Introducción a la Inferencia Bayesiana

con Aplicaciones en Ciencias Cognitivas

Introducción a la Inferencia Bayesiana

con Aplicaciones en Ciencias Cognitivas

¡el debut!

Operacionales

- Días y horarios
- Evaluación
- Licenciatura/doctorado, carrera.
- Correlativas
- ¿Otras?

Programa
(grano grueso)

- Primera parte: inferencia bayesiana
- Segunda parte: los seres humanos como máquinas de inferencia

Bibliografía

- Gelman *et al*.
  Bayesian Data Analysis
- MacKay
  Information Theory, Inference, and Learning Algorithms
- Jaynes
  Probability Theory: The Logic of Science
- Wagenmakers & Lee
  A Course in Bayesian Graphical Modeling for Cognitive Science
- Marr, Anderson, ...

Experimento mental (gedankenexperiment)

Si tiramos 6 veces la moneda,
y sale 6 veces cara...

¿diríamos que la moneda está cargada?

Si tiramos 6 veces la moneda,
y sale 6 veces cara...

¿diríamos que la moneda está cargada?

La “ciencia” hoy dice “*sí*”.

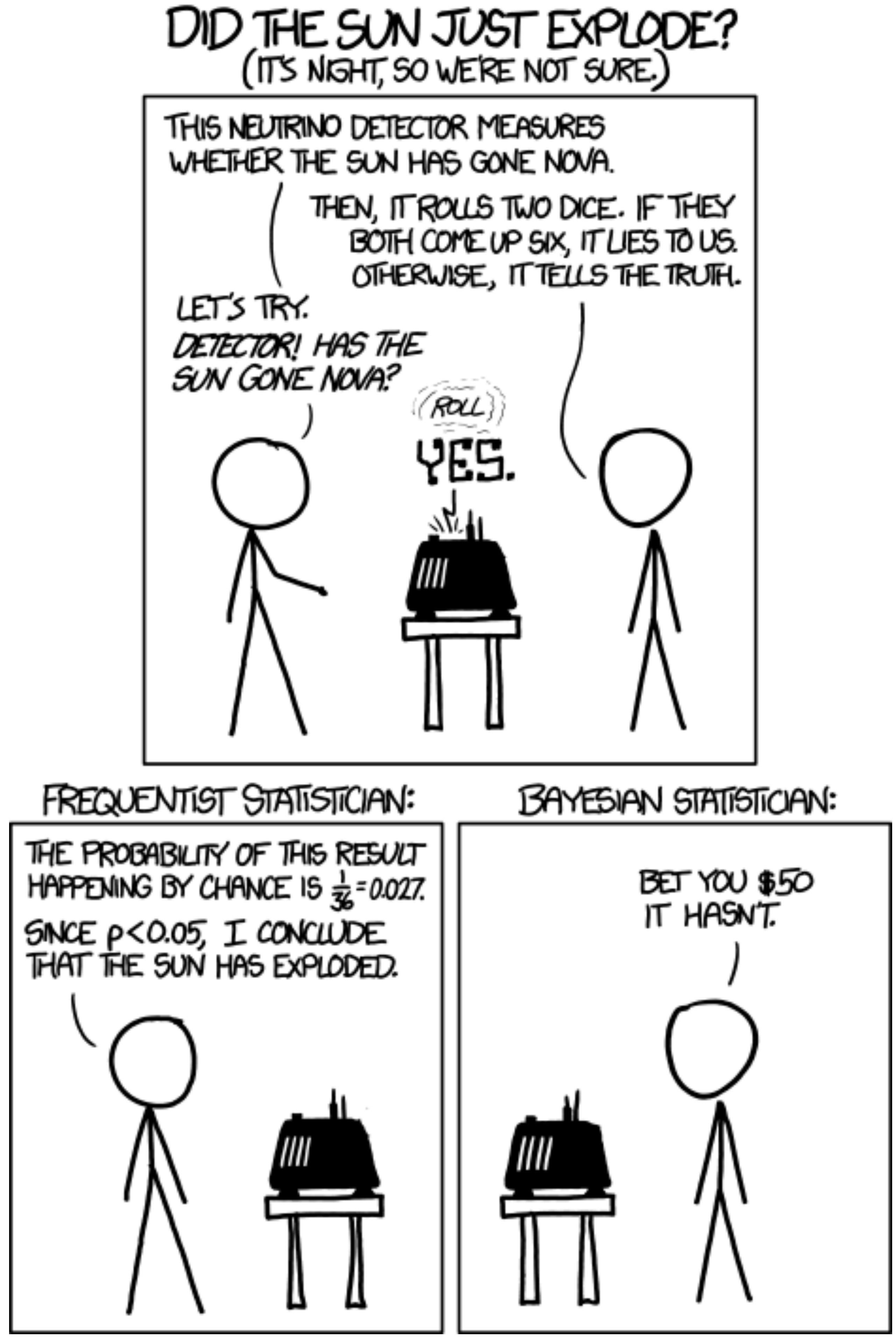
Si tiramos 6 veces la moneda,
y sale 6 veces cara...

¿diríamos que la moneda está cargada?

La “ciencia” hoy dice “*sí*”.

Más precisamente:

“Hay evidencia significativa de que está cargada ($p < 0.5$, binomial test)”


DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)
THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.
THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.
LET'S TRY.
DETECTOR! HAS THE SUN GONE NOVA?
ROLL
YES.
FREQUENTIST STATISTICIAN:
THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS $\frac{1}{36} = 0.027$.
SINCE $p < 0.05$, I CONCLUDE THAT THE SUN HAS EXPLODED.
BAYESIAN STATISTICIAN:
BET YOU $50 IT HASN'T.

Estadística (o inferencia, o análisis de datos) bayesiana como alternativa a la estadística frecuentista o “clásica”

Estadística (o inferencia, o análisis de datos) bayesiana como alternativa a la estadística frecuentista o “clásica”

Permite (entre otras cosas) incorporar nuestro conocimiento previo

Estadística (o inferencia, o análisis de datos) bayesiana como alternativa a la estadística frecuentista o "clásica"

Permite (entre otras cosas) incorporar nuestro conocimiento previo

Probabilidad *subjetiva*

Estadística (o inferencia, o análisis de datos) bayesiana como alternativa a la estadística frecuentista o "clásica"

Permite (entre otras cosas) incorporar nuestro conocimiento previo

Probabilidad *subjetiva*

Podemos hacer afirmaciones sobre cuán probable es que la moneda esté cargada (prohibido en frecuentismo)

inferencia bayesiana vs. estadística frecuentista

inferencia bayesiana vs. estadística frecuentista

$\updownarrow$ mayormente

probabilidad subjetiva vs. probabilidad objetiva

inferencia bayesiana vs. estadística frecuentista

mayormente

probabilidad subjetiva vs. probabilidad objetiva

interpretación de la probabilidad

inferencia bayesiana vs. estadística frecuentista

mayormente

probabilidad subjetiva vs. probabilidad objetiva

interpretación de la probabilidad

p(*lo que yo quiera*) p sólo para muestreo

inferencia bayesiana vs. estadística frecuentista

↕ mayormente

probabilidad subjetiva vs. probabilidad objetiva

interpretación de la probabilidad

p(*lo que yo quiera*)

p sólo para muestreo

"Bayesian statistics is about making probability statements, frequentist statistics is about evaluating probability statements." (Gelman)

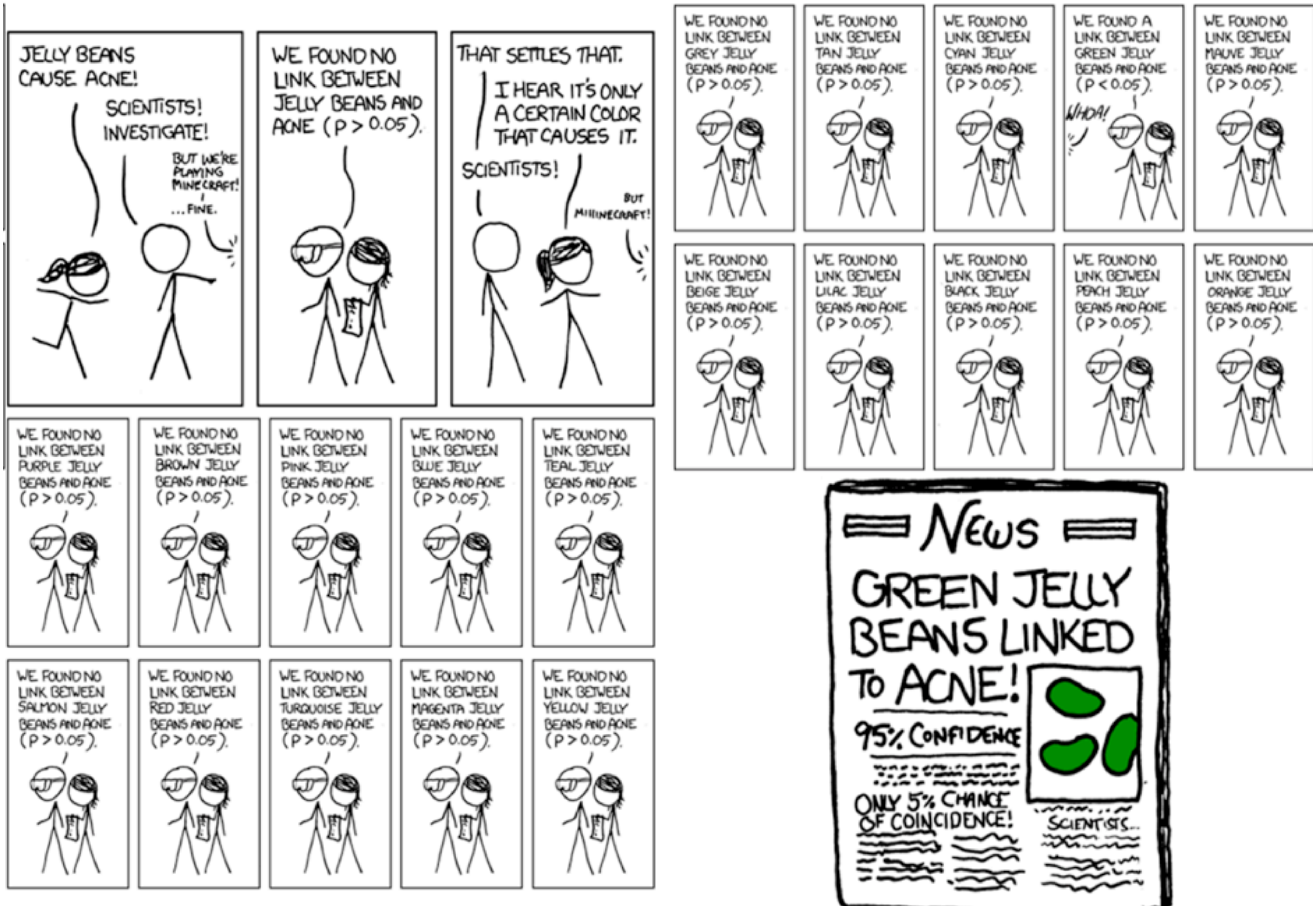
Críticas al frecuentismo: 2 niveles

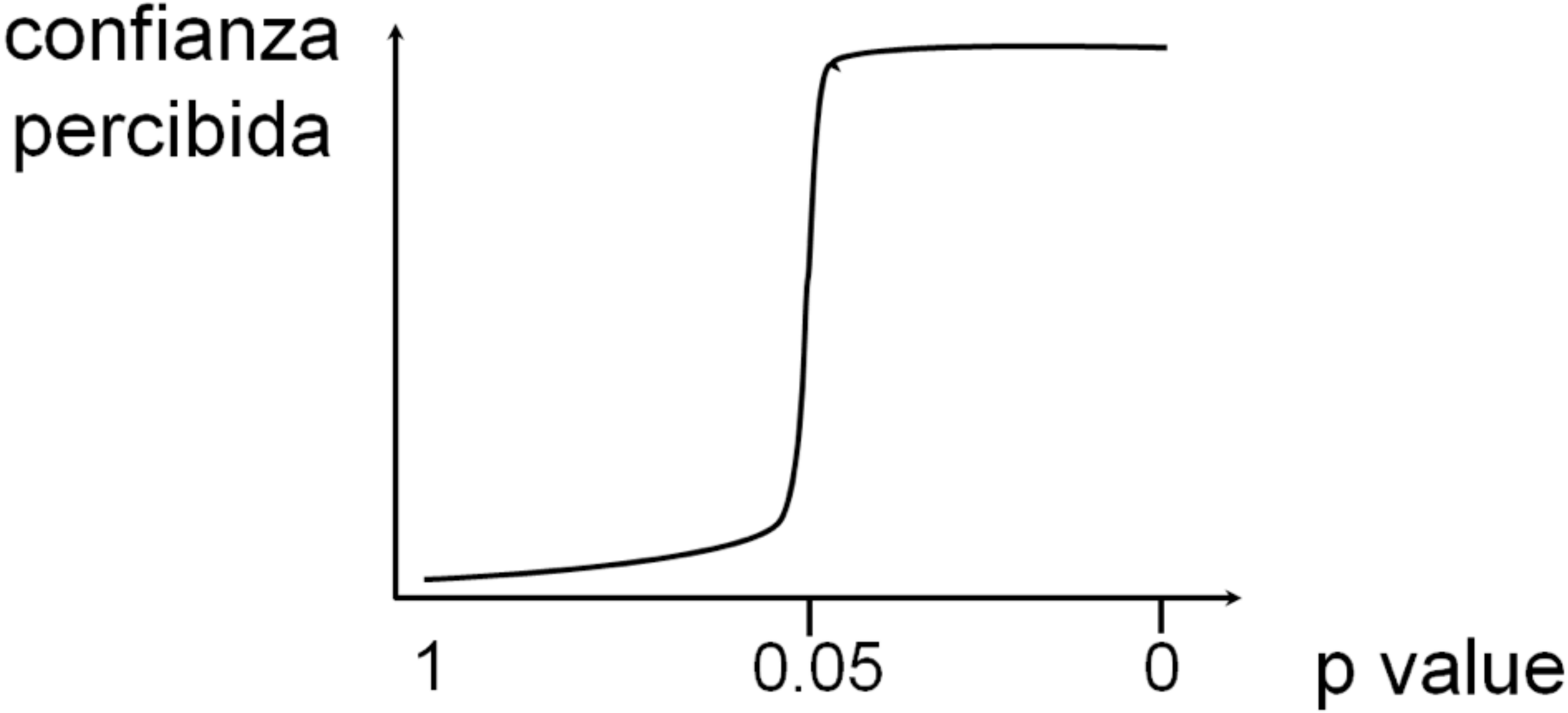
Críticas al frecuentismo: 2 niveles

- críticas de fondo:
  - ignora conocimiento previo
  - inconsistente
  - dilema Fisher (significance testing) vs Neyman/Pearson (hypothesis testing)
  - ... (discusión todavía abierta)

Críticas al frecuentismo: 2 niveles

- críticas de fondo:
  - ignora conocimiento previo
  - inconsistente
  - dilema Fisher (significance testing) vs Neyman/Pearson (hypothesis testing)
  - ... (discusión todavía abierta)
- críticas al (ab)uso:
  - *p hacking*
  - uso ciego en general
  - no específico del frecuentismo


JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
...FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MIIIINECRAFT!
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).
NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...


confianza
percibida
1
0.05
0
p value

Inferencia Bayesiana

$$p(H, D) = p(D|H)p(H) = p(H|D)p(D)$$

Inferencia Bayesiana

$$p(H,D) = p(D|H)p(H) = p(H|D)p(D)$$

$$\Downarrow$$

Inferencia Bayesiana

$$p(H, D) = p(D|H)p(H) = p(H|D)p(D)$$

$$\Downarrow$$

$$p(H|D) = \frac{p(D|H)p(H)}{p(D)}$$

Inferencia Bayesiana

$$p(H,D) = p(D|H)p(H) = p(H|D)p(D)$$

$$\Downarrow$$

$$p(H|D) = \frac{p(D|H)p(H)}{p(D)}$$

Teorema de Bayes

Inferencia Bayesiana

$$p(H, D) = p(D|H)p(H) = p(H|D)p(D)$$

$$\Downarrow$$

prior

$$p(H|D) = \frac{p(D|H)p(H)}{p(D)}$$

Teorema de Bayes

Inferencia Bayesiana

$$p(H, D) = p(D|H)p(H) = p(H|D)p(D)$$

$$\Downarrow$$

likelihood *prior*

$$p(H|D) = \frac{p(D|H)p(H)}{p(D)}$$

Teorema de Bayes

Inferencia Bayesiana

$$p(H,D) = p(D|H)p(H) = p(H|D)p(D)$$

$$\Downarrow$$

likelihood *prior*

posterior

$$p(H|D) = \frac{p(D|H)p(H)}{p(D)}$$

Teorema de Bayes

Inferencia Bayesiana

Inferencia Bayesiana

Modelo

Inferencia Bayesiana

Modelo

likelihood:

Inferencia Bayesiana

Modelo

likelihood:

$$p(k|\theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

Inferencia Bayesiana

Modelo

likelihood:

$$p(k|\theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$$

$$k \sim \mathrm{Binomial}(\theta, n)$$

Inferencia Bayesiana

Modelo

likelihood:

$$p(k|\theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

$$k \sim \mathrm{Binomial}(\theta, n)$$

prior:

Inferencia Bayesiana

Modelo

likelihood:

$$p(k|\theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

$$k \sim \text{Binomial}(\theta, n)$$

prior:

$$\theta \sim \text{Uniform}(0,1) = \text{Beta}(1,1)$$

Inferencia Bayesiana

Modelo

likelihood:

$$p(k|\theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$$

$$k \sim \mathrm{Binomial}(\theta, n)$$

prior:

$$\theta \sim \mathrm{Uniform}(0,1) = \mathrm{Beta}(1,1)$$

$$\theta \sim \mathrm{Beta}(100,100)$$

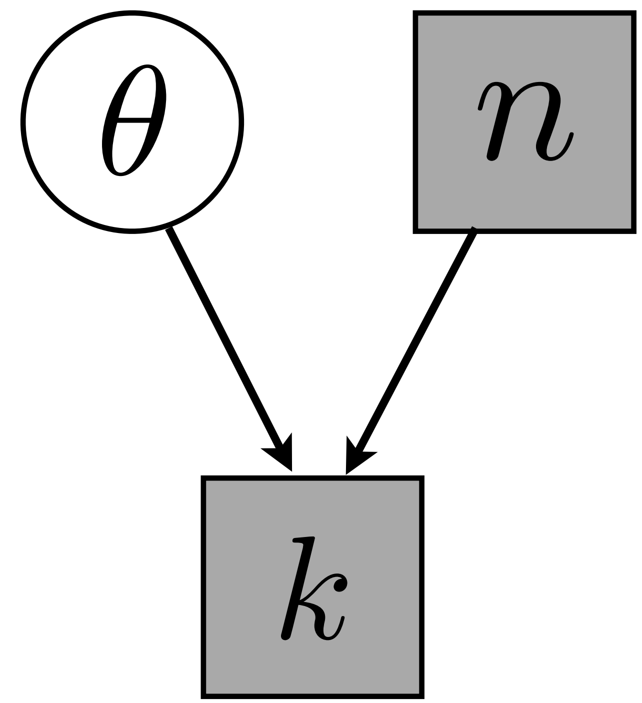
Inferencia Bayesiana

Modelo

likelihood:

$$p(k|\theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$$

$$k \sim \text{Binomial}(\theta, n)$$



prior:

$$\theta \sim \text{Uniform}(0,1) = \text{Beta}(1,1)$$

$$\theta \sim \text{Beta}(100,100)$$

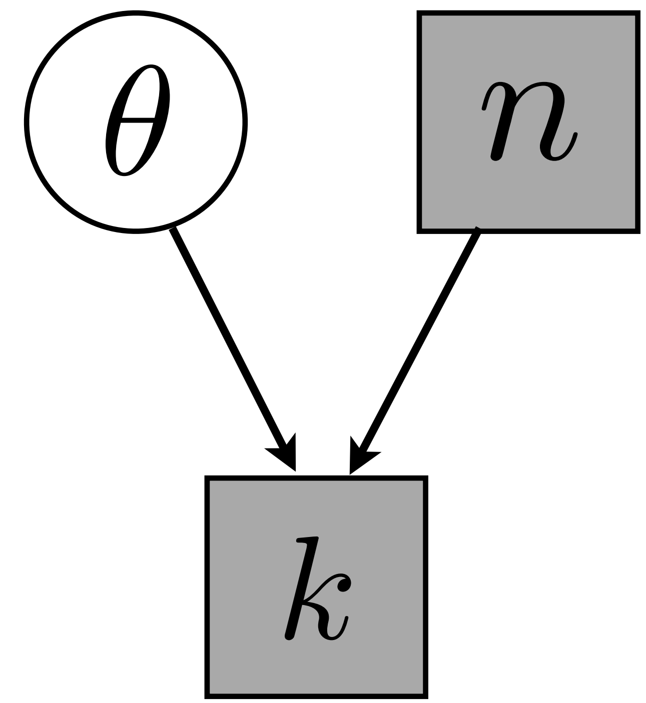
Inferencia Bayesiana

Modelo

likelihood:

$$p(k|\theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

$$k \sim \text{Binomial}(\theta, n)$$

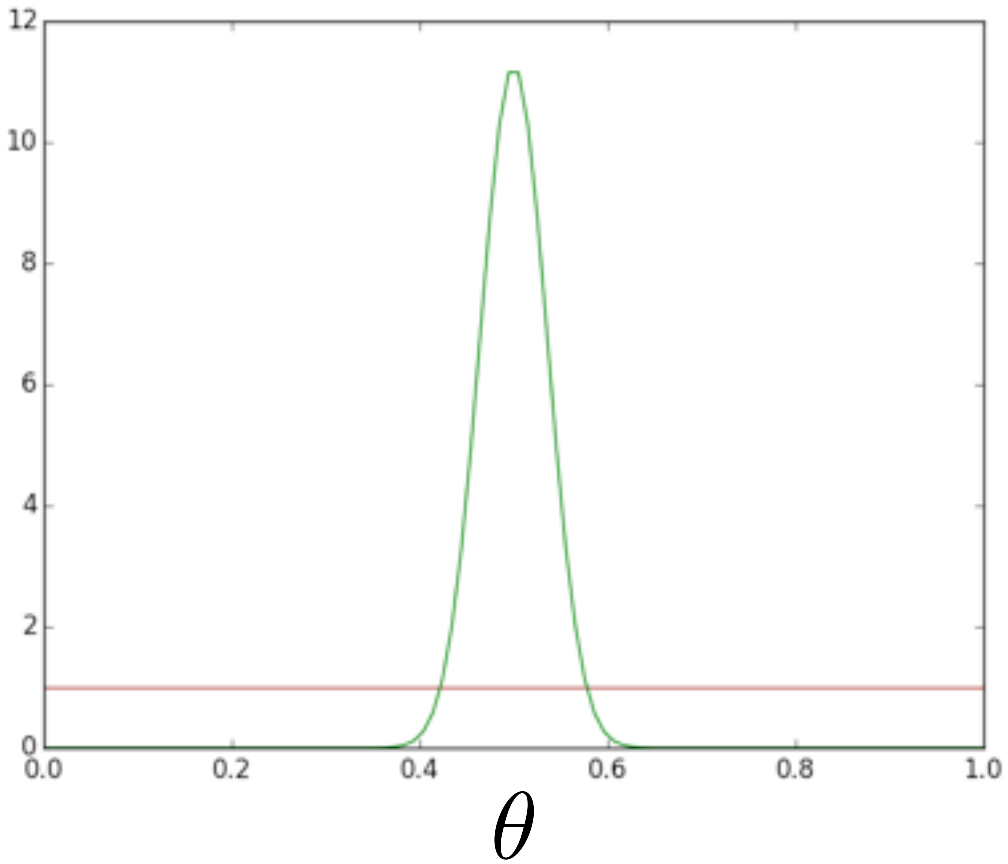


prior:

$$\theta \sim \text{Uniform}(0,1) = \text{Beta}(1,1)$$

$$\theta \sim \text{Beta}(100, 100)$$

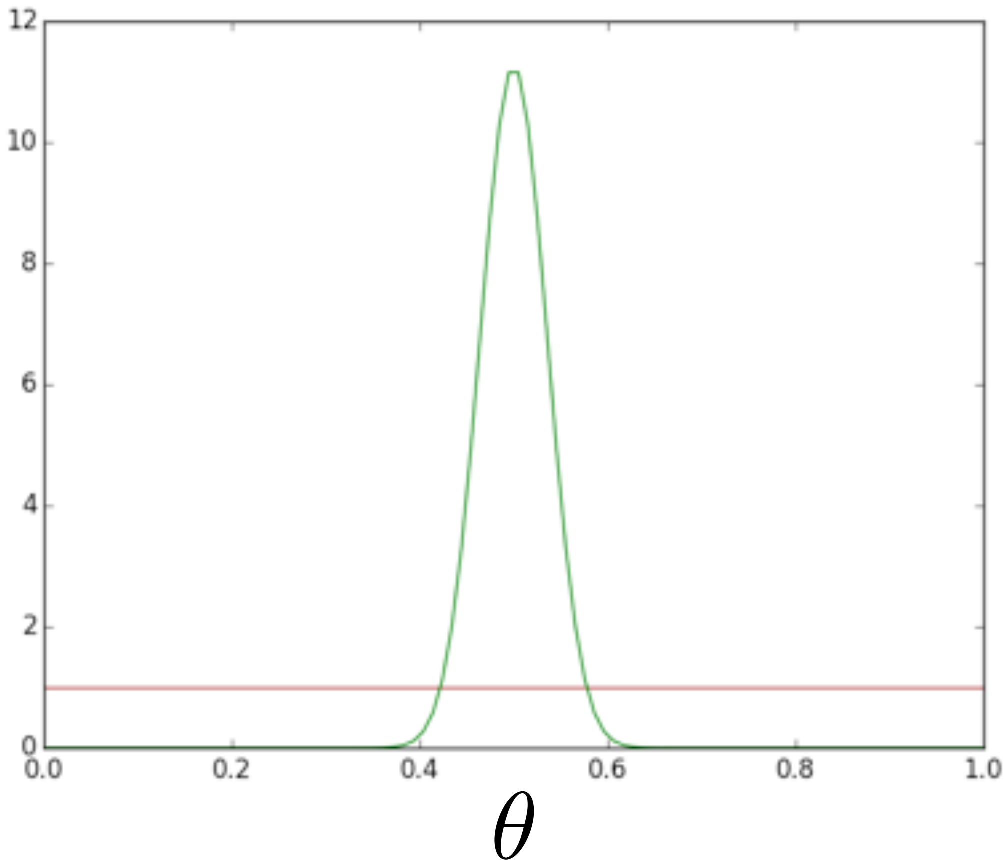
posterior con Bayes: $$p(\theta|k) = \frac{p(k|\theta)p(\theta)}{p(k)}$$



prior

$$\theta \sim \mathrm{Uniform}(0,1) = \mathrm{Beta}(1,1)$$

$$\theta \sim \mathrm{Beta}(100,100)$$

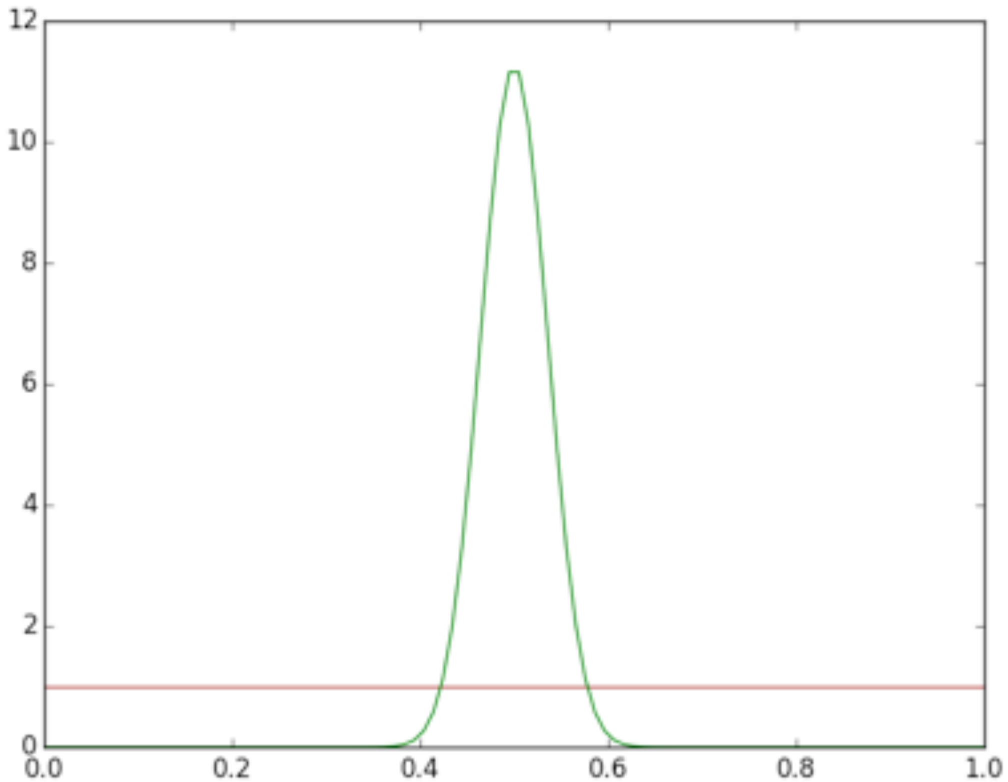


prior

$$\theta \sim \text{Uniform}(0, 1) = \text{Beta}(1, 1)$$

$$\theta \sim \text{Beta}(100, 100)$$

posterior
(luego de 6 caras)



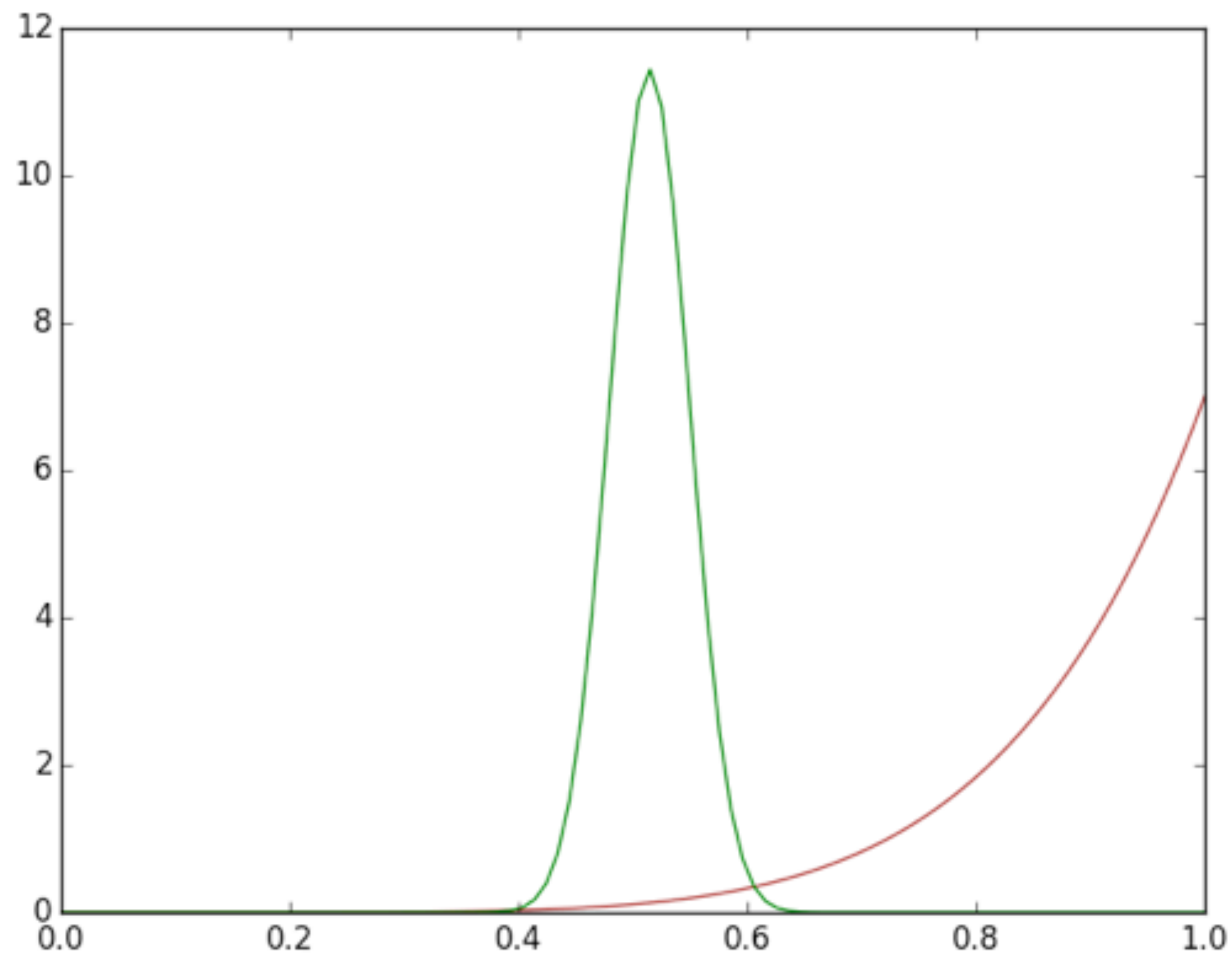
θ

prior

$\theta \sim \mathrm{Uniform}(0,1) = \mathrm{Beta}(1,1)$

$\theta \sim \mathrm{Beta}(100,100)$

posterior
(luego de 6 caras)



θ

Acumulación de Evidencia

$$p(H|D_1) = \frac{p(D_1|H)P(H)}{p(D_1)}$$

Acumulación de Evidencia

$$p(H|D_1) = \frac{p(D_1|H)P(H)}{p(D_1)}$$

$$p(H|D_2D_1) = \frac{p(D_2|HD_1)P(H|D_1)}{p(D_2|D_1)}$$

Acumulación de Evidencia

$$p(H|D_1) = \frac{p(D_1|H)P(H)}{p(D_1)}$$

$$p(H|D_2D_1) = \frac{p(D_2|HD_1)P(H|D_1)}{p(D_2|D_1)}$$

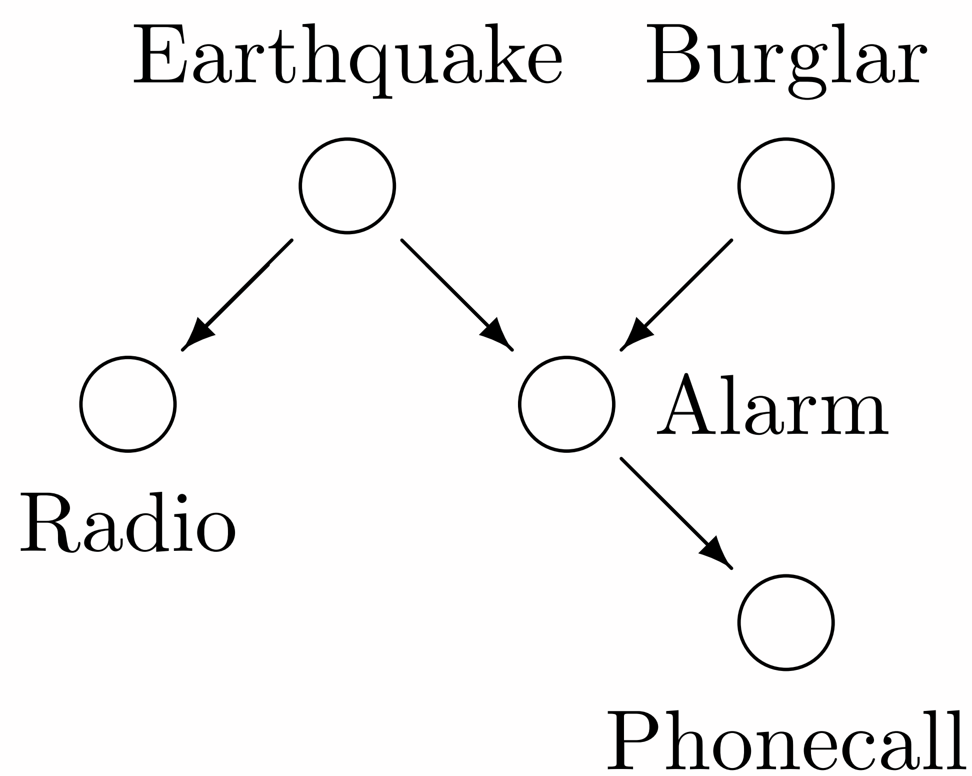
"The posterior is the new prior"

Redes Bayesianas

$$p(B, E, A, P, R) = p(B)p(E|B)p(A|B, E)p(P|A, B, E)p(R|P, A, B, E)$$

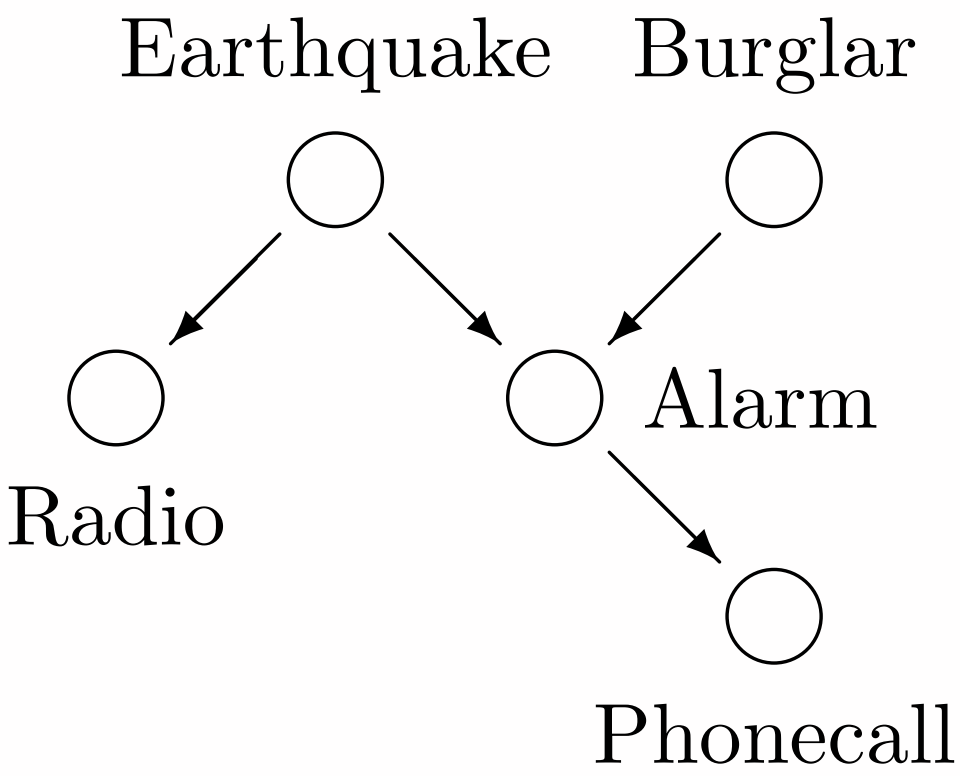
Redes Bayesianas

$$p(B, E, A, P, R) = p(B)p(E|B)p(A|B, E)p(P|A, B, E)p(R|P, A, B, E)$$



Redes Bayesianas

$$p(B, E, A, P, R) = p(B)p(E|B)p(A|B, E)p(P|A, B, E)p(R|P, A, B, E)$$



$$p(B, E, A, P, R) = p(B)p(E)p(A|B, E)p(P|A)p(R|E)$$

Modelos Jerárquicos

Teorías

$\downarrow$

Hipótesis

$\downarrow$

Datos

Modelos Jerárquicos

Teorías

↓

Hipótesis

↓

Datos

Form

↓

Structure

↓

Data

Modelos Jerárquicos

Teorías

↓

Hipótesis

↓

Datos

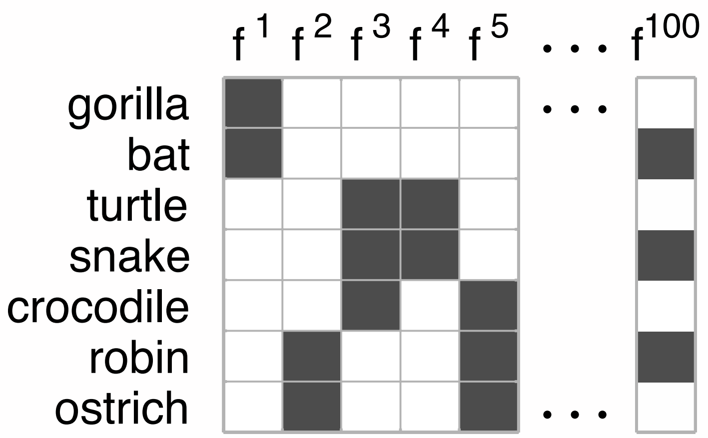
Form

↓

Structure

↓

Data



Modelos Jerárquicos

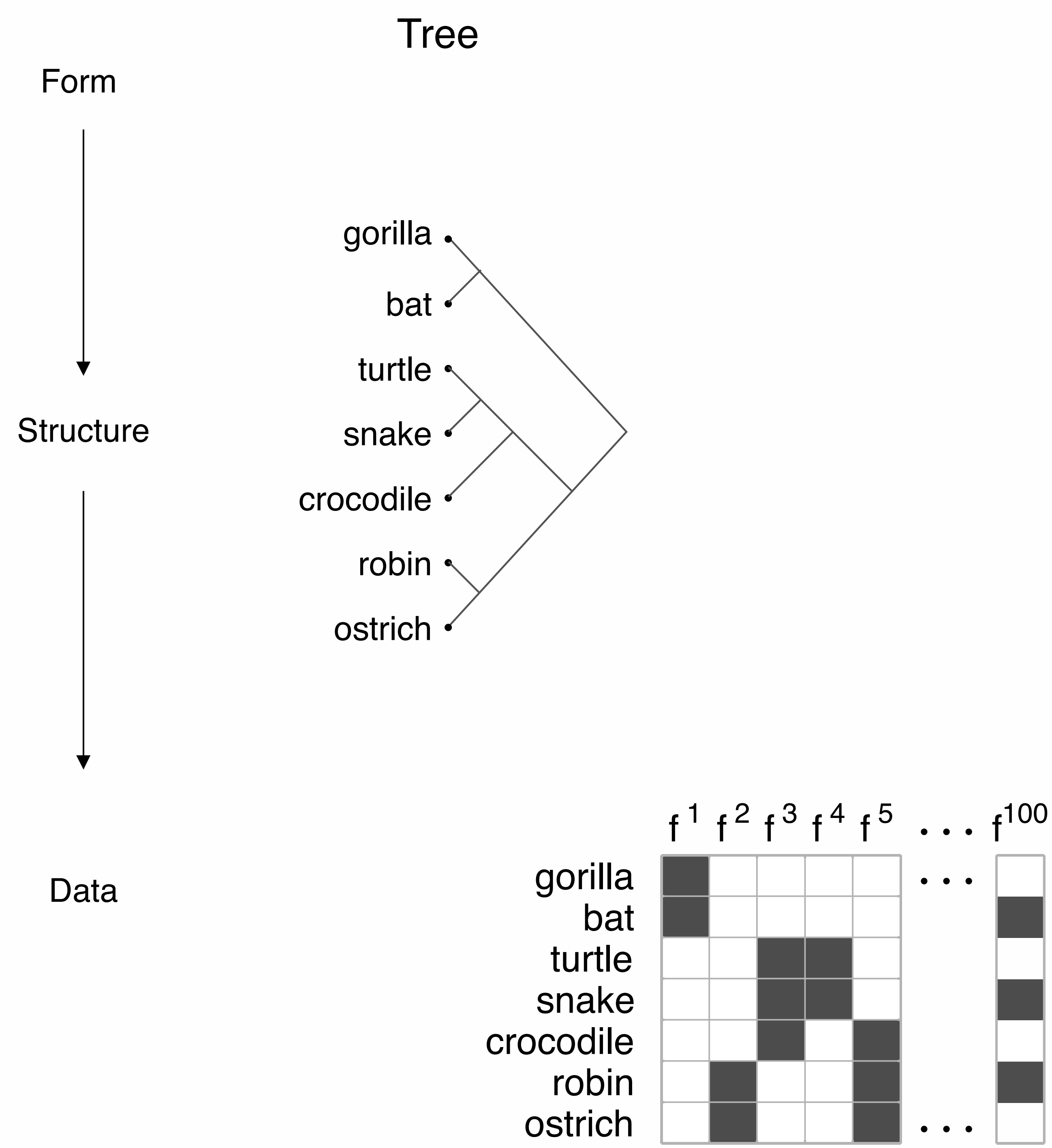
Teorías

↓

Hipótesis

↓

Datos



Modelos Jerárquicos

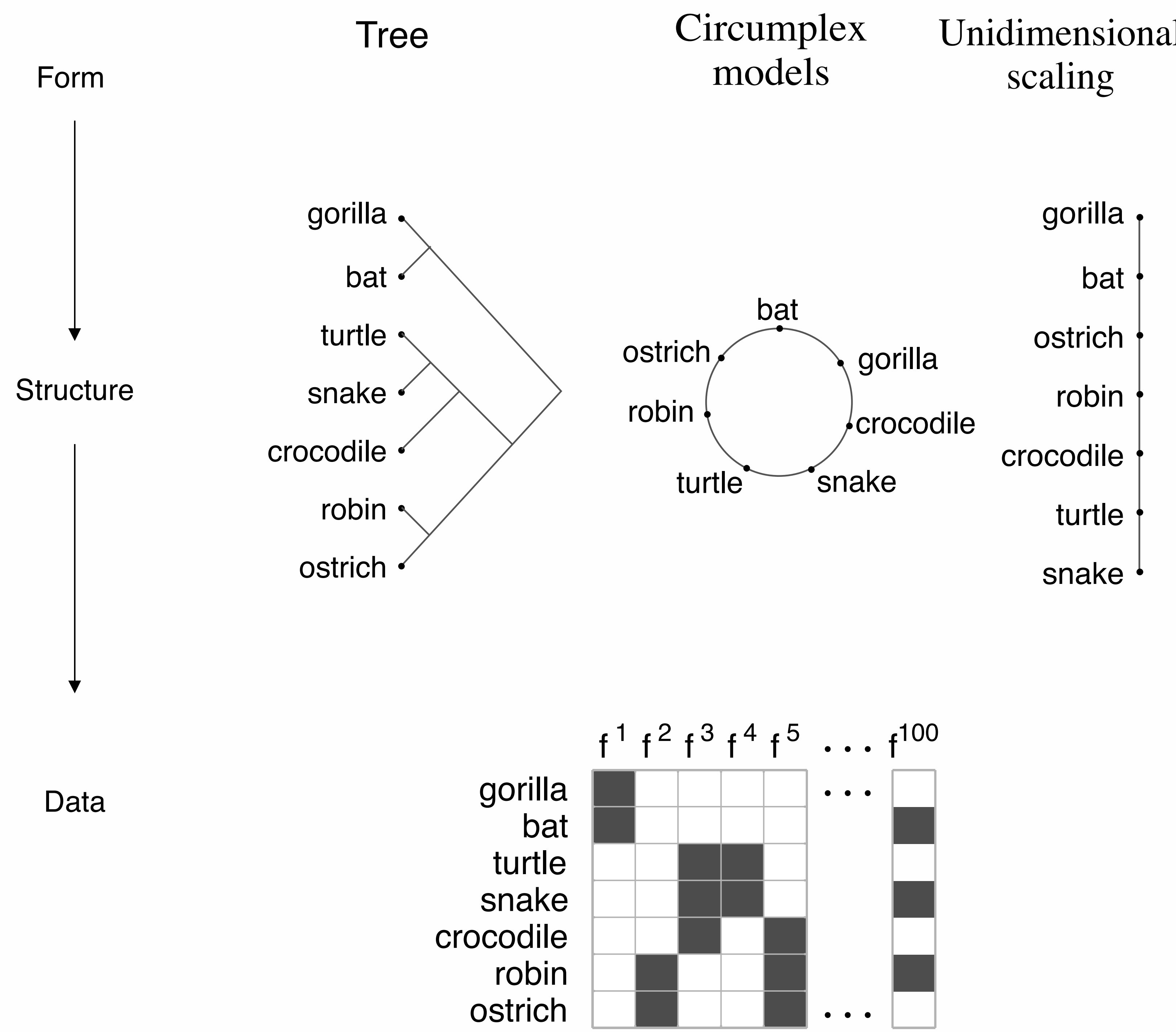
Teorías

↓

Hipótesis

↓

Datos



Modelos Jerárquicos

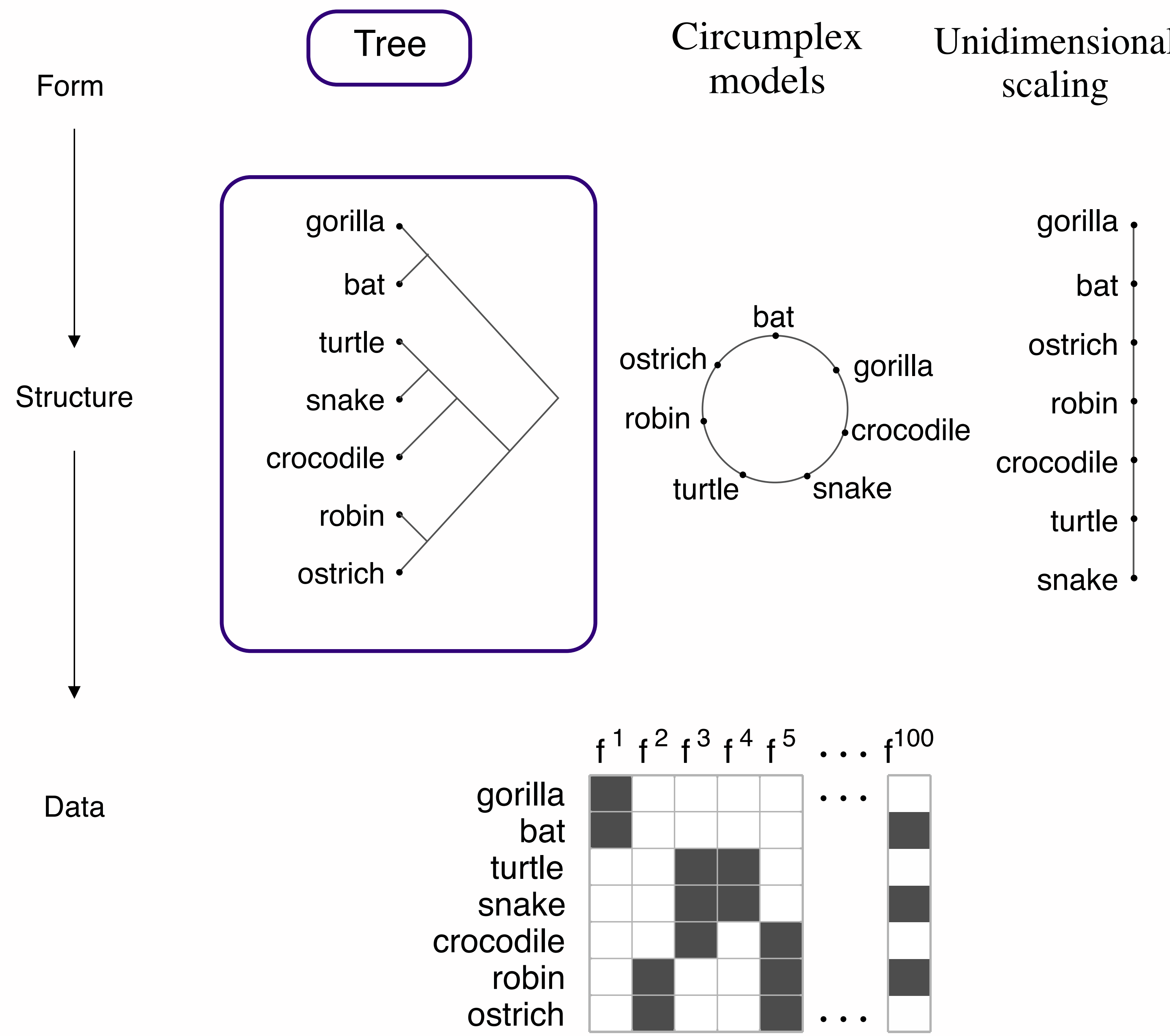
Teorías

↓

Hipótesis

↓

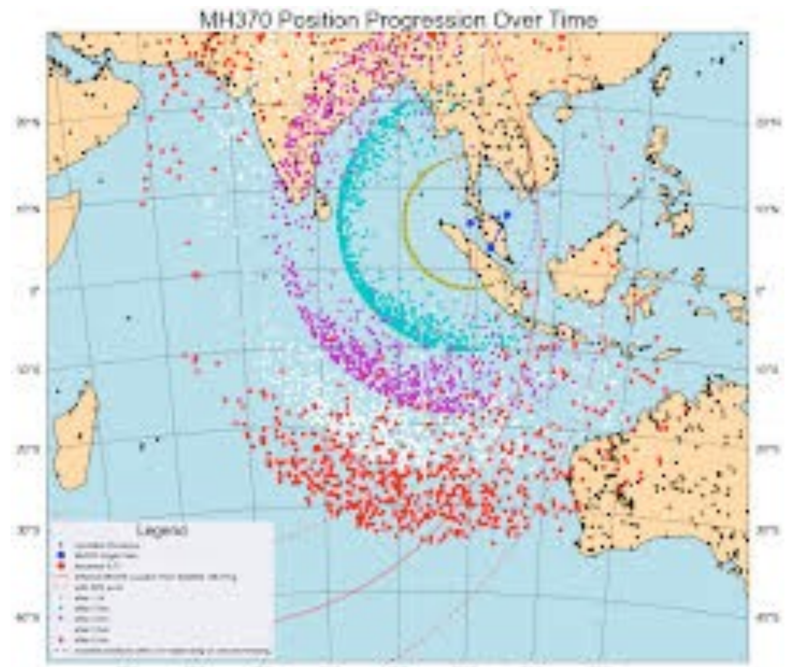
Datos

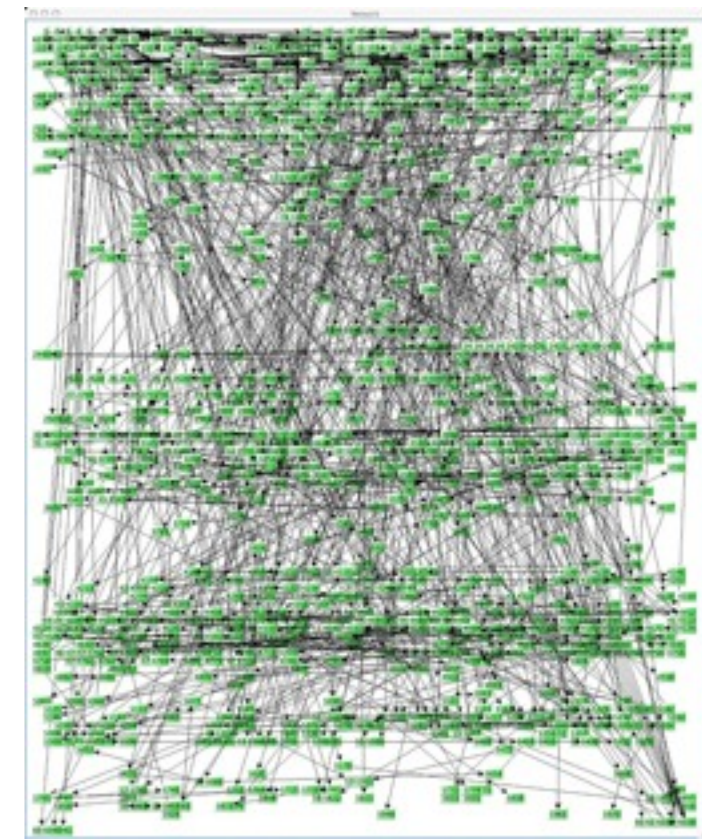


Inferencia Bayesiana

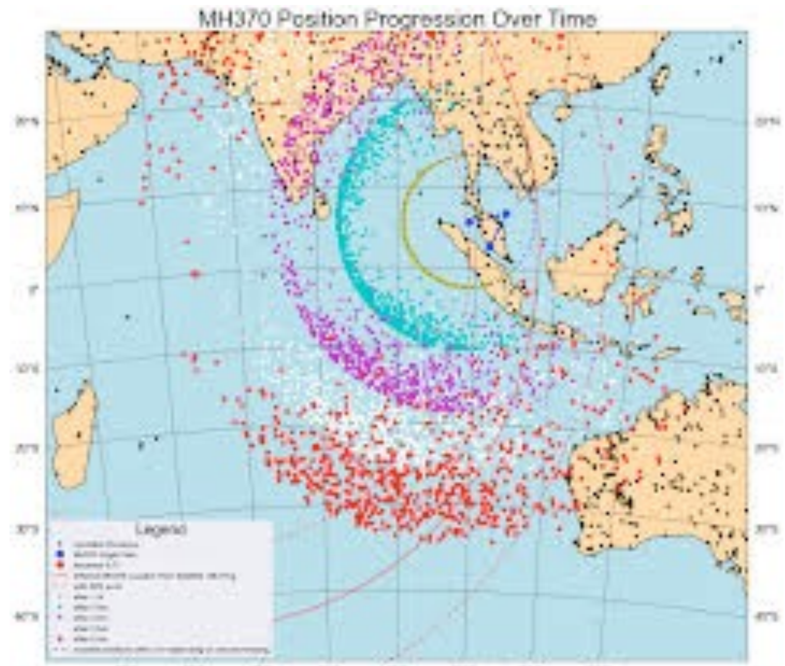
- Incorpora conocimiento *a priori* en forma natural (¡y obligatoria!)
- Elude los *p values*
- Responde en términos de distribuciones de probabilidad (nuestro *grado de creencia)*
- Datos secuenciales: modelo de aprendizaje
- Redes, modelos jerárquicos

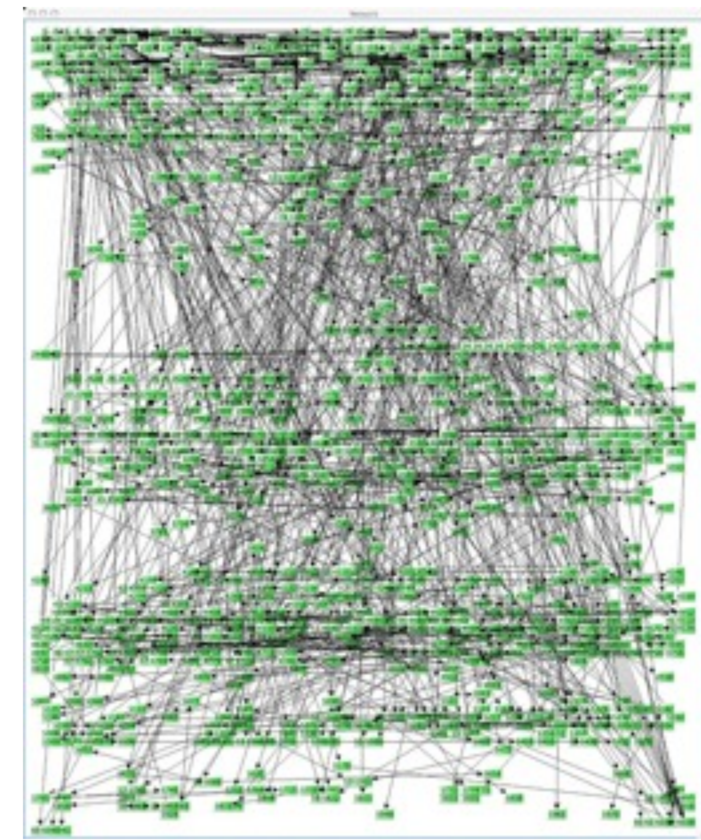
¿Por qué ahora?
(de Finetti 1930, *Probabilismo)*





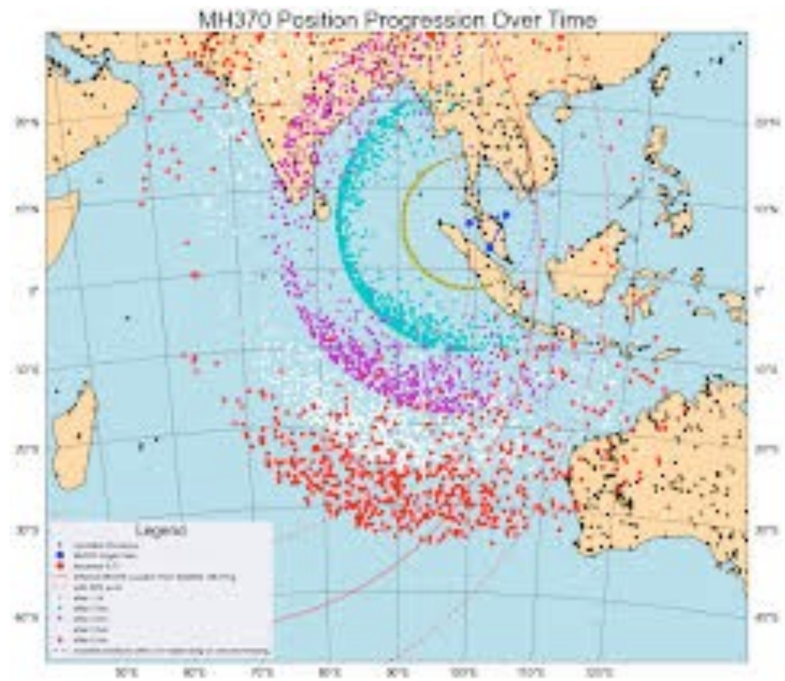
¿Por qué ahora?
(de Finetti 1930, *Probabilismo)*





computadoras + algoritmos

¿Por qué ahora?
(de Finetti 1930, *Probabilismo*)



computadoras + algoritmos

↓

uso práctico

inferencia bayesiana vs. estadística frecuentista

abordaje moderno: convivencia

inferencia bayesiana vs. estadística frecuentista

abordaje moderno: convivencia

- proyectos de largo plazo, con un especialista en el campo
- fuerte en modelado
- ¿más ‘honesta’?

inferencia bayesiana vs. estadística frecuentista

abordaje moderno: convivencia

- proyectos de largo plazo, con un especialista en el campo
- fuerte en modelado
- ¿más ‘honesta’?

- *software bundles*
- uso repetido
- modelado mínimo

¿y por qué probabilidades? ¿por qué Bayes*?*

dutch book arguments -coherencia

¿y por qué probabilidades? ¿por qué Bayes*?*

dutch book arguments -coherencia

más en general:

estadística en términos de toma de decisiones

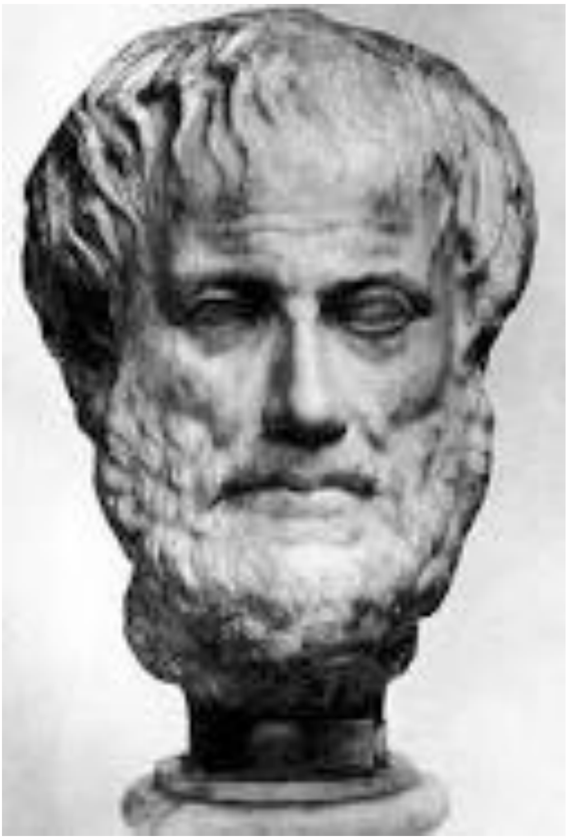
minimizar la péridida, maximizar la utilidad

¿y por qué probabilidades? ¿por qué Bayes*?*
dutch book arguments -coherencia

más en general:
estadística en términos de toma de decisiones
minimizar la péridida, maximizar la utilidad

racionalidad en contextos de incertidumbre...
¡modelo del pensamiento humano!

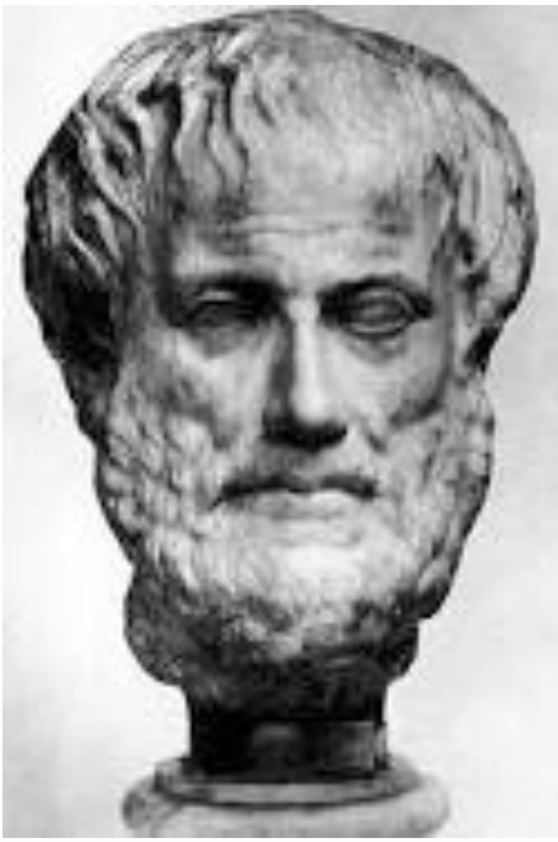
Desde Aristóteles…



Lógica

Todo **hombre** es **mortal**
Sócrates es **hombre**
Ergo **Sócrates** es **mortal**

Desde Aristóteles…

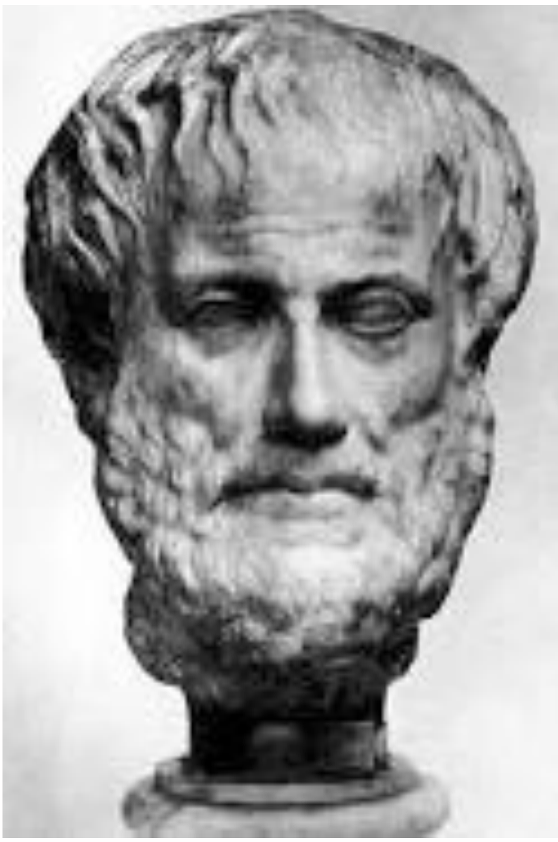


Lógica

Todo **hombre** es **mortal**
Sócrates es **hombre**
Ergo **Sócrates** es **mortal**

Planteada como modelo del pensamiento

Desde Aristóteles…



Lógica

Todo **hombre** es **mortal**
Sócrates es **hombre**
Ergo **Sócrates** es **mortal**

Planteada como modelo del pensamiento

pero…

La Paradoja de Linda

La Paradoja de Linda

$$p(cajera) \qquad\qquad p(cajera \,\&\, feminista)$$

La Paradoja de Linda

$$p(cajera) \quad < \quad p(cajera \,\&\, feminista)$$

Sesgos y Heurísticas, hombre **irracional**
(Tversky & Kahneman, 1980s)

La Paradoja de Linda

$$p(cajera) \quad > \quad p(cajera \,\&\, feminista)$$

(racional)

Sesgos y Heurísticas, hombre **irracional**
(Tversky & Kahneman, 1980s)

Programa de la Cognición Bayesiana

Programa de la Cognición Bayesiana

- Lógica es adecuada en contextos de **certidumbre**

Programa de la Cognición Bayesiana

- Lógica es adecuada en contextos de **certidumbre**
- Cuando hay **incertidumbre**, el lenguaje racional es la teoría de probabilidad

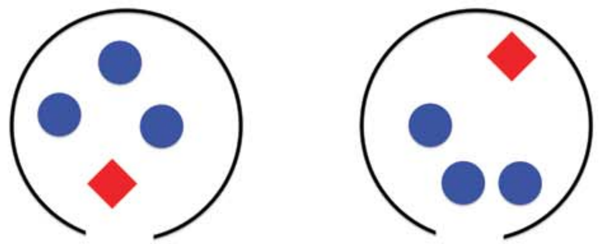
Programa de la Cognición Bayesiana

- Lógica es adecuada en contextos de **certidumbre**
- Cuando hay **incertidumbre**, el lenguaje racional es la teoría de probabilidad
- **Racionalidad acotada** por nuestros recursos de cómputo: ilusión de irracionalidad

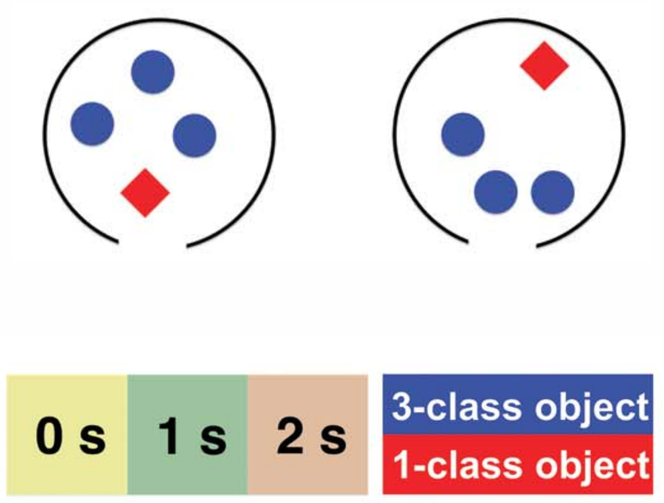
Programa de la Cognición Bayesiana

- Lógica es adecuada en contextos de **certidumbre**
- Cuando hay **incertidumbre**, el lenguaje racional es la teoría de probabilidad
- **Racionalidad acotada** por nuestros recursos de cómputo: ilusión de irracionalidad
- Programa **general** y **cuantitativo**

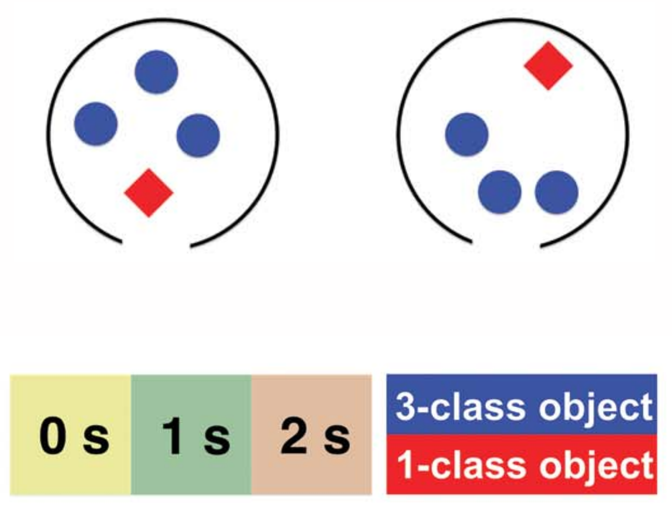
Inferencia Causal

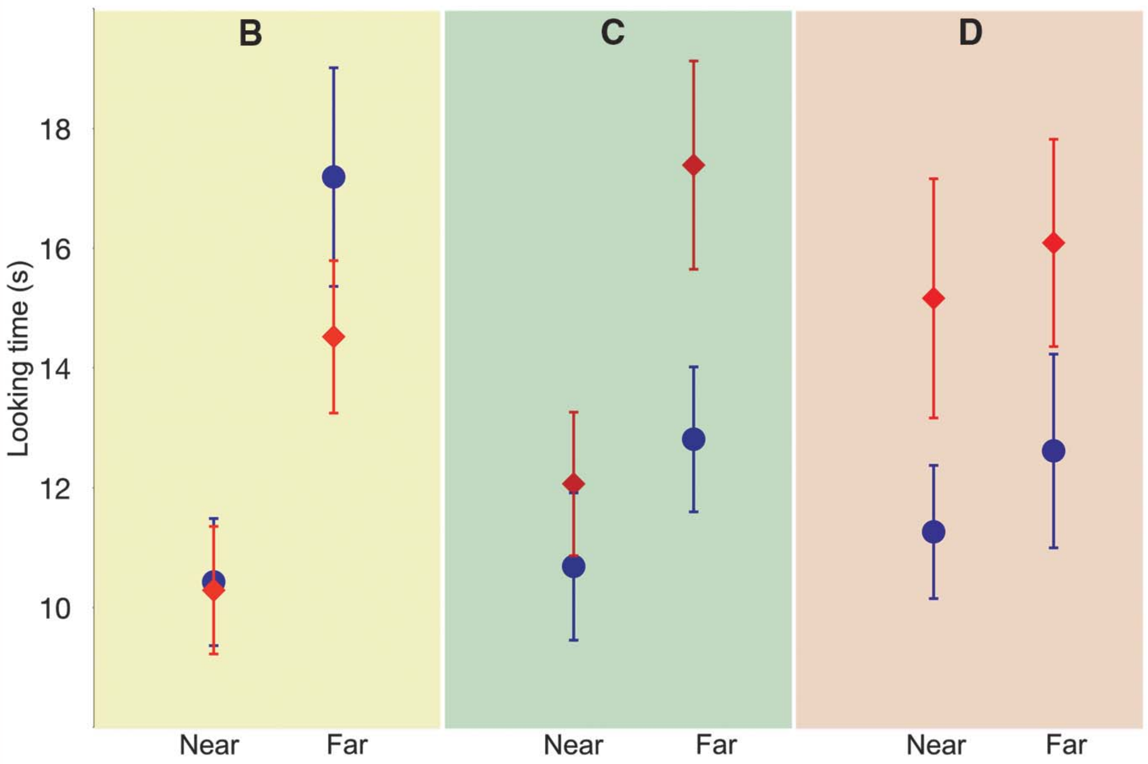


Inferencia Causal

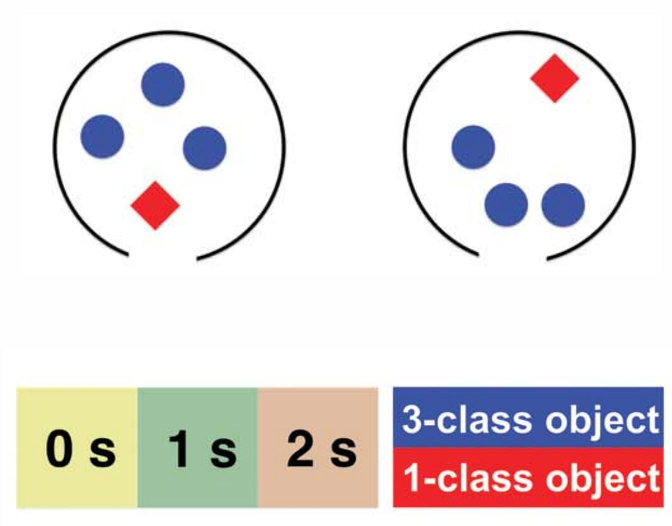


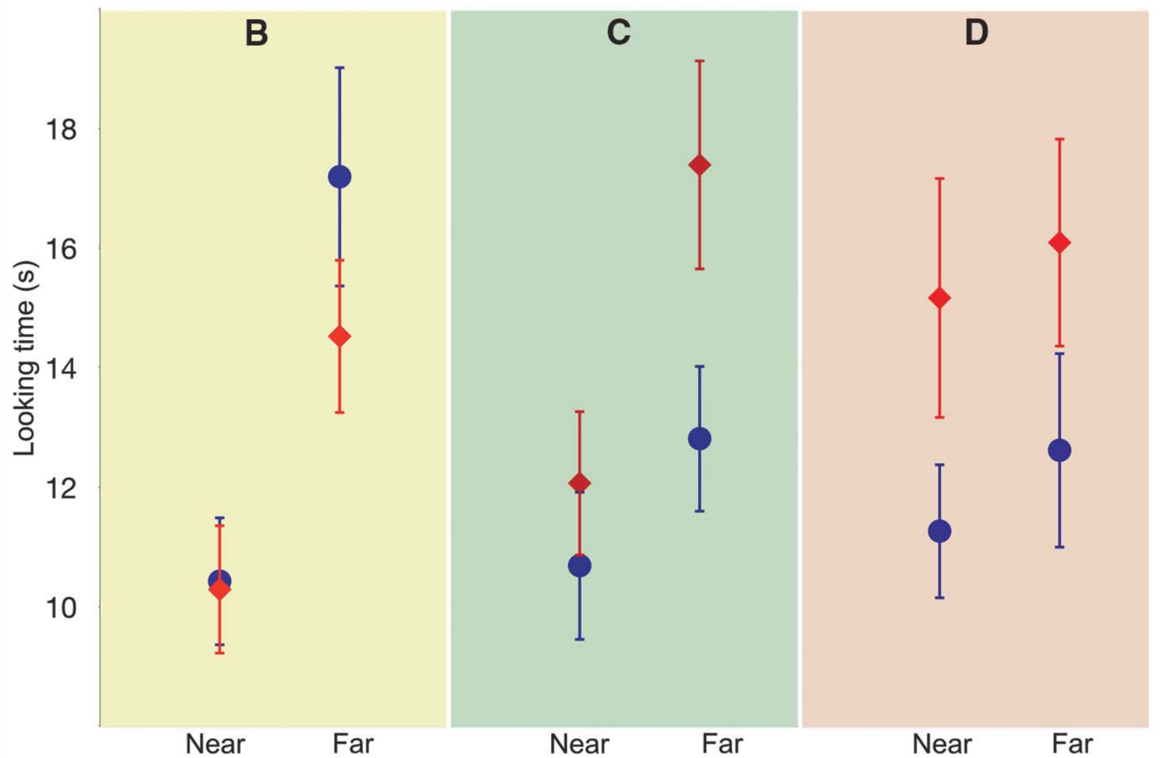
Inferencia Causal

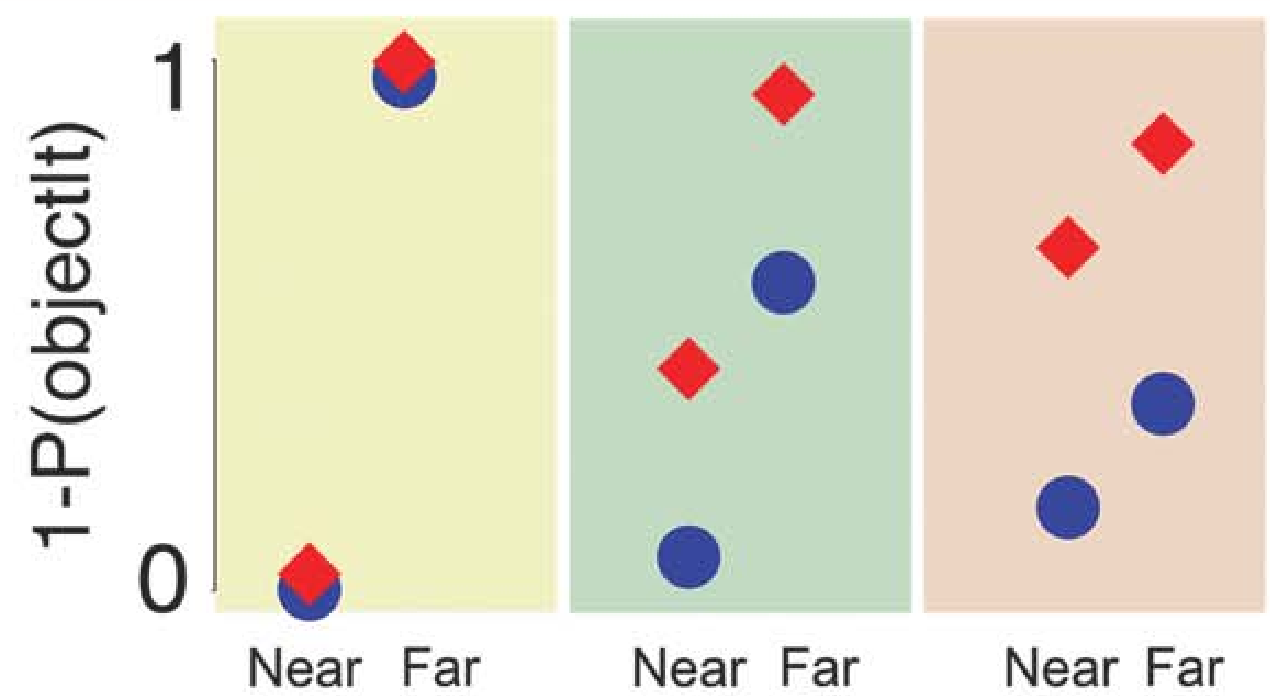




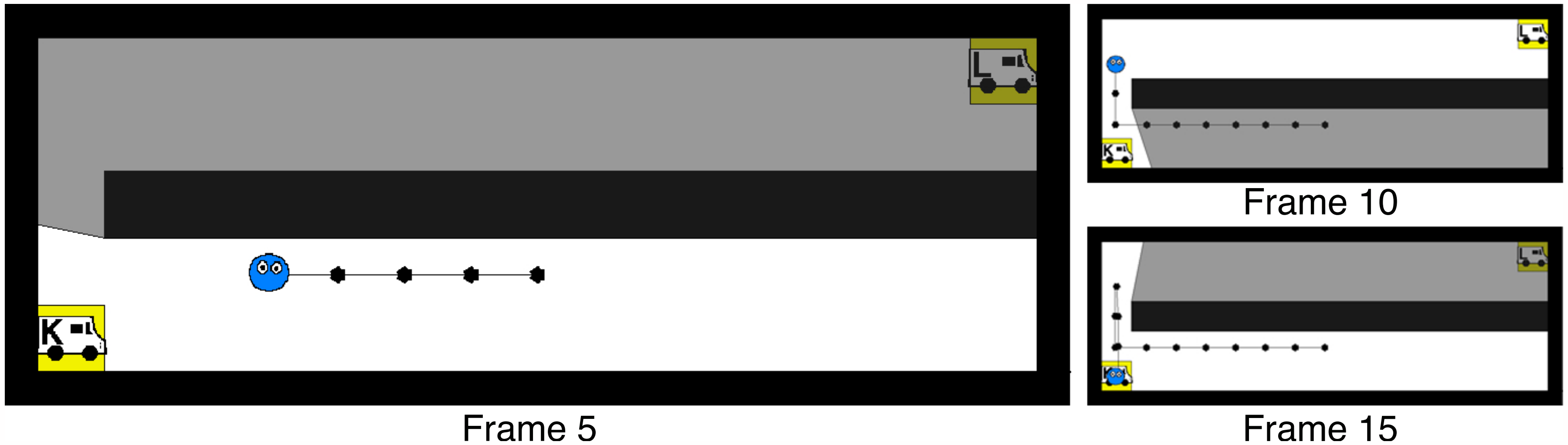
Inferencia Causal







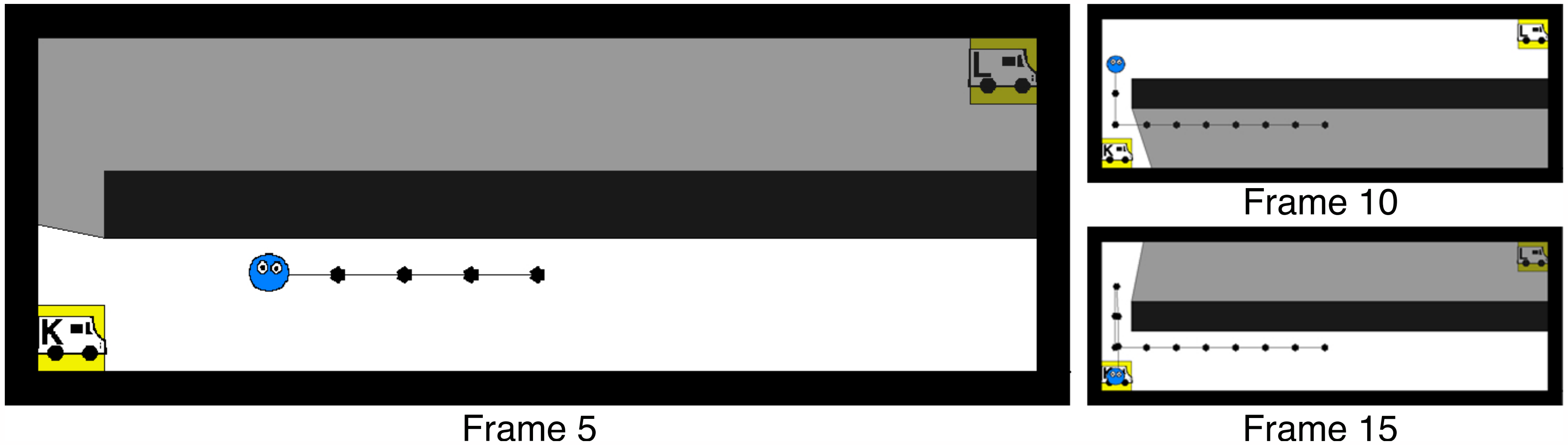
Teoria de la Mente (joint Belief-Desire)

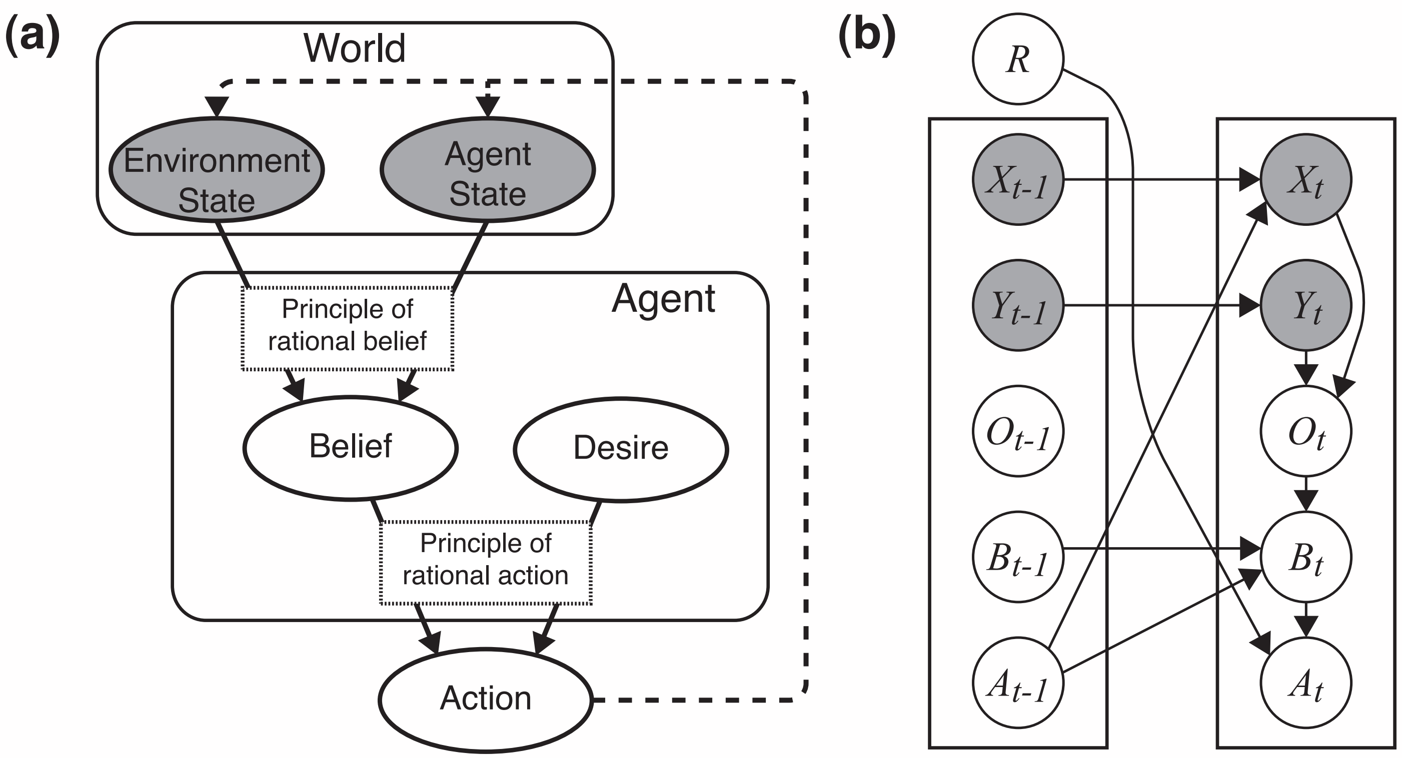


Desires

Beliefs

Teoria de la Mente (joint Belief-Desire)

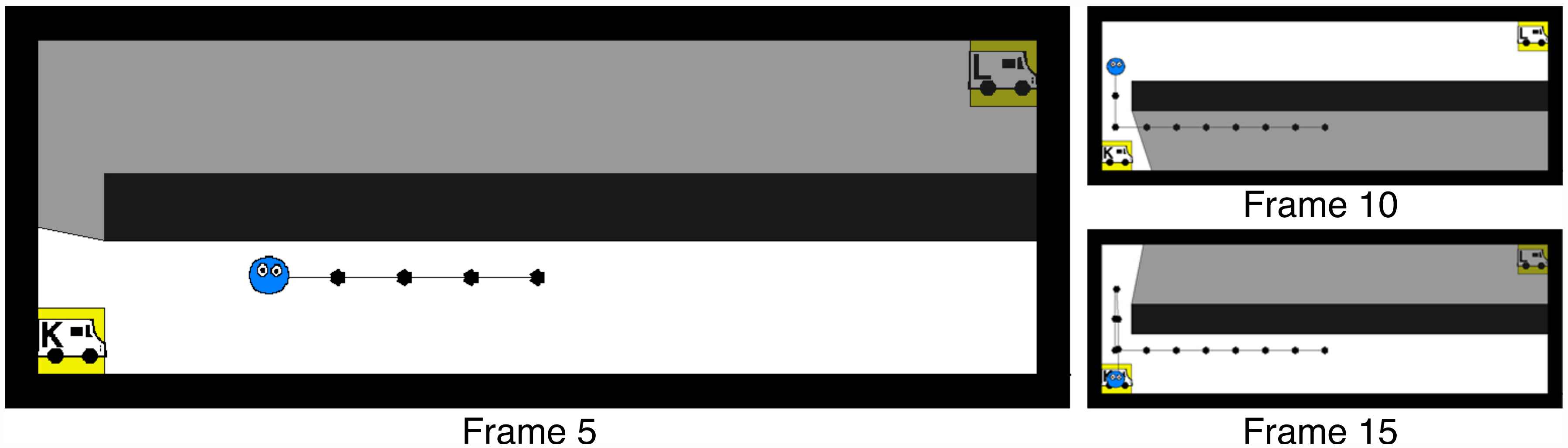


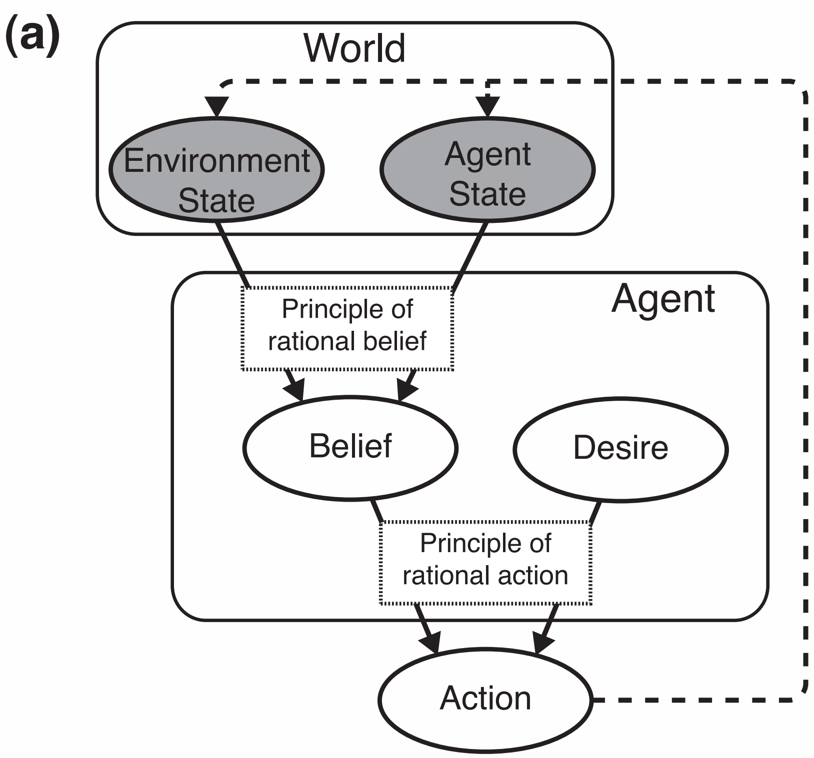


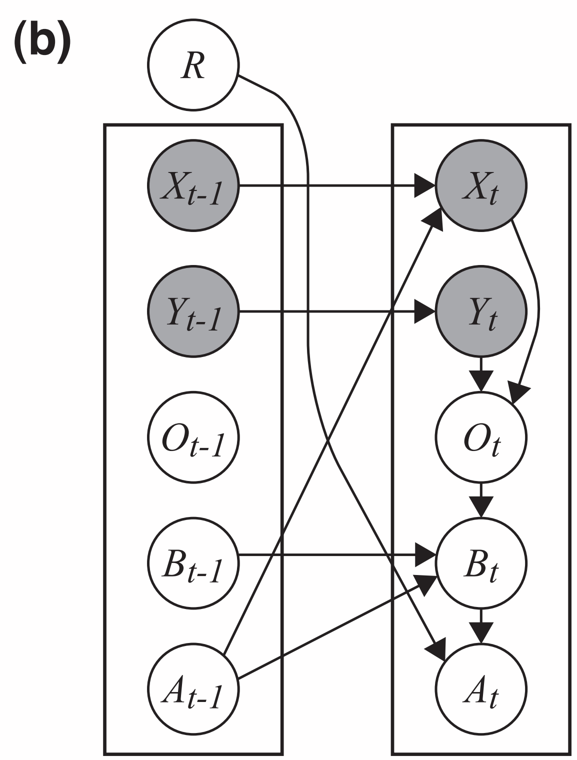
Desires

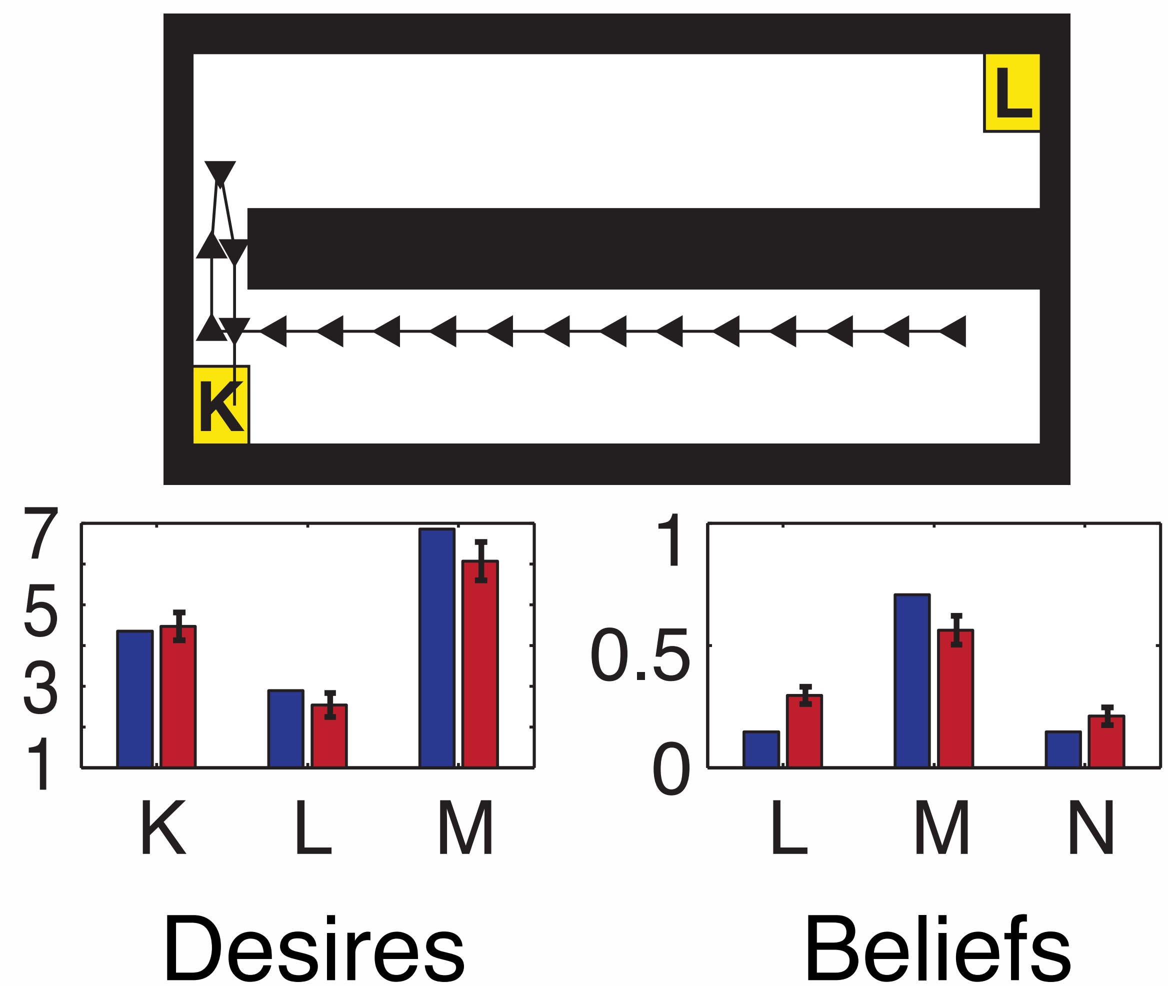
Beliefs

Teoria de la Mente (joint Belief-Desire)









False Belief

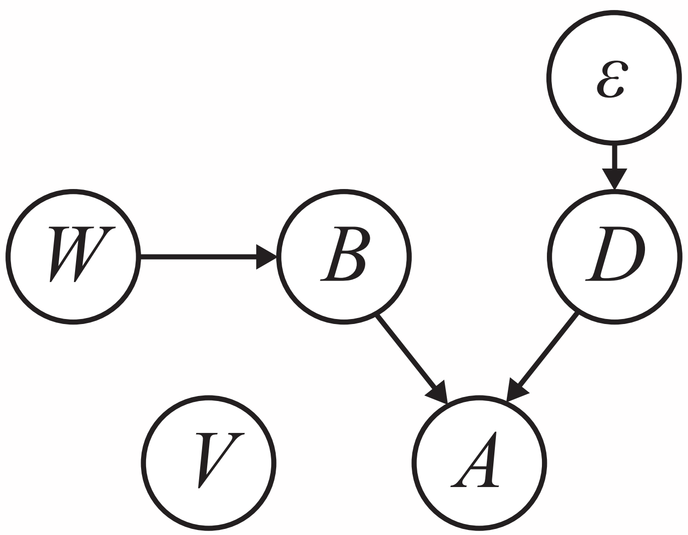
False Belief

Variable	Description	States
World (W)	Location of the toy.	0: Original location, 1: New location.
Access (V)	Could Sally see the toy moved?	0: No, 1: Yes.
Action (A)	Where Sally looks for her toy.	0: Original location, 1: New location.
Belief (B)	Where Sally thinks the toy is.	0: Original location, 1: New location.
Desire (D)	Sally's primary desire.	1: To find the toy, 0: Anything else.

False Belief

Variable	Description	States
World (W)	Location of the toy.	0: Original location, 1: New location.
Access (V)	Could Sally see the toy moved?	0: No, 1: Yes.
Action (A)	Where Sally looks for her toy.	0: Original location, 1: New location.
Belief (B)	Where Sally thinks the toy is.	0: Original location, 1: New location.
Desire (D)	Sally's primary desire.	1: To find the toy, 0: Anything else.

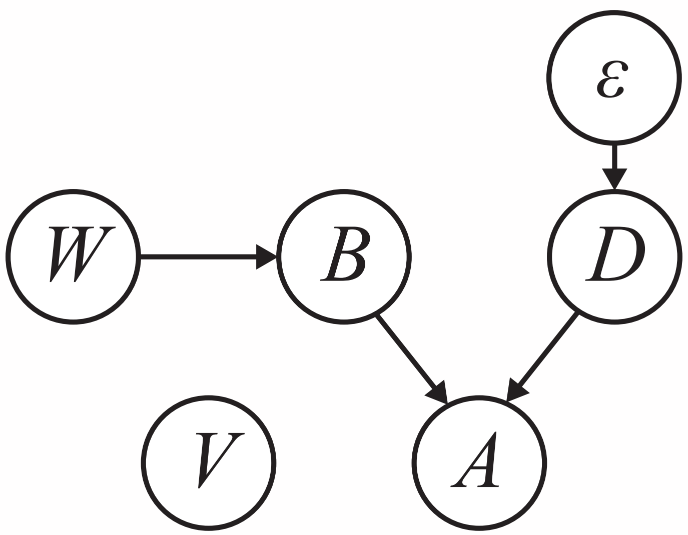
Copy Theorist



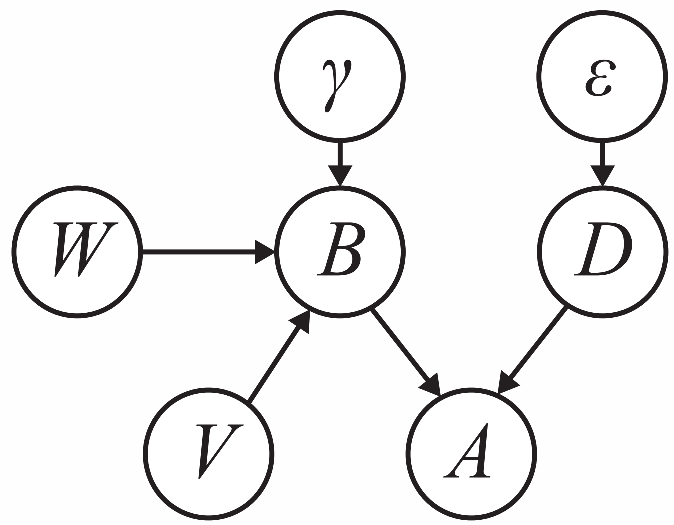
False Belief

Variable	Description	States
World (W)	Location of the toy.	0: Original location, 1: New location.
Access (V)	Could Sally see the toy moved?	0: No, 1: Yes.
Action (A)	Where Sally looks for her toy.	0: Original location, 1: New location.
Belief (B)	Where Sally thinks the toy is.	0: Original location, 1: New location.
Desire (D)	Sally's primary desire.	1: To find the toy, 0: Anything else.

Copy Theorist



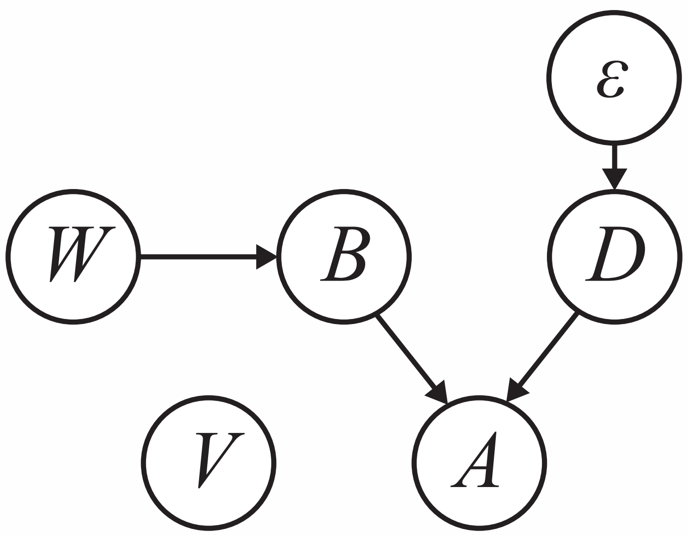
Perspective Theorist



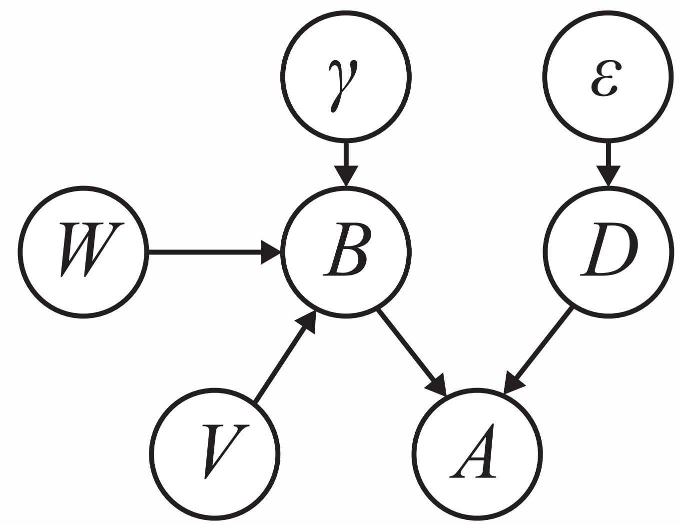
False Belief

Variable	Description	States
World (W)	Location of the toy.	0: Original location, 1: New location.
Access (V)	Could Sally see the toy moved?	0: No, 1: Yes.
Action (A)	Where Sally looks for her toy.	0: Original location, 1: New location.
Belief (B)	Where Sally thinks the toy is.	0: Original location, 1: New location.
Desire (D)	Sally's primary desire.	1: To find the toy, 0: Anything else.

Copy Theorist



Perspective Theorist

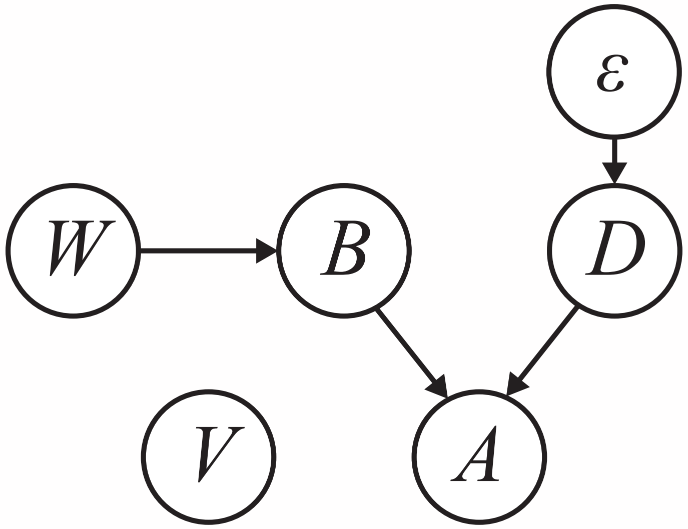


$P(A=1\mid B,D)$	B	D
0	0	1
1	1	1
0.5	0	0
0.5	1	0

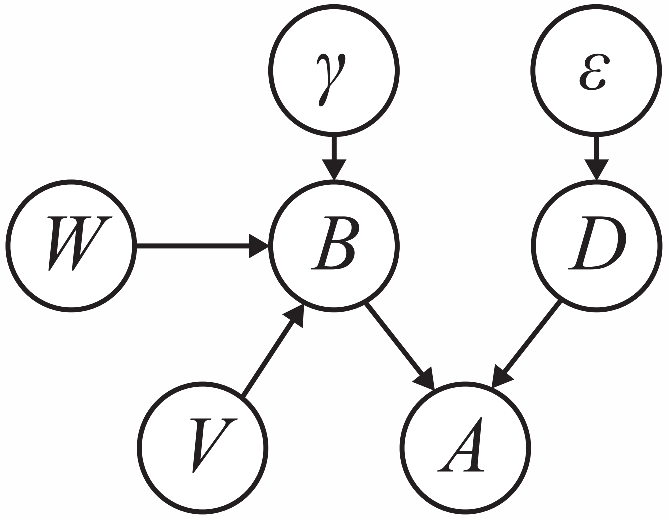
False Belief

Variable	Description	States
World (W)	Location of the toy.	0: Original location, 1: New location.
Access (V)	Could Sally see the toy moved?	0: No, 1: Yes.
Action (A)	Where Sally looks for her toy.	0: Original location, 1: New location.
Belief (B)	Where Sally thinks the toy is.	0: Original location, 1: New location.
Desire (D)	Sally's primary desire.	1: To find the toy, 0: Anything else.

Copy Theorist



Perspective Theorist



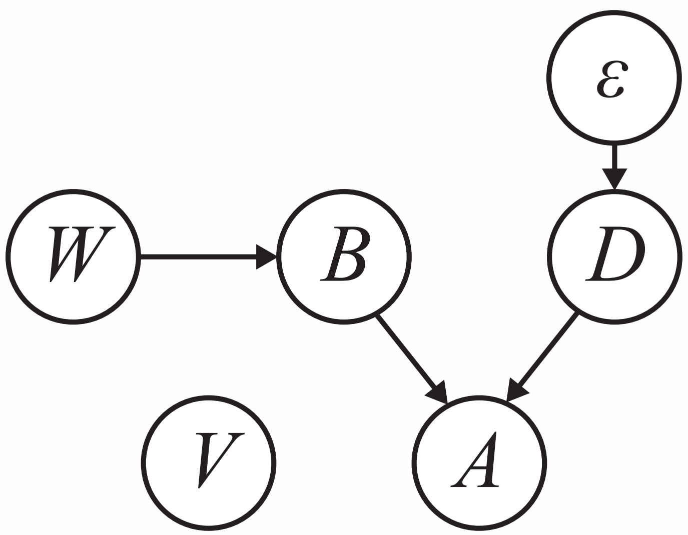
$P(A=1 \vert B,D)$	B	D
0	0	1
1	1	1
0.5	0	0
0.5	1	0

$P_{\mathrm{PT}}(B=1 \vert W,V)$	W	V
0	0	1
1	1	1
γ	0	0
γ	1	0

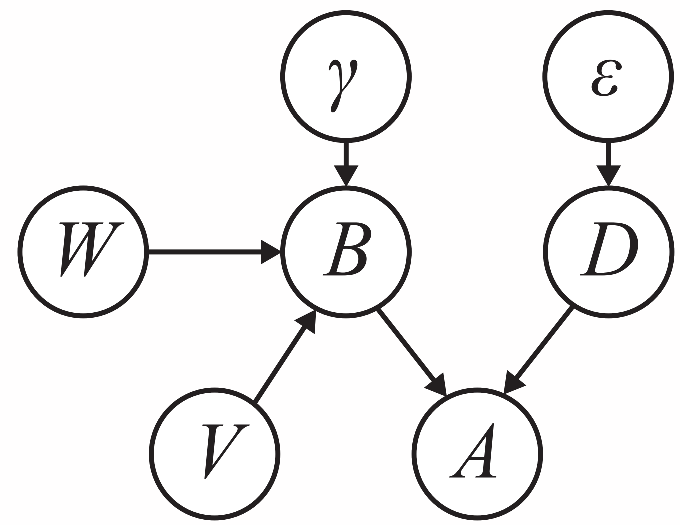
False Belief

Variable	Description	States
World (W)	Location of the toy.	0: Original location, 1: New location.
Access (V)	Could Sally see the toy moved?	0: No, 1: Yes.
Action (A)	Where Sally looks for her toy.	0: Original location, 1: New location.
Belief (B)	Where Sally thinks the toy is.	0: Original location, 1: New location.
Desire (D)	Sally's primary desire.	1: To find the toy, 0: Anything else.

Copy Theorist

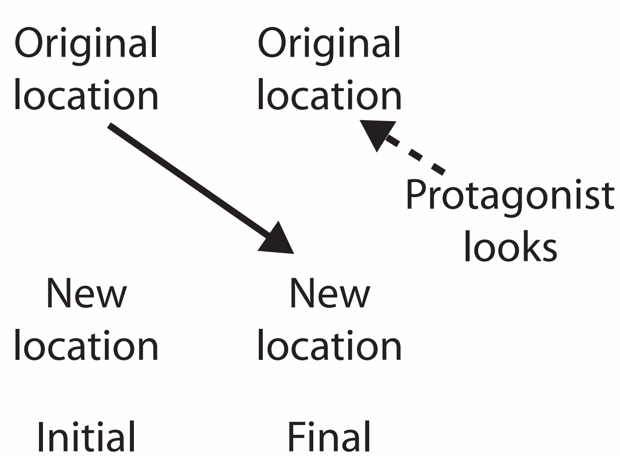


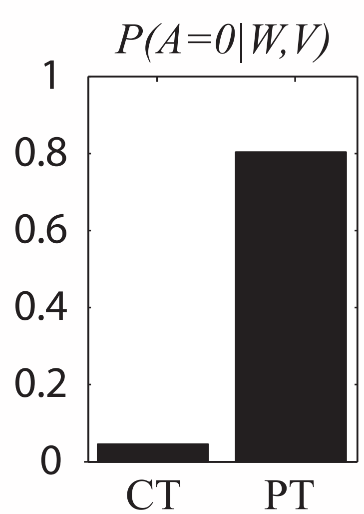
Perspective Theorist



$P(A=1\|B,D)$	B	D
0	0	1
1	1	1
0.5	0	0
0.5	1	0

$P_{\mathrm{PT}}(B=1\|W,V)$	W	V
0	0	1
1	1	1
γ	0	0
γ	1	0

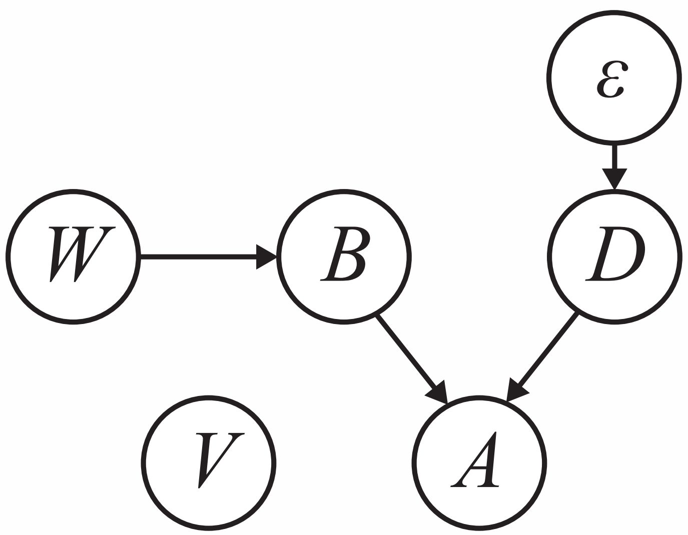




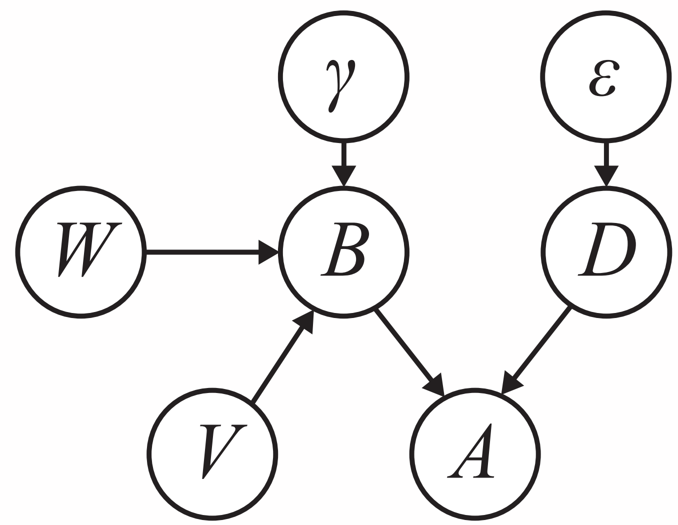
False Belief

Variable	Description	States
World (W)	Location of the toy.	0: Original location, 1: New location.
Access (V)	Could Sally see the toy moved?	0: No, 1: Yes.
Action (A)	Where Sally looks for her toy.	0: Original location, 1: New location.
Belief (B)	Where Sally thinks the toy is.	0: Original location, 1: New location.
Desire (D)	Sally's primary desire.	1: To find the toy, 0: Anything else.

Copy Theorist

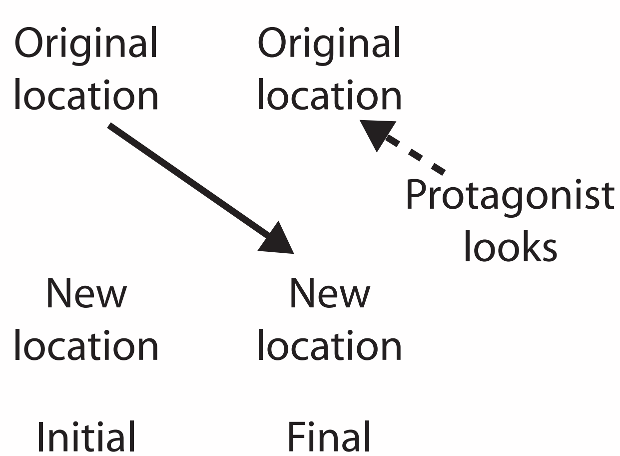


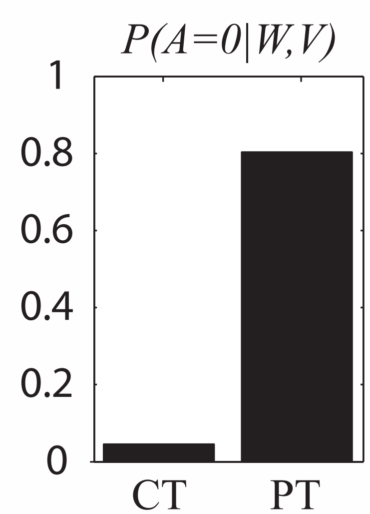
Perspective Theorist

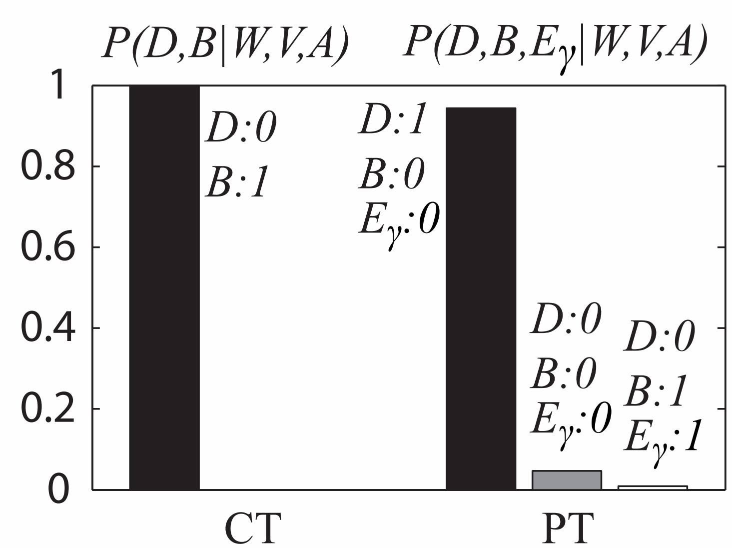


$P(A=1 \mid B,D)$	B	D
0	0	1
1	1	1
0.5	0	0
0.5	1	0

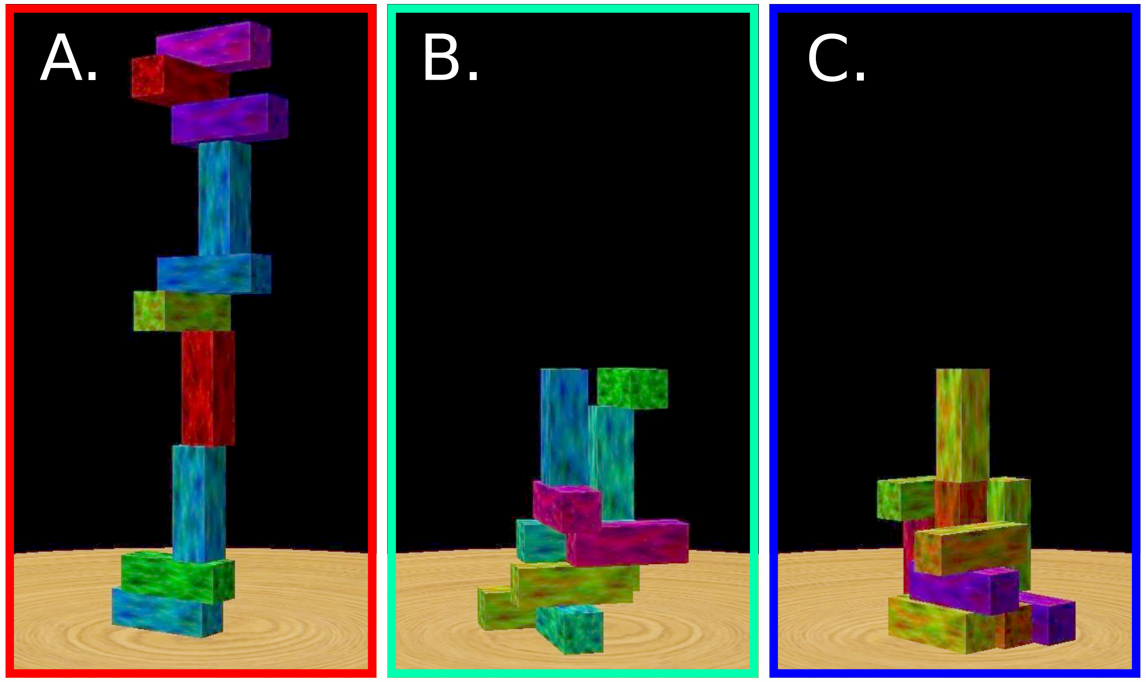
$P_{\mathrm{PT}}(B=1 \mid W,V)$	W	V
0	0	1
1	1	1
γ	0	0
γ	1	0







Física Intuitiva



Física Intuitiva