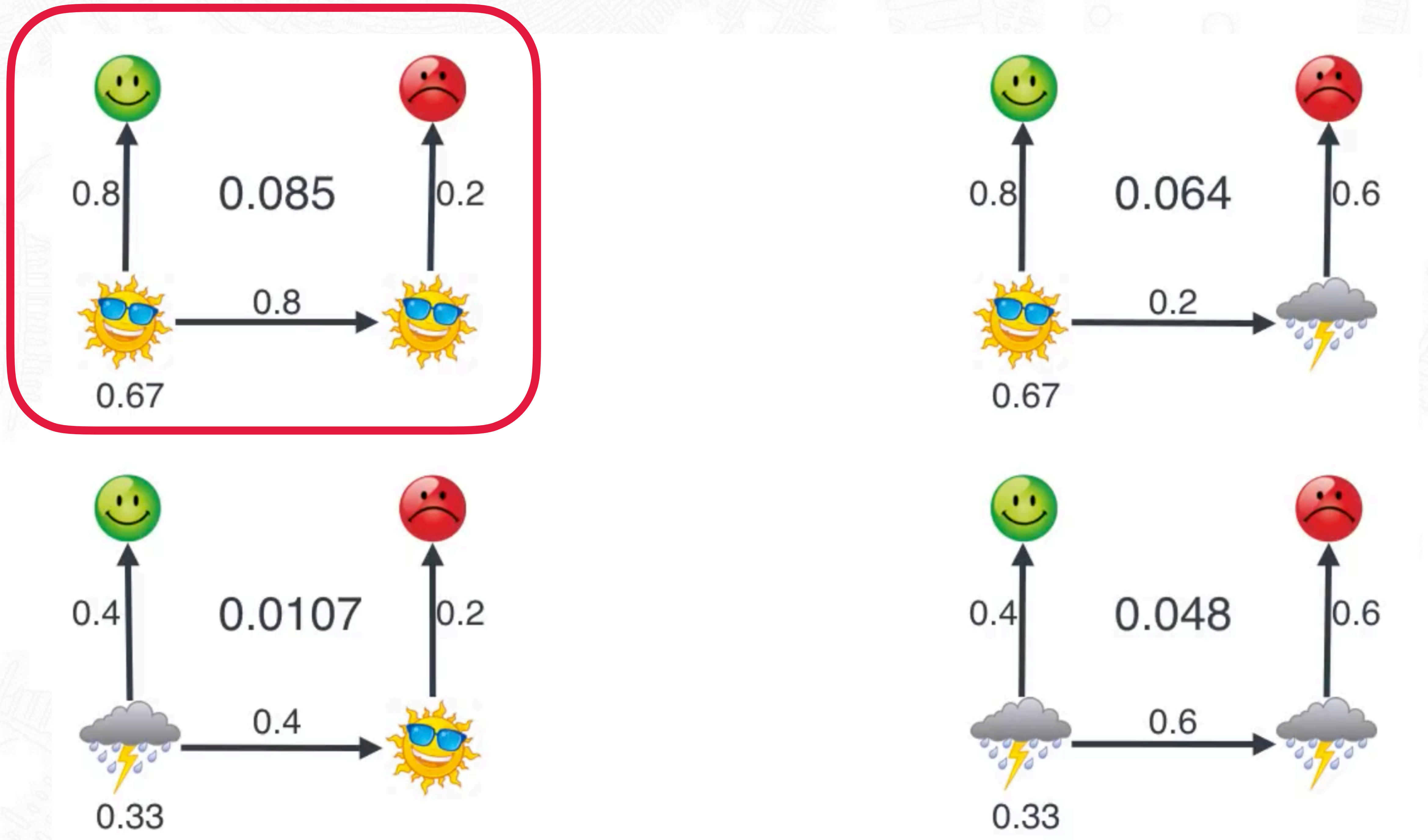
Ejemplo

Si solo queremos saber la secuencia del clima en dos días consecutivos, necesitamos evaluar 4 condiciones





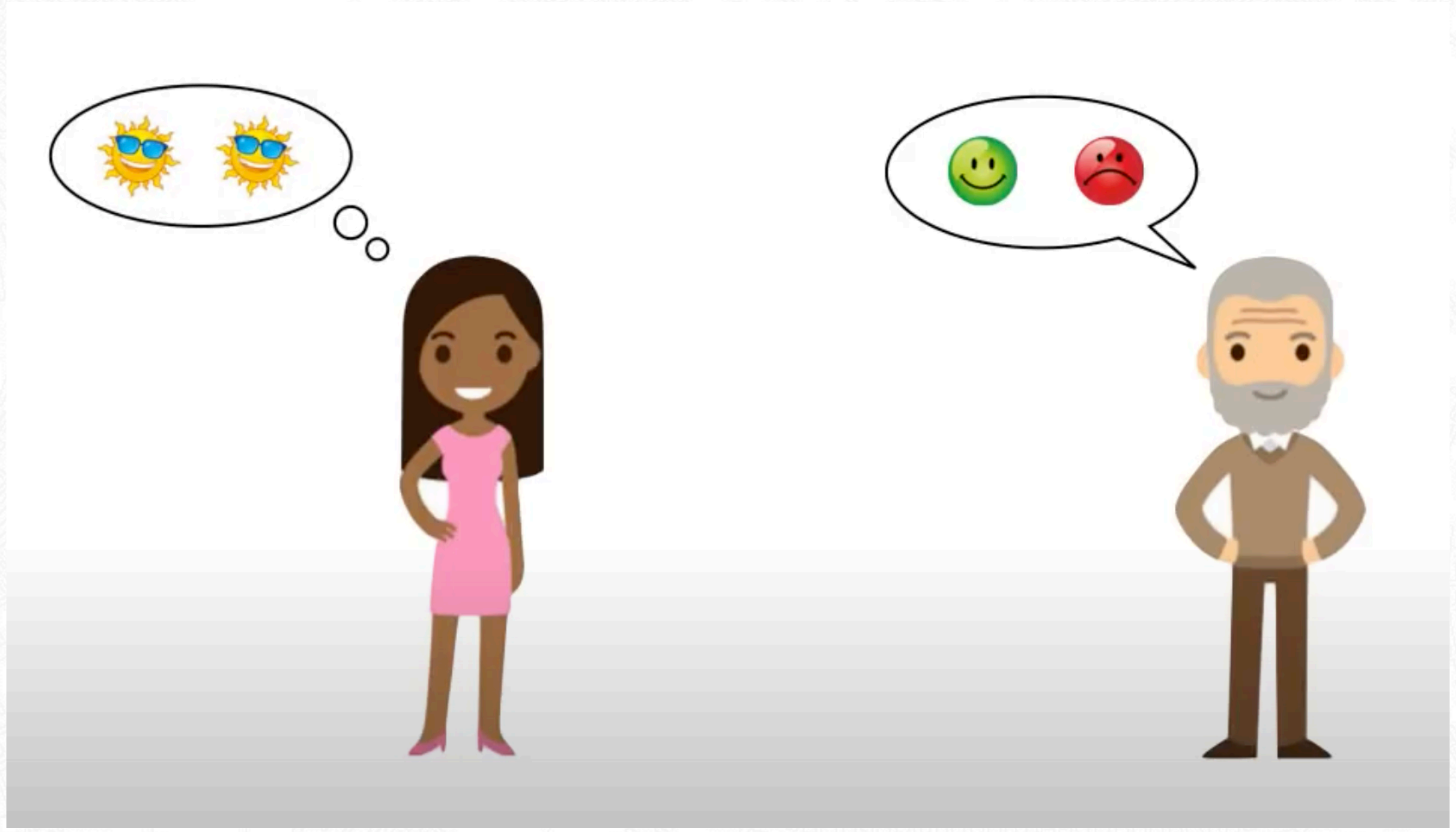
Al final se selecciona la secuencia que tiene una mayor probabilidad posterior. (Maximum Likelihood)



Source: https://www.youtube.com/watch?v=kqSzLo9fenk

Ejemplo

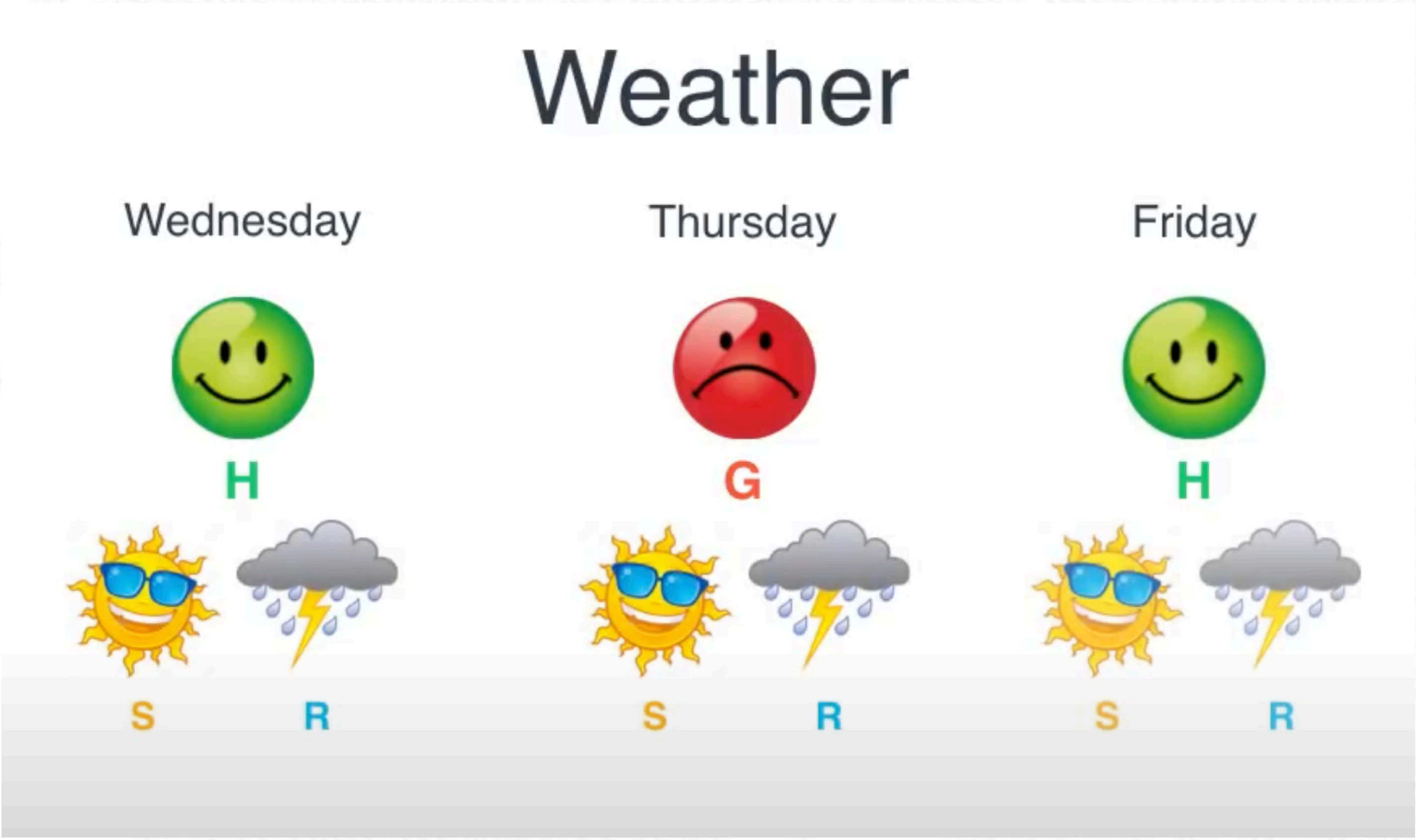
Por lo tanto, basado en la evidencia, y las observaciones del estado de animo de Bob, la secuencia de estados del clima más probable es que ambos días hayan sido soleados.

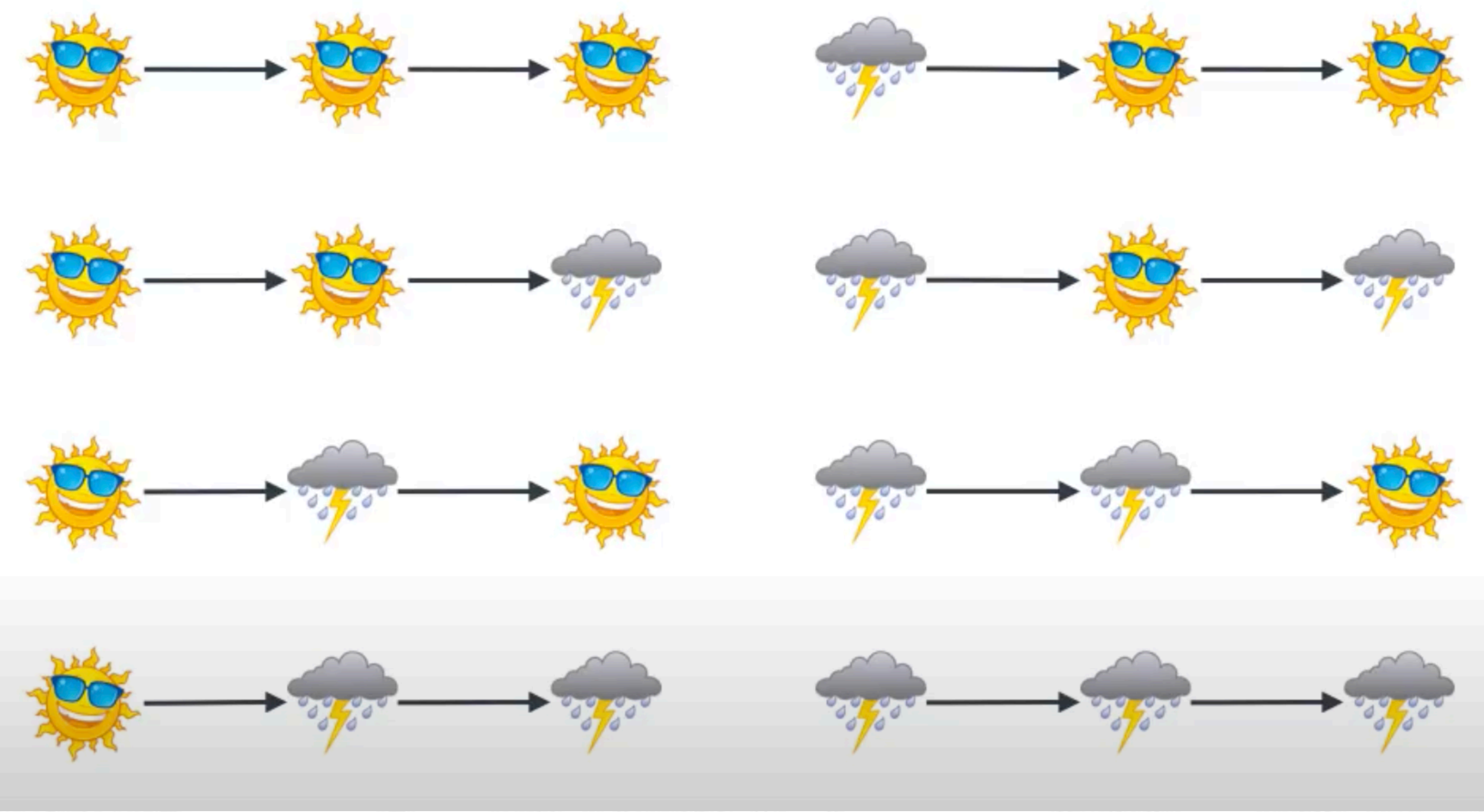


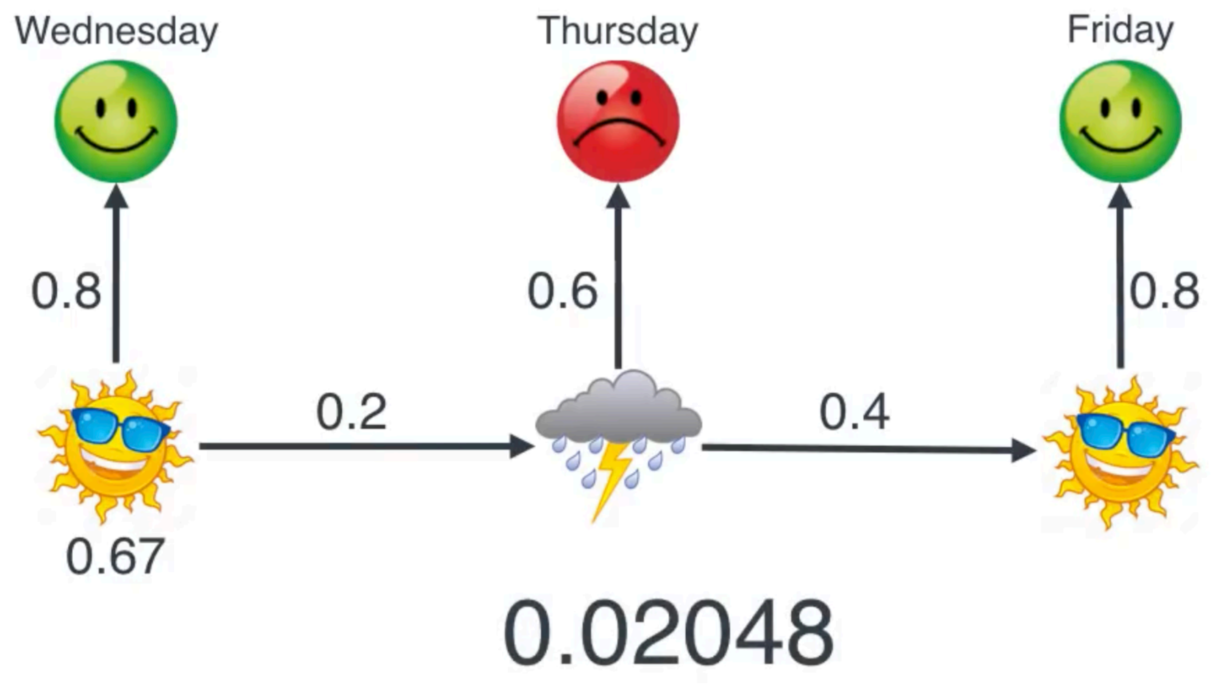
Source: https://www.youtube.com/watch?v=kqSzLo9fenk

Ejemplo

¿Qué pasa si queremos saber cuál es la secuencia de estados del clima, dada 3 observaciones del estado de animo de Bob?





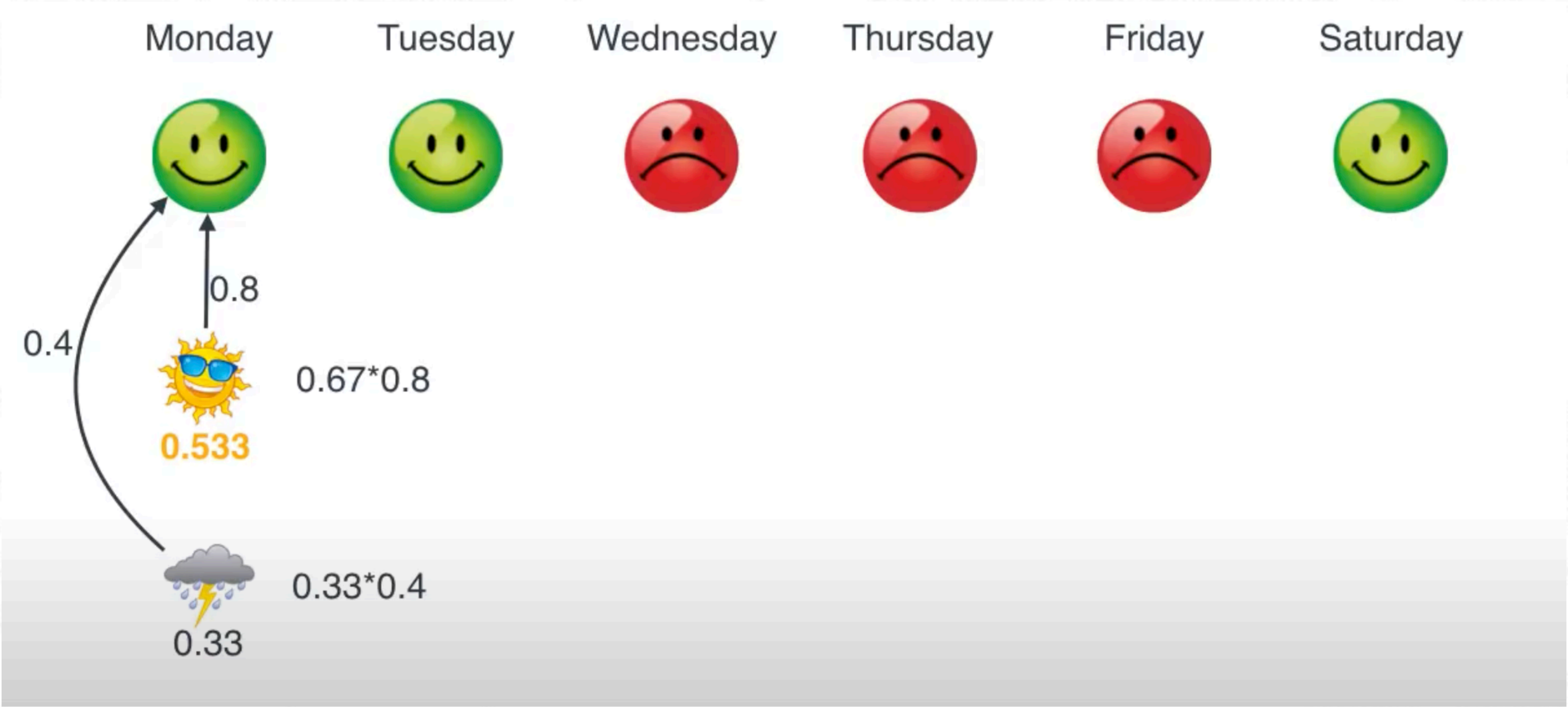


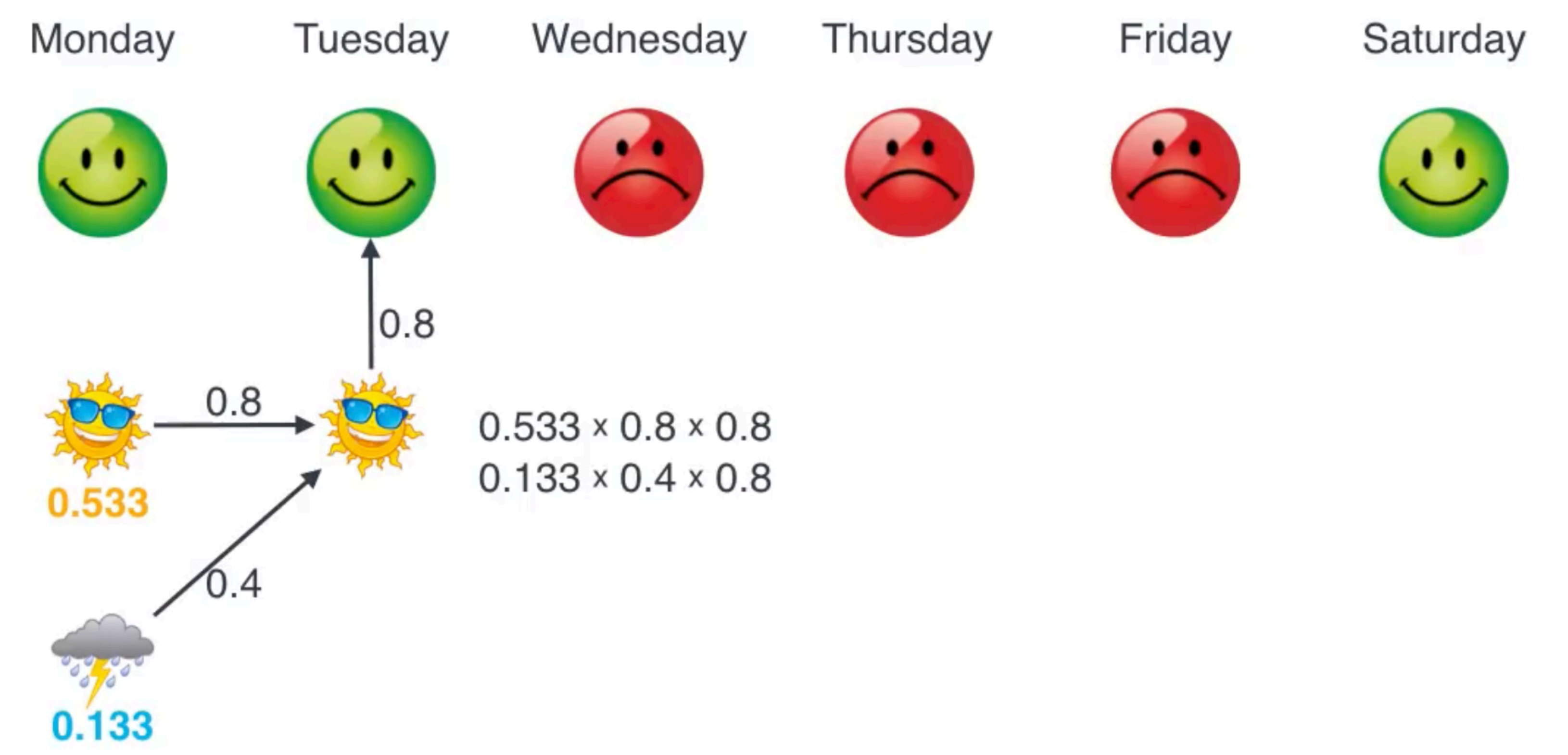
En este caso debemos evaluar 8 posibles combinaciones. El número de combinaciones a evaluar crece de forma exponencial.

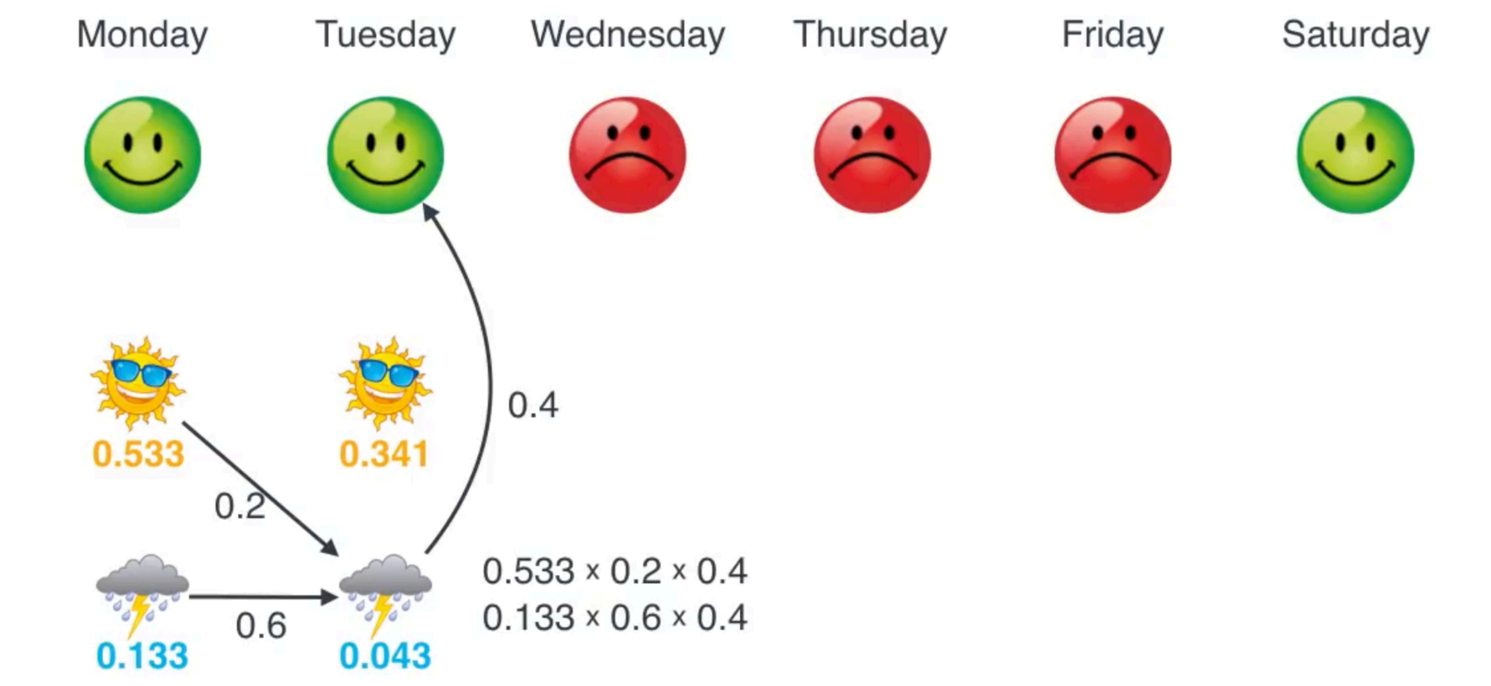
Source: https://www.youtube.com/watch?v=kqSzLo9fenk

Ejemplo

Para evitar realizar todas las posibles combinaciones, se utiliza el algoritmo **Viterbi**. Se empieza a calcular las probabilidades para cada estado, y solo se mantiene la mayor de las probabilidades en cada posición de la secuencia.



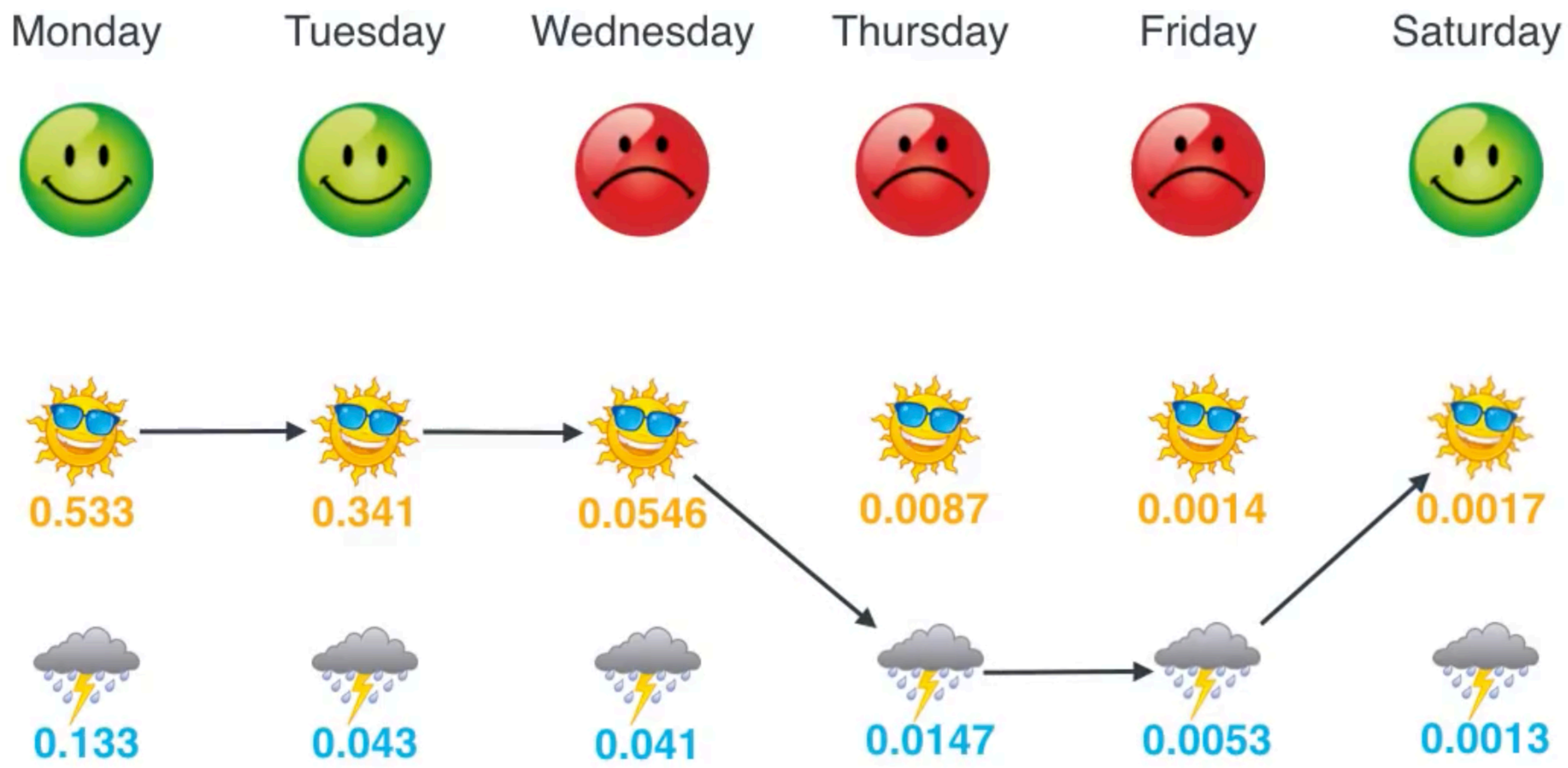


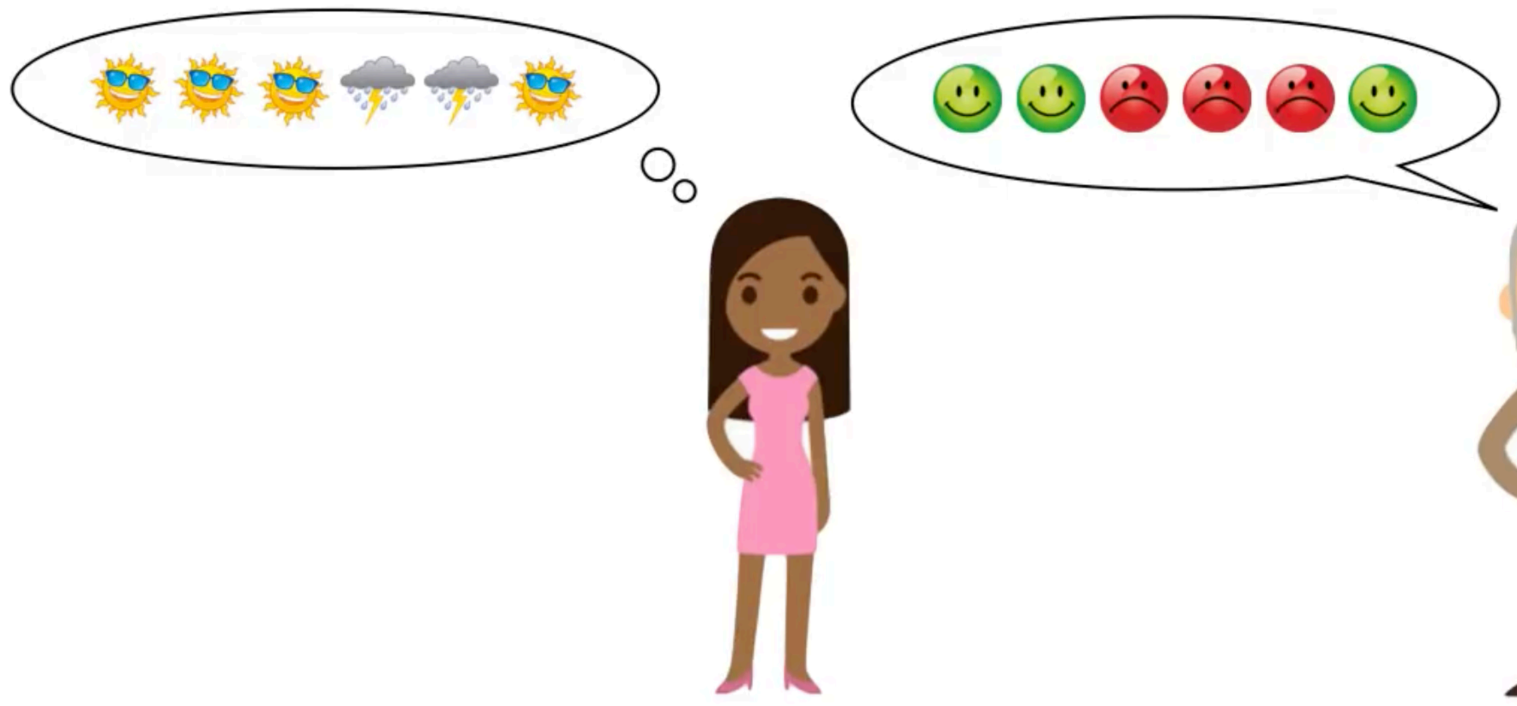


Source: https://www.youtube.com/watch?v=kqSzLo9fenk

Ejemplo

Al final, para cada punto de la secuencia, se escoge el estado que tenga l mayor probabilidad posterior. De esa forma se puede encontrar la secuencia de estados del clima, más probable, que produjeron las observaciones del estado de animo de Bob.





Source: https://www.youtube.com/watch?v=kqSzLo9fenk

Contenido

- HMM.
- HMM para ASR systems.

Hidden Markov Models

¿Pero cómo se aplica esto para Speech Recognition?

1. Suponga que N es el número de fonemas en el lenguaje.
2. Se modela las transiciones entre los fonemas (incluyendo el silencio) basado en datos escritos de entrenamiento.
3. Se tienen datos de pronunciación de los textos escritos, se realiza análisis mel-cepstral.
4. Se calcula la probabilidad que para un estado determinado (un fonema escrito) corresponda un vector mel-cepstral determinado.
5. Para un espectograma determinado, se determina cual es la secuencia más probable que produzca este conjunto de observaciones. (Esto puede ser computacionalmente muy costoso).
6. Una vez se tienen las secuencias de fonemas, se busca en un diccionario las palabras correspondientes, y finalmente se pasa por un modelo de lenguaje.

Algoritmo Viterbi

1. **Initialization**

$$
\begin{aligned}
\text{Set } t &= 2; \\
\delta_1(i) &= \pi_i b_i(\mathbf{o}_1), \quad 1 \leq i \leq N \\
\psi_1(i) &= 0, \quad 1 \leq i \leq N
\end{aligned}
$$

2. **Induction**

$$
\begin{aligned}
\delta_t(j) &= b_j(\mathbf{o}_t) \max_{1 \leq i \leq N} \delta_{t-1}(i) a_{ij}, \quad 1 \leq j \leq N \\
\psi_t(j) &= \arg \max_{1 \leq i \leq N} \left[\delta_{t-1}(i) a_{ij}\right], \quad 1 \leq j \leq N
\end{aligned}
$$

3. **Update time**

   Set $t = t + 1$;
   Return to step 2 if $t \leq T$;
   Otherwise, terminate the algorithm (goto step 4).

4. **Termination**

$$
\begin{aligned}
P^* &= \max_{1 \leq i \leq N} \left[\delta_T(i)\right] \\
q_T^* &= \arg \max_{1 \leq i \leq N} \left[\delta_T(i)\right]
\end{aligned}
$$

5. **Path (state sequence) backtracking**

   (a) **Initialization**

   Set $t = T - 1$

   (b) **Backtracking**

$$
q_t^* = \psi_{t+1}\left(q_{t+1}^*\right)
$$

   (c) **Update time**

   Set $t = t - 1$;
   Return to step (b) if $t \geq 1$;
   Otherwise, terminate the algorithm.

Un problema es que se trabaja con producto de probabilidades, para evitar errores numéricos se puede trabajar con el logaritmo de estas probabilidades.

Source: https://www.diva-portal.org/smash/get/diva2:831263/FULLTEXT01.pdf

Gracias

Preguntas?