The Logical Essentials of Bayesian Reasoning

Bart Jacobs
Radboud University Nijmegen

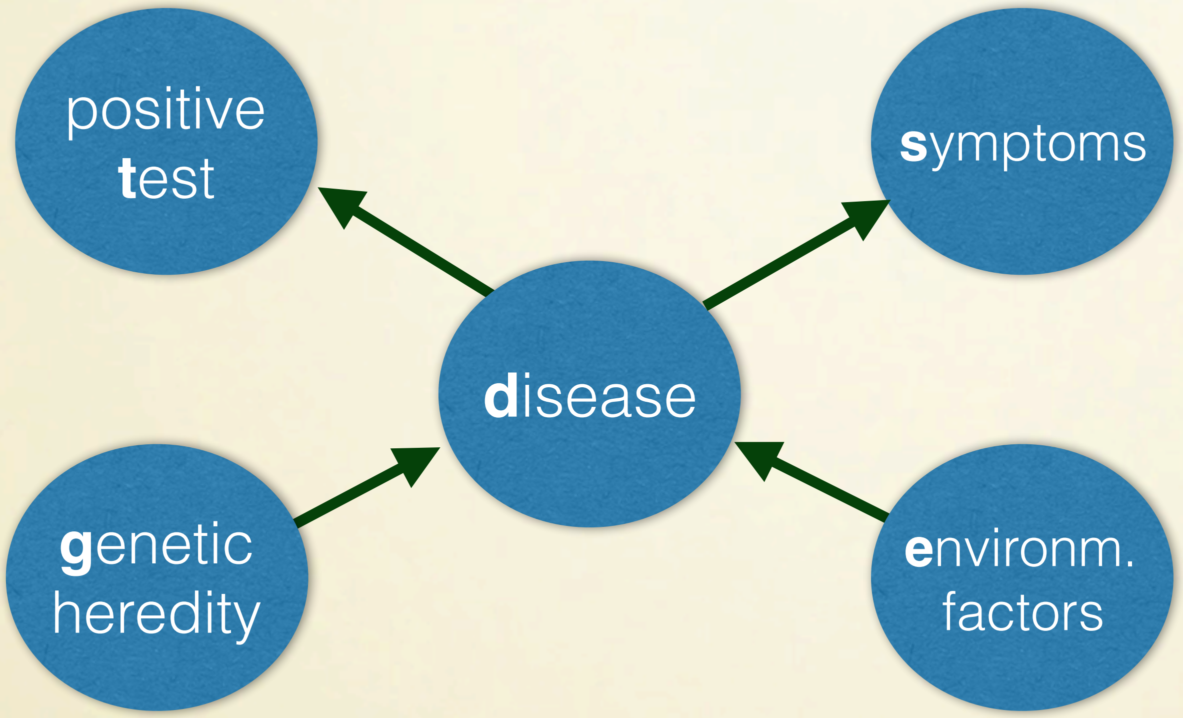
Fabio Zanasi
University College London

arxiv.org/abs/1804.01193

In a nutshell

- We develop a categorical approach to Bayesian probability theory.
- Our methodology is driven by programming language semantics.
- It offers a principled, compositional way of performing the fundamental Bayesian reasoning tasks, such as inference and learning.

Bayesian Networks



g
1/50

e
1/10

	d
g e	9/10
g e$^{\perp}$	8/10
g$^{\perp}$ e	4/10
g$^{\perp}$ e$^{\perp}$	0

	t
d	9/10
d$^{\perp}$	1/20

	s
d	9/10
d$^{\perp}$	1/15

Inference questions

P(t)

What is the a priori probability of a positive test?

P(t | s)

*What is the probability of a positive test **given** the symptoms?*

Toolbox

State	$\omega \in \mathcal{D}(X)$	State of affairs	$1 \rightsquigarrow X$
Predicate	$p : X \to [0,1]$	Observation, (fuzzy) event	$X \rightsquigarrow 2$
Conditioning	$\omega\|_p \in \mathcal{D}(X)$	Revision due to an observation	$1 \rightsquigarrow X$
Channel	$f : X \to \mathcal{D}(Y)$	Change of base/ message passing	$X \rightsquigarrow Y$

State trans

Predicate tra

be as arrows of $\mathrm{kl}(\boldsymbol{\mathcal{D}})$

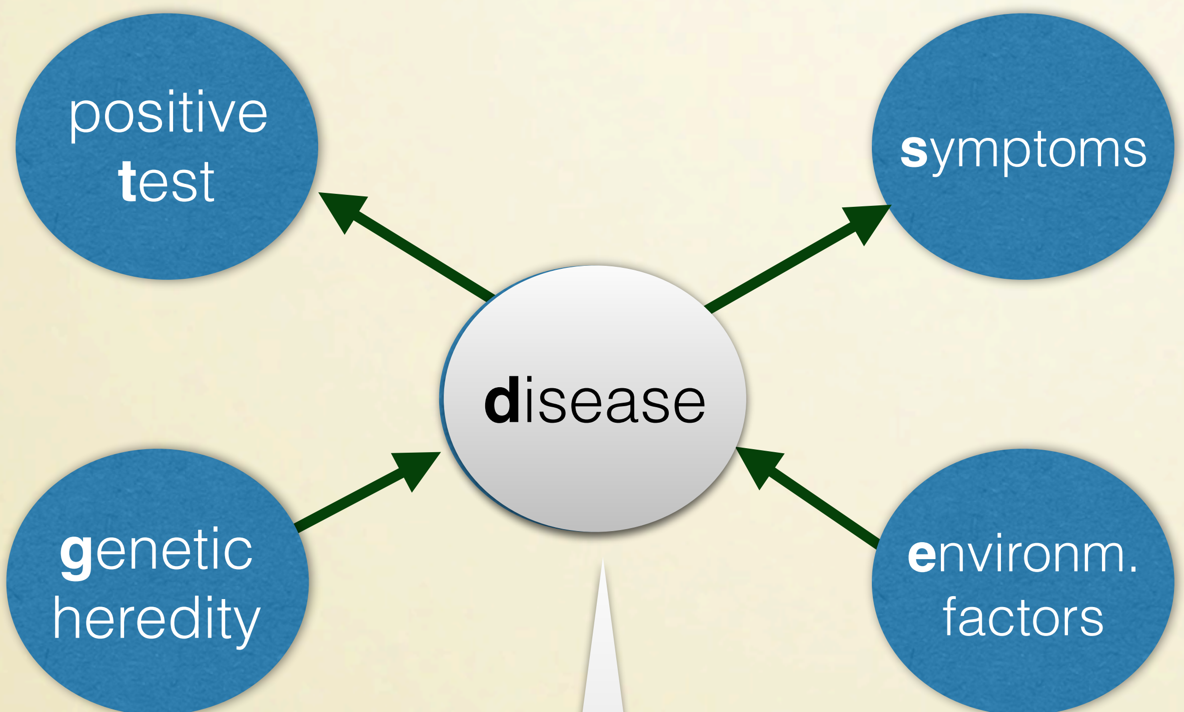
These notions make sense in other categories as well.

Continuous case: $\mathrm{kl}(\boldsymbol{\mathcal{G}})$
states are probability measures, predicates are measurable functions to [0,1].

Quantum case: $\mathbb{vNA}^{op}$
states are…quantum states, predicates are effects.

Bayesian Networks in Kleisli

We interpret a Bayesian network as an arrow of kl($\mathcal{D}$).



g
1/50

e
1/10

	d
g e	9/10
g e$^\perp$	8/10
g$^\perp$ e	4/10
g$^\perp$ e$^\perp$	0

	t
d	9/10
d$^\perp$	1/20

	s
d	9/10
d$^\perp$	1/15

Kl($\mathcal{D}$)

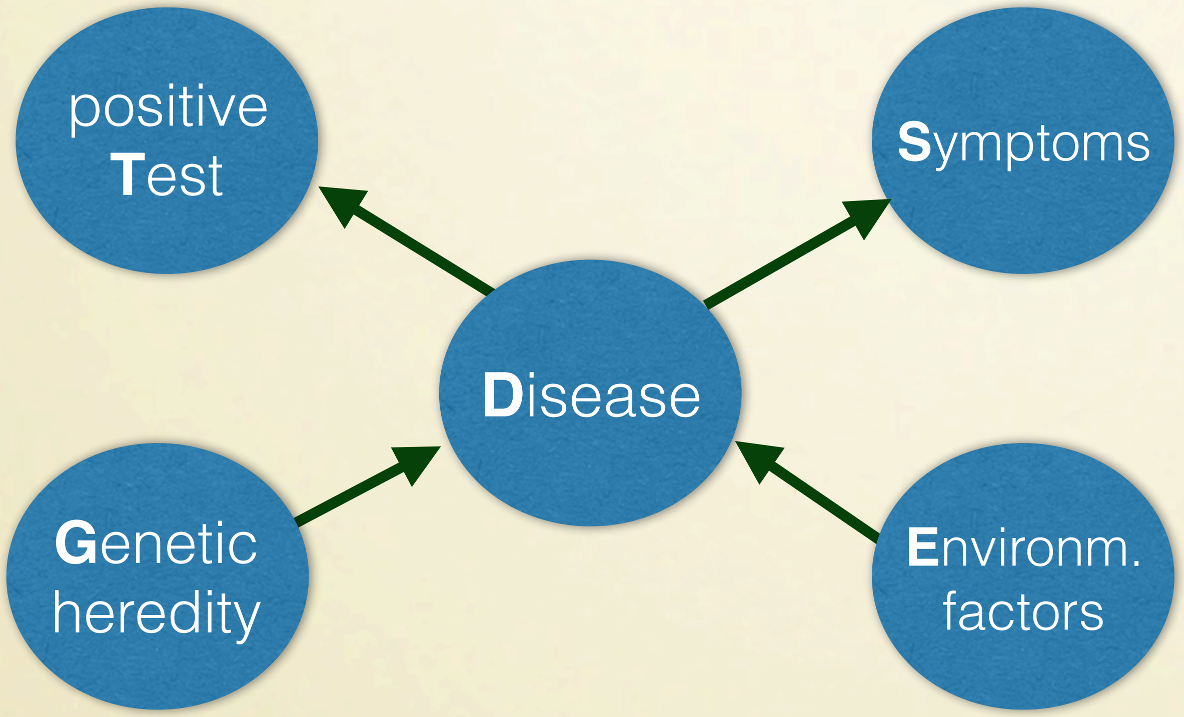
$$\{g, g^\perp\} \times \{e, e^\perp\} \overset{D}{\to} \{d, d^\perp\}$$

$$2_g \otimes 2_e \to 2_d$$

$$(g, e) \mapsto 9/10|d\rangle + 1/10|d^\perp\rangle$$
$$(g, e^\perp) \mapsto 8/10|d\rangle + 2/10|d^\perp\rangle$$
$$(g^\perp, e) \mapsto 4/10|d\rangle + 6/10|d^\perp\rangle$$
$$(g^\perp, e^\perp) \mapsto 1|d^\perp\rangle$$

Bayesian Networks in Kleisli

We interpret a Bayesian network as an arrow of **kl**($\boldsymbol{\mathcal{D}}$).



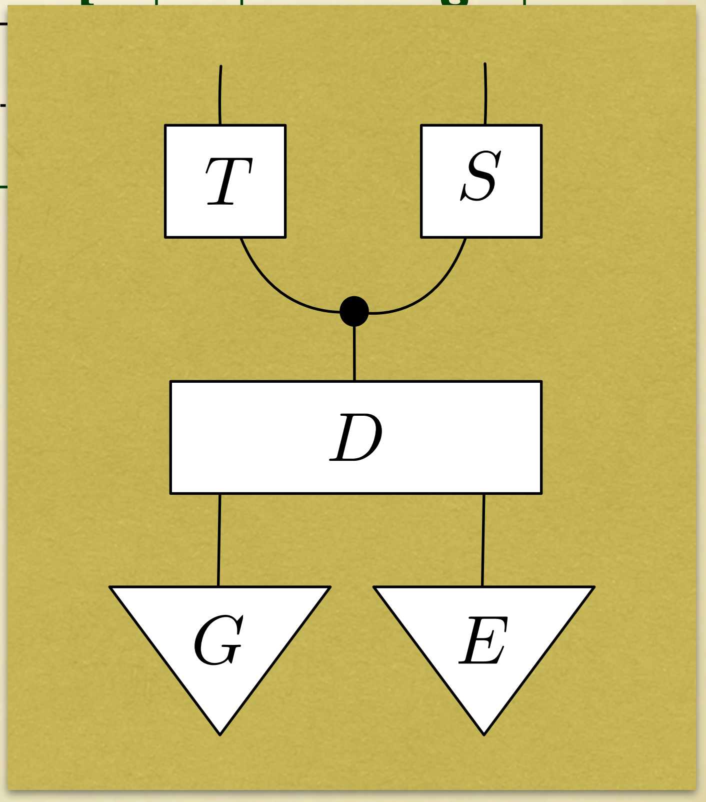
g
1/50

e
1/10

	d
g e	9/10
g e$^\perp$	8/10
g$^\perp$ e	4/10
g$^\perp$ e$^\perp$	0

	t
d	
d$^\perp$	

	s



Kl($\mathcal{D}$)

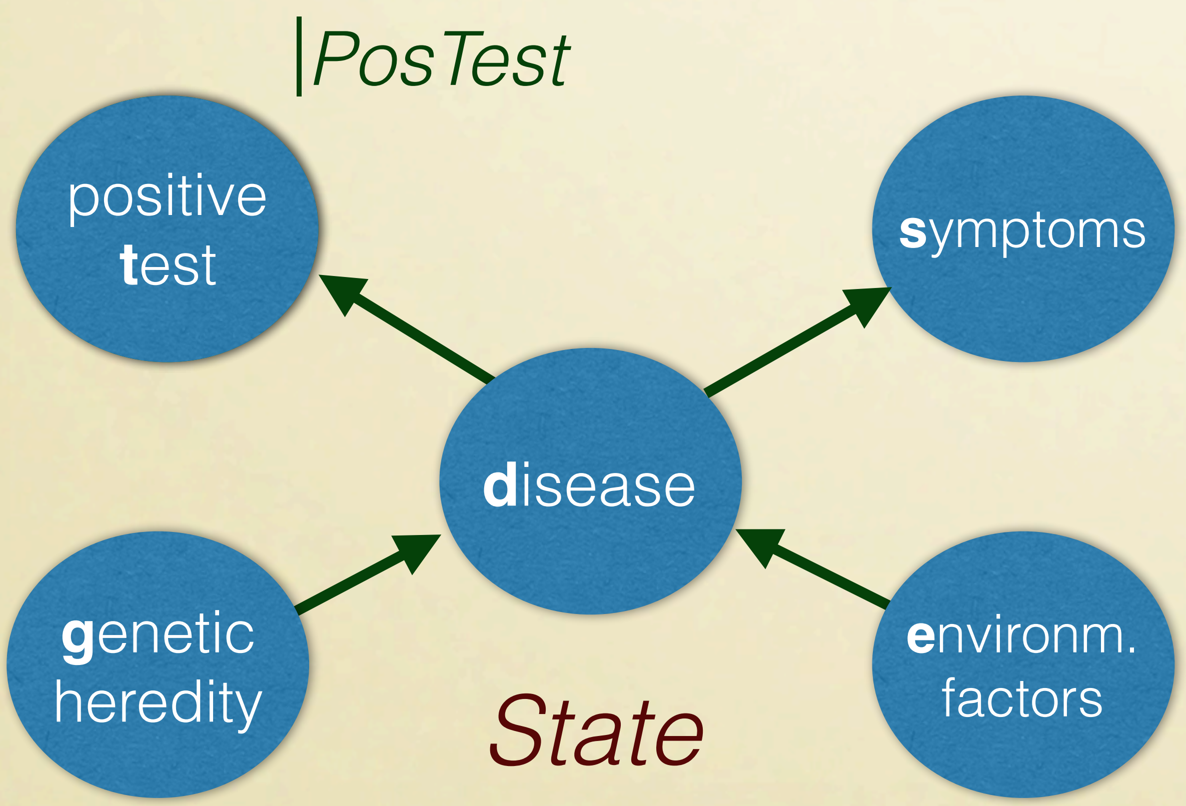
$$1 \xrightarrow{G\otimes E} 2_G \otimes 2_E \xrightarrow{D} 2_D \xrightarrow{\Delta} 2_D \otimes 2_D \xrightarrow{T\otimes S} 2_T \otimes 2_S$$

Bayesian Inference in Kleisli

Both Bayesian networks and our toolbox live in **kl**($\mathcal{D}$).

We shall now compute the two inference questions in **kl**($\mathcal{D}$).

The calculation will have a `dynamical' flavour:



Inference questions

What is the a priori probability of a positive test?

What is the probability of a positive test ***given*** *the symptoms?*

Bayesian Inference in Kleisli

Inference question I *What is the a priori probability of a positive test?*

$$\begin{tikzcd}
1 \arrow[r, "G\otimes E"] \arrow[r, bend right] \arrow[rrr, bend right, "T\gg D\gg(G\otimes E)"'] & 2_G \otimes 2_E \arrow[r, "D"] & 2_D \arrow[r, "T"] & 2_T \arrow[d, "PosTest"] \\
 & & & 2
\end{tikzcd}$$

$t \longmapsto 1$

$t^{\perp} \longmapsto 0$

1. Consider state $G\otimes E \in \mathcal{D}(2_G \otimes 2_E)$

2. Use channel T and D as state transformers

$$T\gg D\gg(G\otimes E) \in \mathcal{D}(2_T)$$

3. Consider the predicate *PosTest*: $2_T \rightarrow [0,1]$

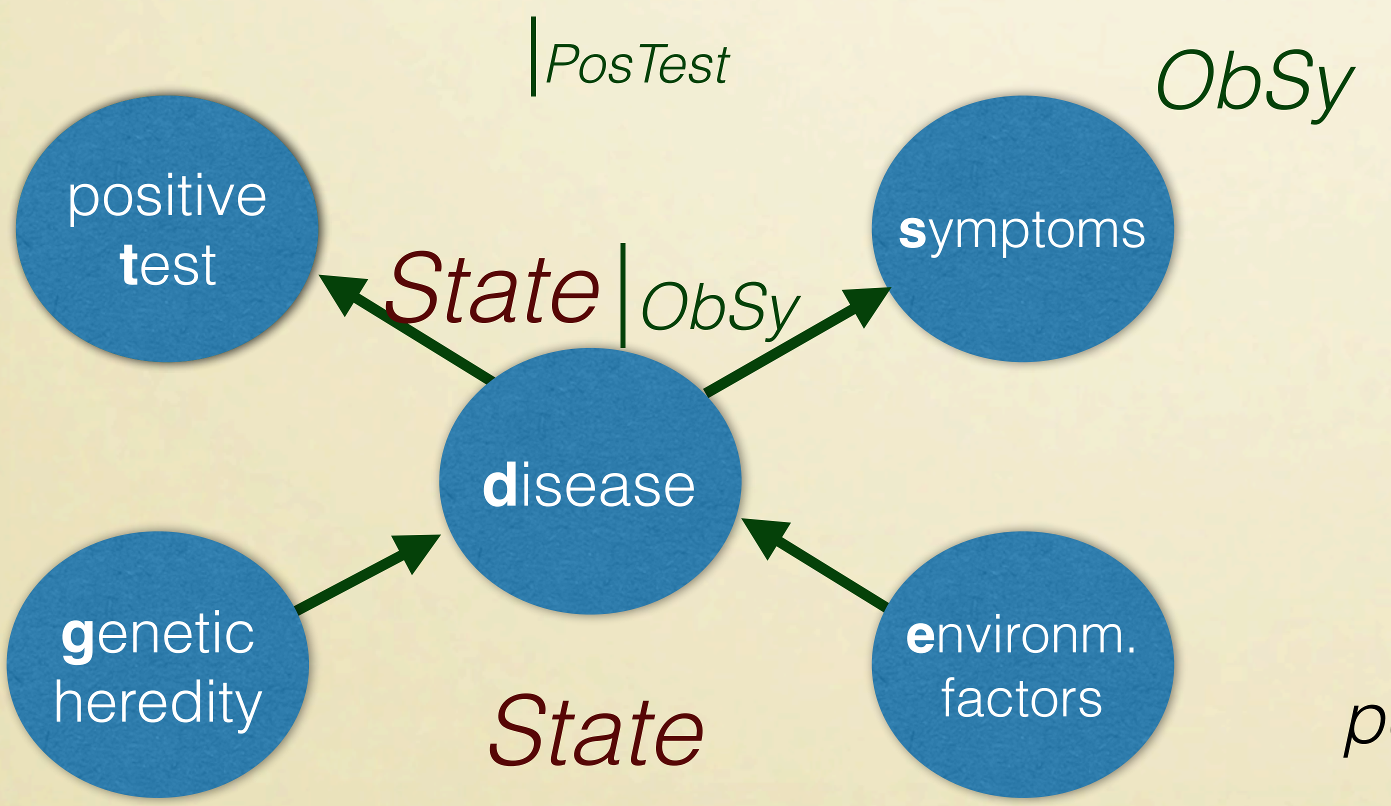
4. Answer is the conditioned state $(T\gg D\gg(G\otimes E))_{PosTest} \in \mathcal{D}(2_T)$

Bayesian Inference in Kleisli

Both Bayesian networks and our toolbox live in $\mathbf{kl}(\mathcal{D})$.

We shall now compute the two inference questions in $\mathbf{kl}(\mathcal{D})$.

The calculation will have a `dynamical' flavour:



Inference questions

What is the a priori probability of a positive test?

*What is the probability of a positive test **given** the symptoms?*

Bayesian Inference in Kleisli

Inference question II *What is the probability of a positive test **given** the symptoms?*

$$\begin{tikzcd}[column sep=large]
1 \arrow[r, "G\otimes E"] \arrow[rrr, bend left, "T\gg(D\gg(G\otimes E))_{S\ll ObSy}"] \arrow[rr, bend right, "(D\gg(G\otimes E))_{S\ll ObSy}"'] & 2_G \otimes 2_E \arrow[r, "D"] & 2_D \arrow[r, "T"] \arrow[d, "S"] \arrow[dr, bend left, "S\ll ObSy"] & 2_T \arrow[r, "PosTest"] & 2 \\
 & & 2_S \arrow[r, "ObSy"] & 2 &
\end{tikzcd}$$

$s \longmapsto 1$

$s^{\perp} \longmapsto 0$

1.Predicate *ObSy:* $2_S \to [0,1]$

2.Predicate transformation $S\ll ObSy : 2_D \to [0,1]$

3.Conditioning $(D\gg(G\otimes E))_{S\ll ObSy} \in \mathcal{D}(2_D)$

4.State transformation $T\gg(D\gg(G\otimes E))_{S\ll ObSy}) \in \mathcal{D}(2_T)$

5.Answer is the conditioned state $(T\gg(D\gg(G\otimes E))_{S\ll ObSy}))_{PosTest} \in \mathcal{D}(2_T)$

Bayesian Inference in Kleisli

Inference question II *What is the probability of a positive test* ***given*** that doctor I is **70% sure** the patient manifests the symptoms and doctor II is **50% sure** he doesn't?

$$\begin{tikzcd}
1 \arrow[r, "G\otimes E"] \arrow[rrr, bend left, "T\gg(D\gg(G\otimes E))_{S\ll ObSy}"] \arrow[rr, bend right, "(D\gg(G\otimes E))_{S\ll ObSy}"'] & 2_G \otimes 2_E \arrow[r, "D"] & 2_D \arrow[r, "T"] \arrow[d, "S"] \arrow[dr, bend left, "S\ll ObSy"] & 2_T \arrow[r] & 2 \\
 & & 2_S \arrow[r, "ObSy"] & 2 &
\end{tikzcd}$$

$s \mapsto 0.7$

$s^{\perp} \mapsto 0.5$

1. Predicate *ObSy*: $2_S \to [0,1]$

2. Predicate transformation $S\ll ObSy : 2_D \to [0,1]$

3. Conditioning $(D\gg(G\otimes E))_{S\ll ObSy} \in \mathcal{D}(2_D)$

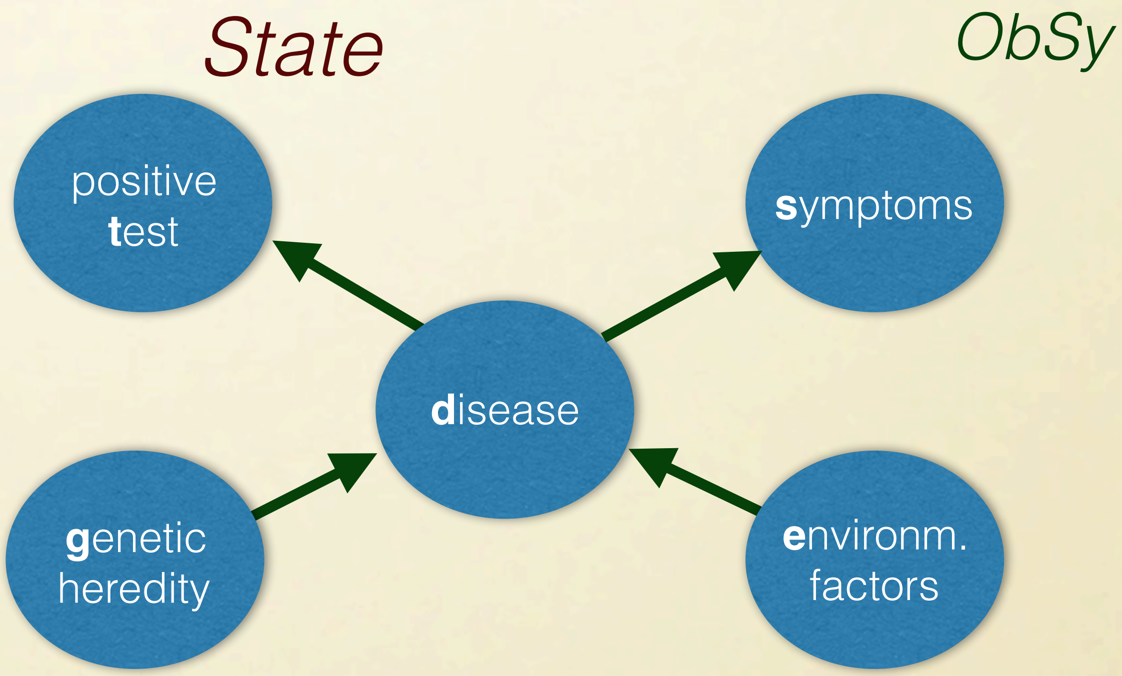
4. State transformation $T\gg(D\gg(G\otimes E))_{S\ll ObSy}) \in \mathcal{D}(2_T)$

5. Answer is the conditioned state $(T\gg(D\gg(G\otimes E))_{S\ll ObSy}))_{PosTest} \in \mathcal{D}(2_T)$

Influence

The two questions show that siblings can influence each others.

$$T \gg D \gg (G \otimes E)$$
$$\neq$$
$$(T \gg (D \gg (G \otimes E))_{S \ll ObSy}))$$

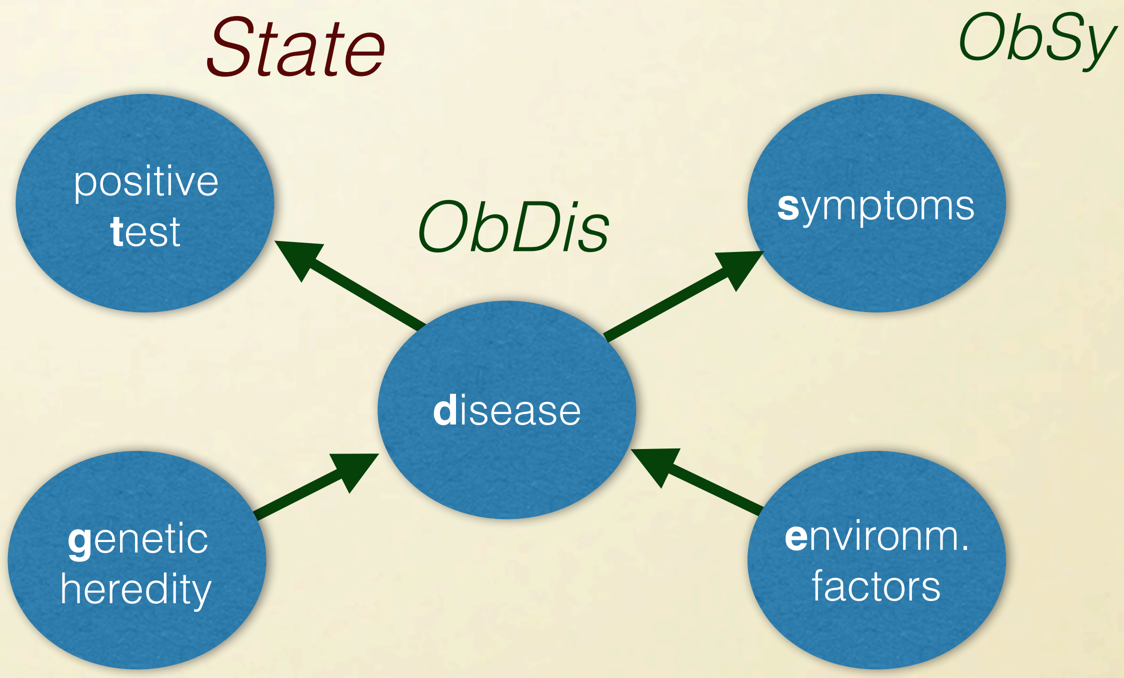


Blocking Influence

But this influence can be **blocked.** The channel language is able to express and formally prove it.

$$T \gg D \gg (G \otimes E)$$
$$\neq$$
$$(T \gg (D \gg (G \otimes E))_{S \ll ObSy}))$$

$$T \gg (D \gg (G \otimes E))_{ObDis}$$
$$=$$
$$(T \gg ((D \gg (G \otimes E))_{ObDis})_{S \ll ObSy}))$$



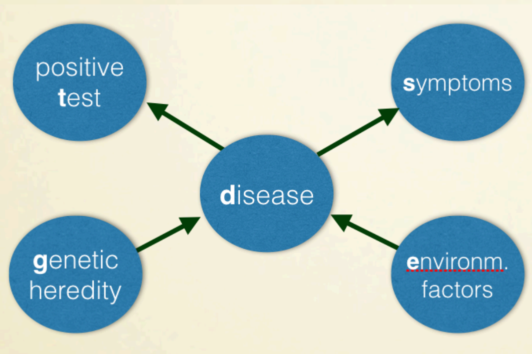
Influence: overview

- More generally, the channel language allows to prove the three **d-separation** scenarios as formal statements.

- Influence can be formally quantified, via a (total variation) distance between states.

Back to inference

Predicate/state transformation in **kl($\mathcal{D}$)** offers a novel, dynamical style of performing Bayesian inference.



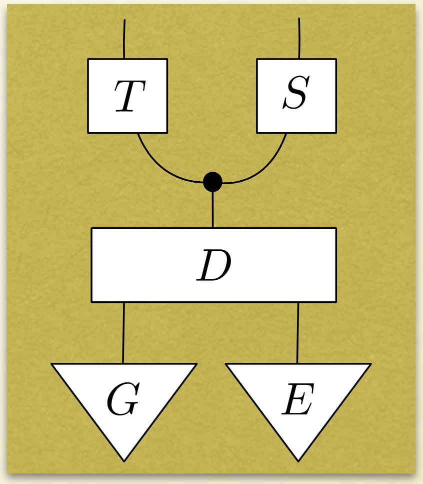
g
1/50

e
1/10

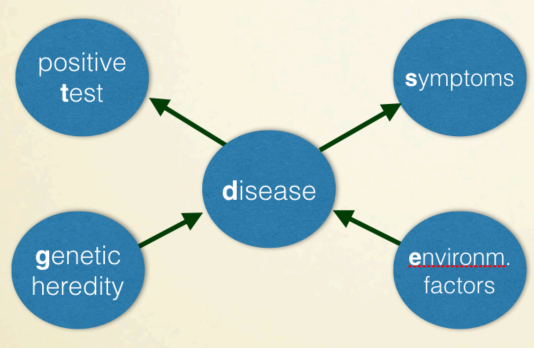
	d
g e	9/10
g e$^\perp$	8/10
g$^\perp$ e	4/10
g$^\perp$ e$^\perp$	0

	t
d	9/10
d$^\perp$	1/20

	s
d	9/10
d$^\perp$	1/15



In **kl($\mathcal{D}$)** we can reproduce also a more traditional account of Bayesian inference, in which belief revision is performed on the whole joint distribution.



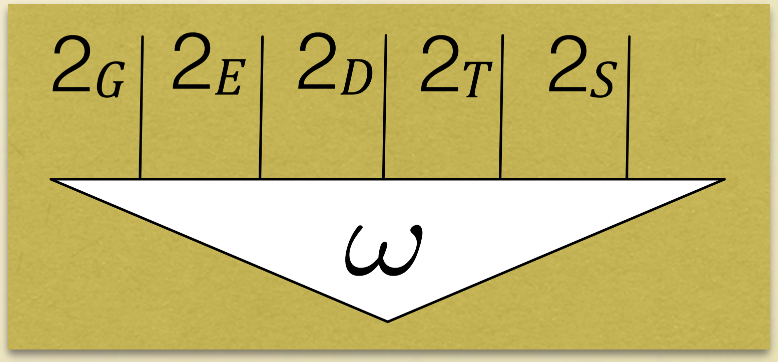
g
1/50

e
1/10

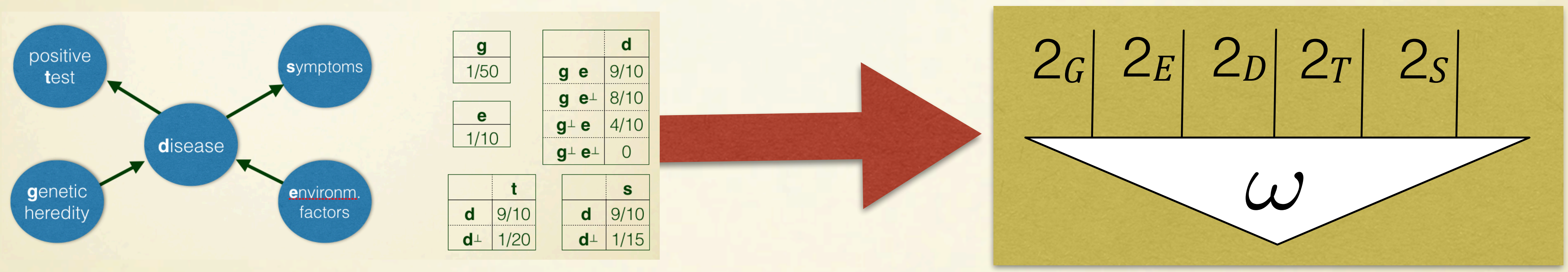
	d
g e	9/10
g e$^\perp$	8/10
g$^\perp$ e	4/10
g$^\perp$ e$^\perp$	0

	t
d	9/10
d$^\perp$	1/20

	s
d	9/10
d$^\perp$	1/15



Back to inference



Inference questions

What is the a priori probability of a positive test?

Answer:

$$M_4(\omega|_{(\mathrm{Id}\otimes\mathrm{Id}\otimes\mathrm{Id}\otimes \mathit{PosTest}\otimes\mathrm{Id})})$$

Fourth marginal (positive test node)

What is the probability of a positive test ***given*** *the symptoms?*

Answer:

$$M_4(\omega|_{(\mathrm{Id}\otimes\mathrm{Id}\otimes\mathrm{Id}\otimes \mathit{PosTest}\otimes \mathit{ObSy})})$$

Back to inference

- It turns out the two styles of inference are provably **equivalent**.

- This can be made formally precise once we explain **disintegration**: the process of factorising a given joint state into a Bayesian network.