

©2009 CBS INTERACTIVE

JOHDATUS TEKOÄLYYN

LUENTO 4.

TODENNÄKÖISYYSMALLINNUS II:

BAYESIN KAAVA

TEEMU ROOS



HELSINGIN YLIOPISTO

PINGVIINI(tweety) :- true.

Wulffmorgenthaler



HS 14.9.2012

TODENNÄKÖISYYS (TN)

EHDOLLINEN TN:

$P(B | A)$ "B:N TODENNÄKÖISYYS, KUN TIEDETÄÄN, ETTÄ A"
"B:N EHDOLLINEN TN ANNETTUNA A"

$$P(\text{"1. SILMÄLUKU ON 5"} | \text{"1. SILMÄLUKU EI 6"}) = \underline{\underline{1/5}}$$

$$P(\text{"1. SILMÄLUKU ON 5"} | \text{"2. SILMÄLUKU EI 6"}) = \underline{\underline{1/6}}$$

$$P(B | A) = P(A, B) / P(A), \quad \text{KUN } P(A) > 0$$

$$\Rightarrow P(A, B) = P(A) \times \underline{\underline{P(B|A)}}$$

TODENNÄKÖISYYS (TN)

1. $P(A, B) = P(A) P(B|A) = P(B) P(A|B)$ // KETJUSÄÄNTÖ

2. $P(A, B, C) = P(A, B) P(C|A, B) = P(A) P(B|A) P(C|A, B)$
// KETJUSÄÄNTÖ

3. $P(A) = P(A, B) + P(A, \neg B)$ // "MARGINALISOINTI"

4. $P(A|B) = P(A, B) / P(B)$ // EHDOLLINEN TN.

5. $A \perp B \Rightarrow P(A|B) = P(A)$ // RIIPPUMATTOMUUS

6. $A \perp B | C \Rightarrow P(A|B, C) = P(A | C)$
// EHDOLLINEN RIIPPUMATTOMUUS

7. $P(B|A) = P(B) P(A|B) / P(A)$ // BAYESIN KAAVA

TODENNÄKÖISYYS (TN)

- * SATUNNAISMUUTTUJA (SM), ESIM. NOPAN SILMÄLUKU, ON MUUTTUJA, JONKA ARVO MÄÄRÄYTYY ALKEISTAPAHTUMAN PERUSTEELLA
- * SM X ON ALKEISTAPAHTUMAN FUNKTIO $X: \Omega \rightarrow X(\Omega)$.
- * $X(\Omega)$ ON X :N ARVOJOUKKO: $X(\omega) \in X(\Omega)$ KAIKILLA $\omega \in \Omega$
- * ESIM.

KAHDEN NOPAN HEITOSSA ALKEISTAPAHTUMIEN JOUKKO ON $\Omega = \{(m,n) \mid m,n \in \{1,2,3,4,5,6\}\}$

SILMÄLUKIJEN SUMMA: $\omega = (m,n) \rightarrow X(\omega) = m+n$
 $X(\Omega) = \{2,3,4,5,6,7,8,9,10,11,12\}$

TODENNÄKÖISYYS (TN)

MUUTTUJAN NIMET

ARVOT

ISOLLA

PIENELLÄ

* SATUNNAISMUUTTUJEN AVULLA VOIDAAN MÄÄRITELLÄ

UUSIA TAPAHTUMIA:

$$A = "X=X"$$

$$B = "X < 4"$$

$$C = "X=X, Y < 5"$$

$$D = "X > Y"$$

* ESIM.

KAHDEN NOPAN HEITOSSA ALKEISTAPAHTUMIEN JOUKKO

$$\text{ON } \Omega = \{(m, n) \mid m, n \in \{1, 2, 3, 4, 5, 6\}\}$$

SILMÄLUKIJEN SUMMA:

$$\omega = (m, n) \rightarrow X(\omega) = m + n$$

$$X(\Omega) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

TODENNÄKÖISYYS (TN)

- * SATUNNAISMUUTTUJIEN AVULLA VOIDAAN MÄÄRITELLÄ UUSIA TAPAHTUMIA:
 $A = "X=X"$ $B = "X < 4"$
 $C = "X=X, Y < 5"$ $D = "X > Y"$

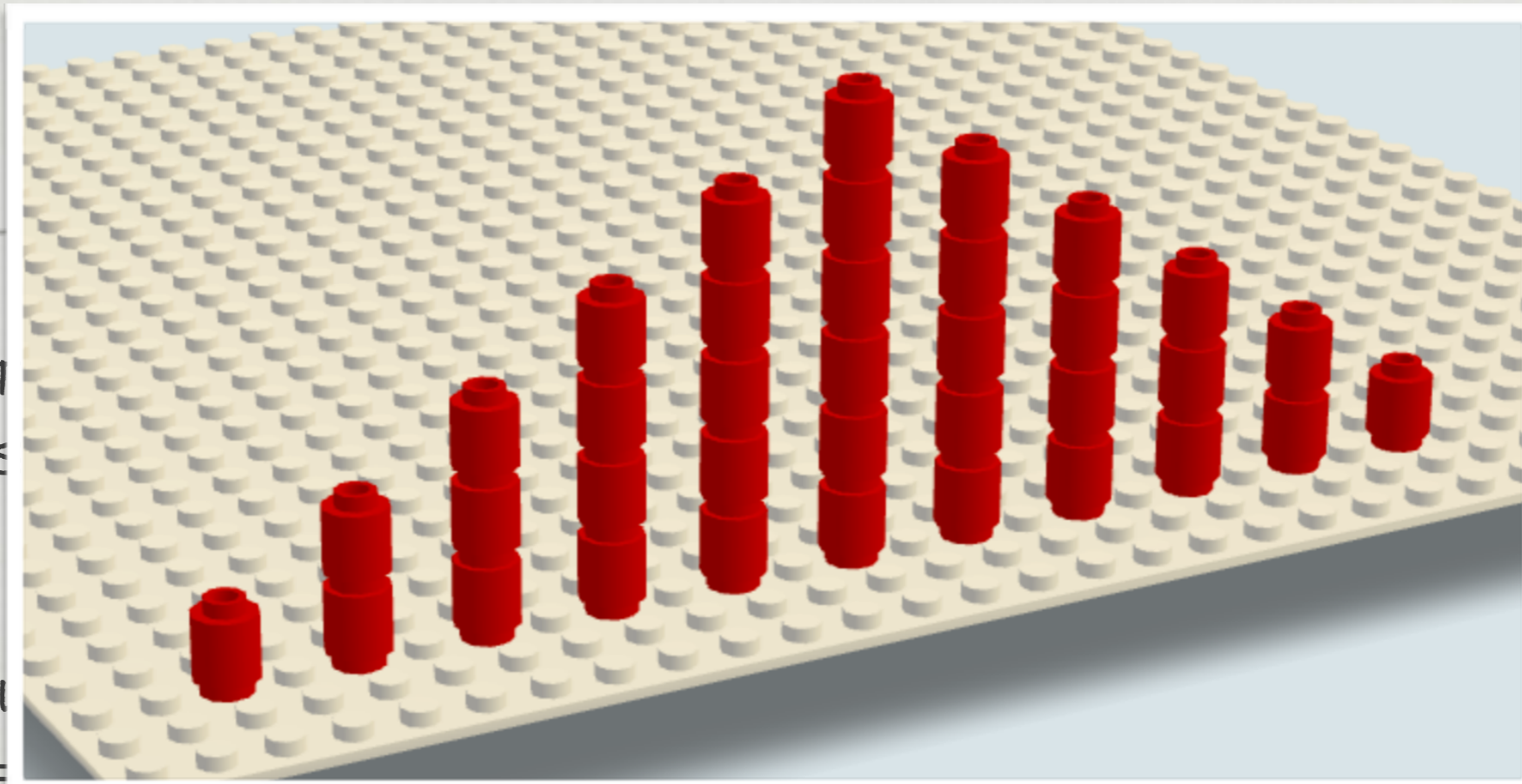
- * MUUTTUJILLA ON JAKAUMA:

$$P_X = (P(X=2), P(X=3), P(X=4), \dots, P(X=12))$$
$$= (1/36, 2/36, 3/36, 4/36, 5/36, 6/36, 5/36, 4/36, 3/36, 2/36, 1/36)$$

- * ESIM.

KAHDEN NOPAN HEITOSSA ALKEISTAPAHTUMIEN JOUKKO ON $\Omega = \{(m,n) \mid m,n \in \{1,2,3,4,5,6\}\}$

SILMÄLUKIJEN SUMMA: $\omega = (m,n) \rightarrow X(\omega) = m+n$
 $X(\Omega) = \{2,3,4,5,6,7,8,9,10,11,12\}$



SAT
LUK

TELLÄ

"
"



MU
 $P_x =$

$$= (1/36, 2/36, 3/36, 4/36, 5/36, 6/36, 5/36, 4/36, 3/36, 2/36, 1/36)$$



ESIM.

KAHDEN NOPAN HEITOSSA ALKEISTAPAHTUMIEN JOUKKO
ON $\Omega = \{(m, n) \mid m, n \in \{1, 2, 3, 4, 5, 6\}\}$

SILMÄLUKIJEN SUMMA:

$$\omega = (m, n) \rightarrow X(\omega) = m + n$$

$$X(\Omega) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

MUUTTUJAT VS TAPAHTUMAT

* TAPAHTUMIEN LASKUSÄÄNNÖT, KUTEN KETJUSÄÄNTÖ, PÄTEVÄT MYÖS MUUTTUJIEN AVULLA MÄÄRITELTYIHIN TAPAHTUMIIN

* ESIM.

$$P(X=x, Y=y) = P(X=x) \times \underline{P(Y=y | X=x)} \quad \forall x, y$$

* USEIN LYHENNETÄÄN

$$P(x, y) = P(x) P(y|x) \quad \forall x, y$$

BAYES?

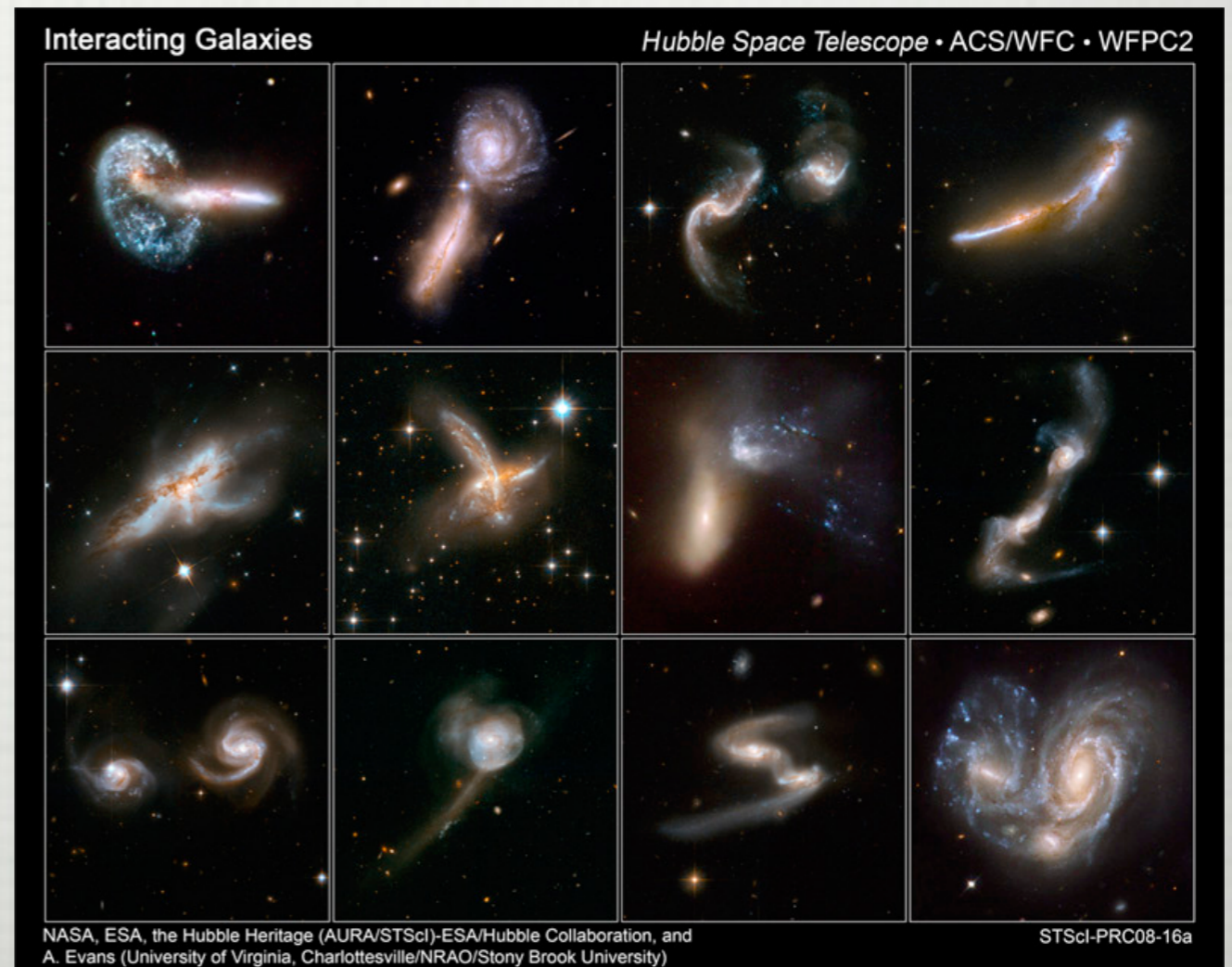
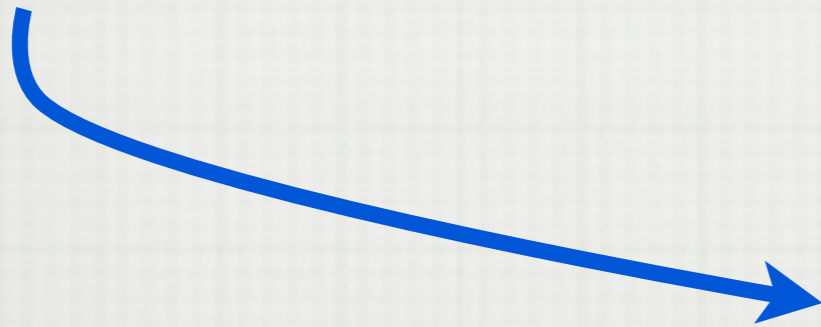


T. BAYES

BAYES?

* $P(B|A) = P(B) P(A|B) / P(A)$

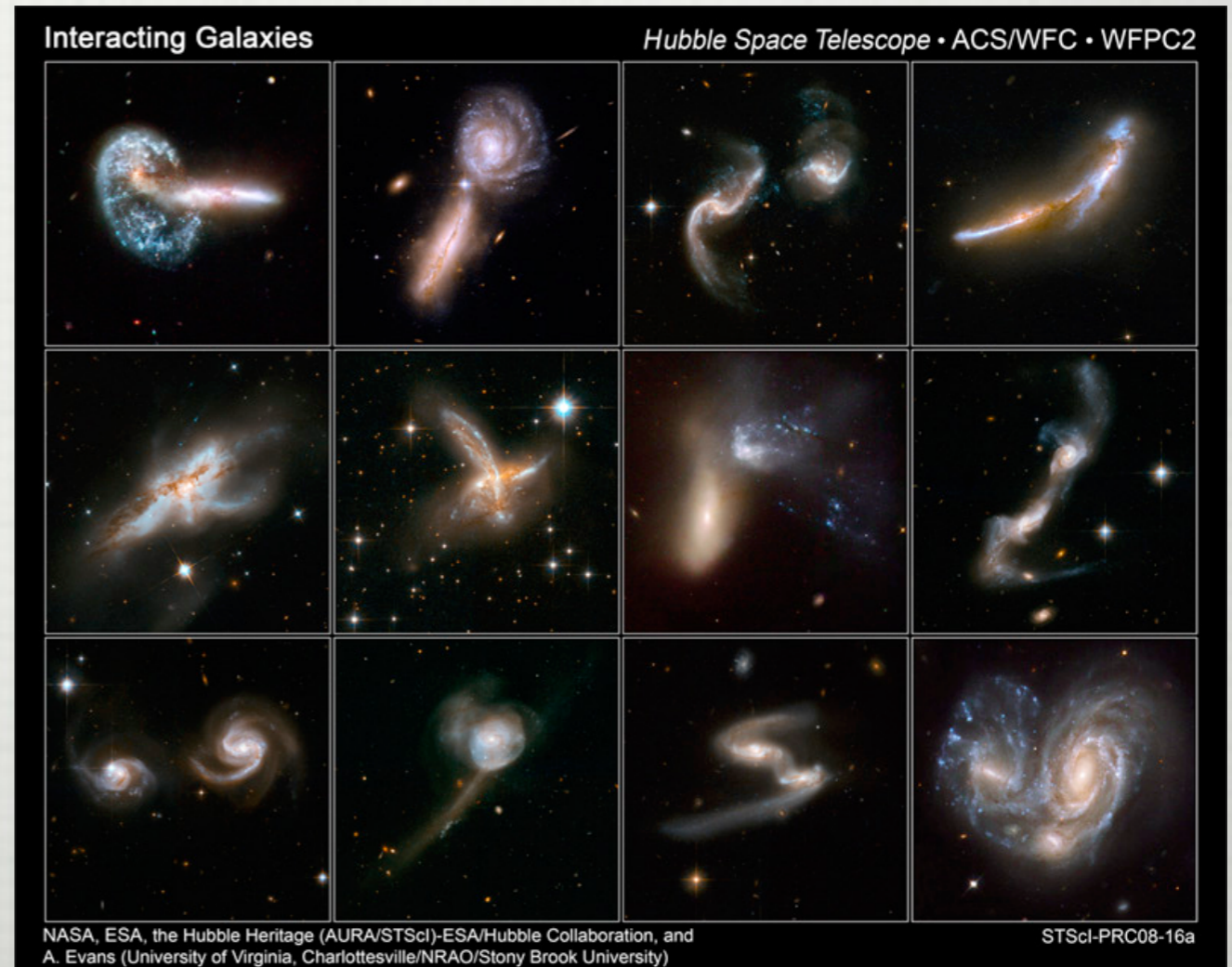
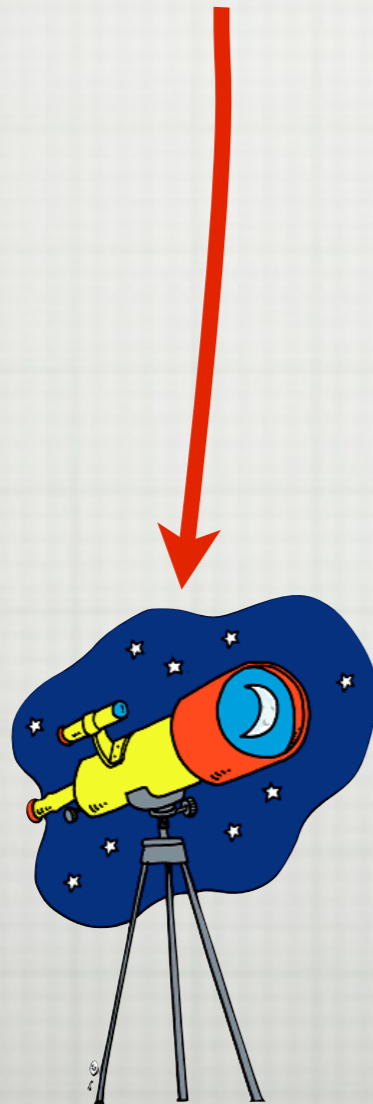
* $P(\text{tíla} | \text{havainto}) = P(\text{tíla}) P(\text{havainto} | \text{tíla}) / P(\text{havainto})$



BAYES?

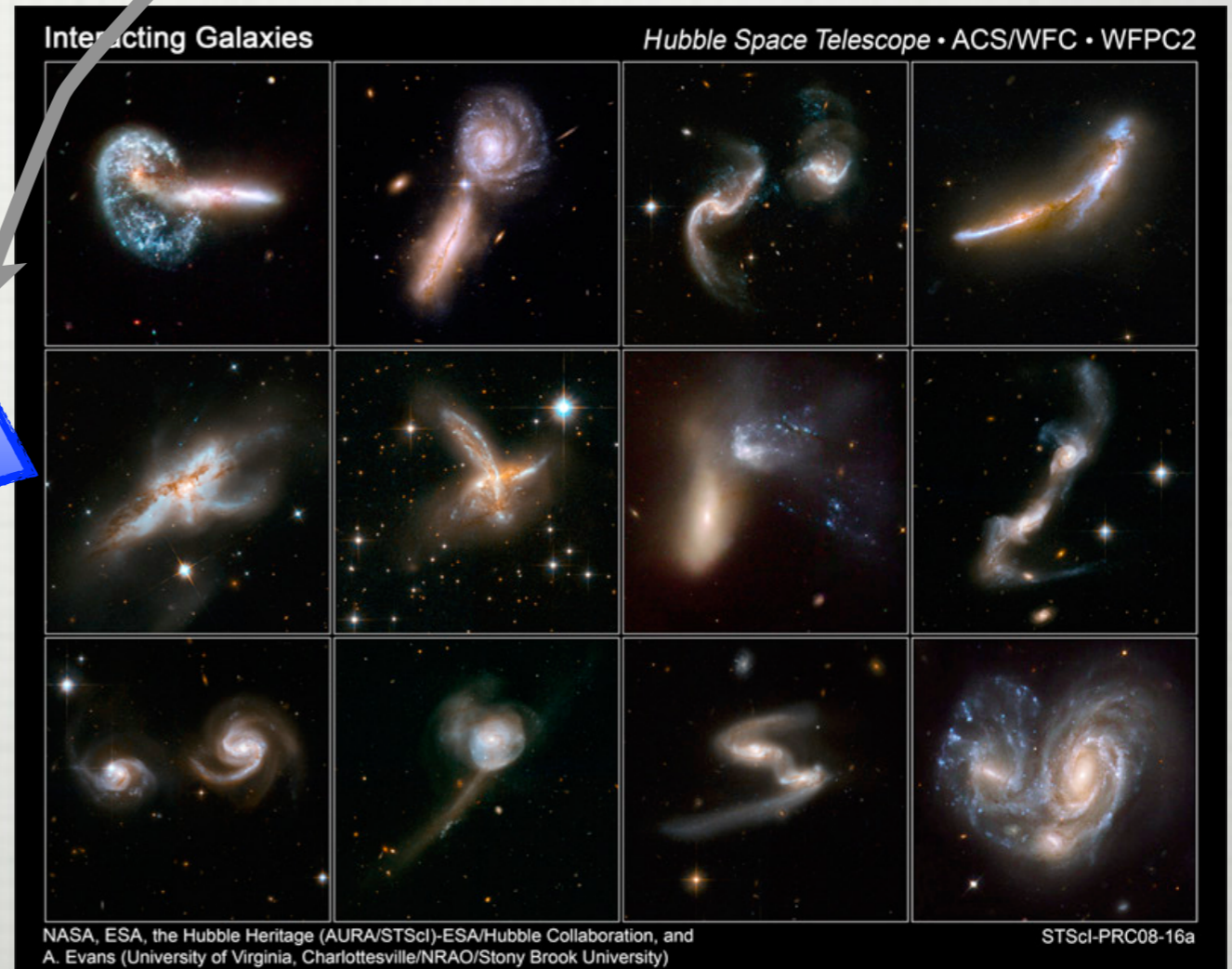
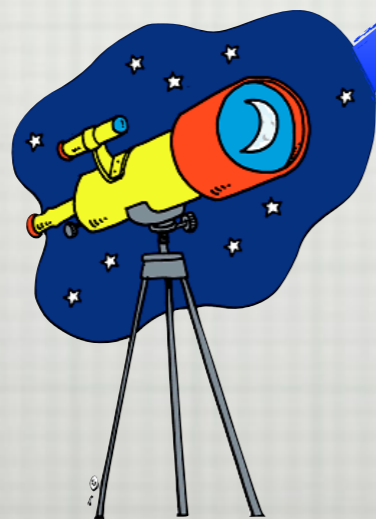
* $P(B|A) = P(B) P(A|B) / P(A)$

* $P(\text{tíla} | \text{havainto}) = P(\text{tíla}) P(\text{havainto} | \text{tíla}) / P(\text{havainto})$



BAYES?

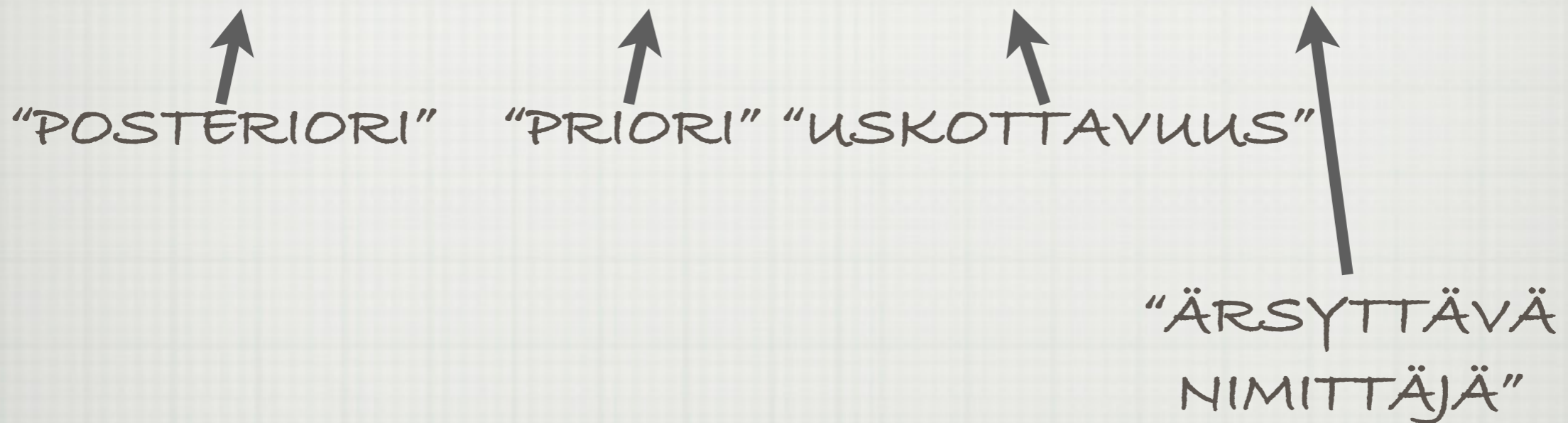
- * $P(B|A) = P(B) P(A|B) / P(A)$
- * $P(\text{tíla} | \text{havainto}) = P(\text{tíla}) P(\text{havainto} | \text{tíla}) / P(\text{havainto})$



BAYES?

* $P(B|A) = P(B) P(A|B) / P(A)$

* $P(\text{tila} | \text{havainto}) = P(\text{tila}) P(\text{havainto} | \text{tila}) / P(\text{havainto})$



BAYES?

* $P(B|A) = P(B) P(A|B) / P(A)$

* $P(\text{tíla} | \text{havainto}) = P(\text{tíla}) P(\text{havainto} | \text{tíla}) / P(\text{havainto})$



BAYES?

- * $P(B|A) = P(B) P(A|B) / P(A)$
- * $P(\text{tíla} | \text{havainto}) = P(\text{tíla}) P(\text{havainto} | \text{tíla}) / P(\text{havainto})$



BAYES?

- * $P(\text{tíla} | \text{havainto}) = P(\text{tíla}) P(\text{havainto} | \text{tíla}) / P(\text{havainto})$
- * $\text{Tíla} \in \{\text{sairas}, \text{terve}\}$ // potilas sairas tai terve
- * $\text{Havainto} \in \{\text{pos}, \text{neg}\}$ // testi positiivinen tai negatiivinen
- * $P(\text{sairas}) = 0.001$ $P(\text{terve}) = \underline{0.999}$
 $P(\text{pos} | \text{sairas}) = 0.9$ $P(\text{pos} | \text{terve}) = 0.01$
 $P(\text{neg} | \text{sairas}) = \underline{0.1}$ $P(\text{neg} | \text{terve}) = \underline{0.99}$
- * $P(\text{sairas} | \text{pos}) = \underline{0.08264463 \approx 8\%}$

SEURAAVAKSI

- **TODENNÄKÖISYYSMALLINNUK**

- **SPAM-SUODATIN**

- **BAYES-VERKOT**

