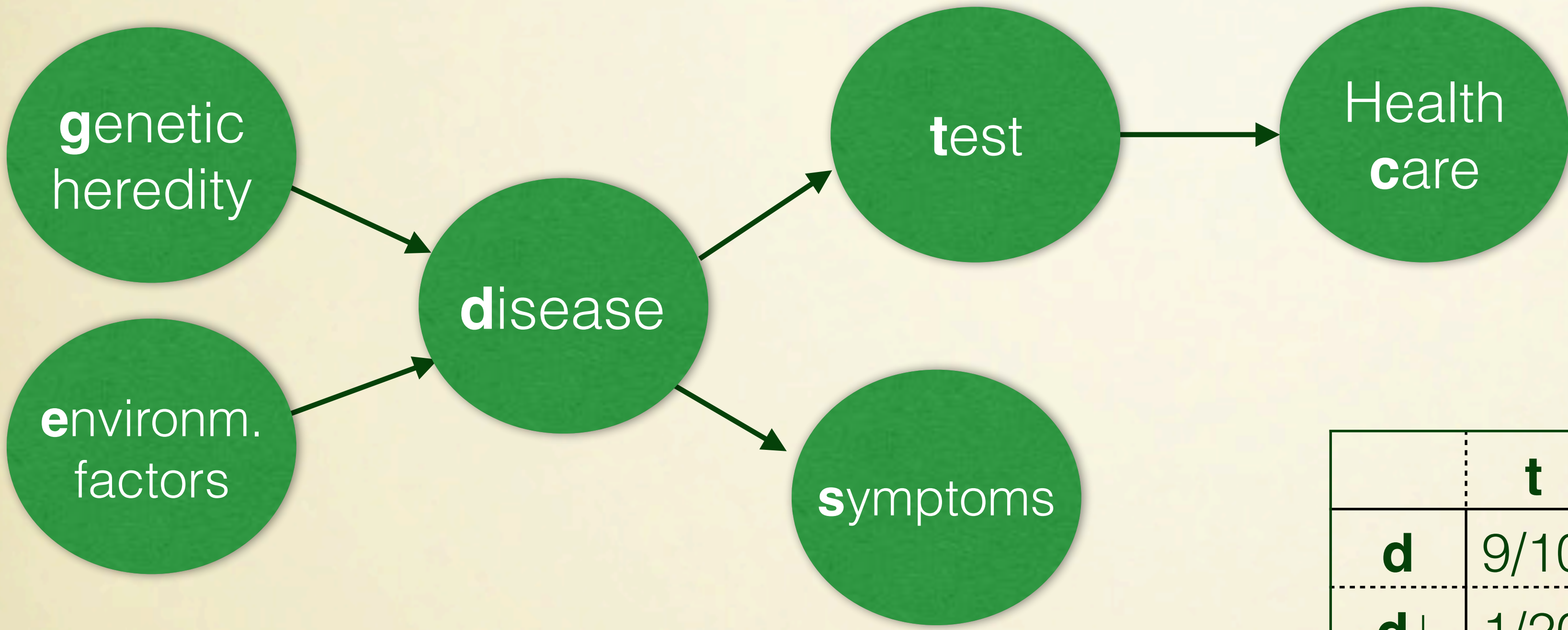
A Formal Semantics for Bayesian Reasoning

Fabio Zanasi
University College London

Joint work with Bart Jacobs

1. Bayesian Reasoning, traditionally

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

	c
t	4/5
$t^{\perp}$	1/10

Learning

Forward inference

P(T|¬G)

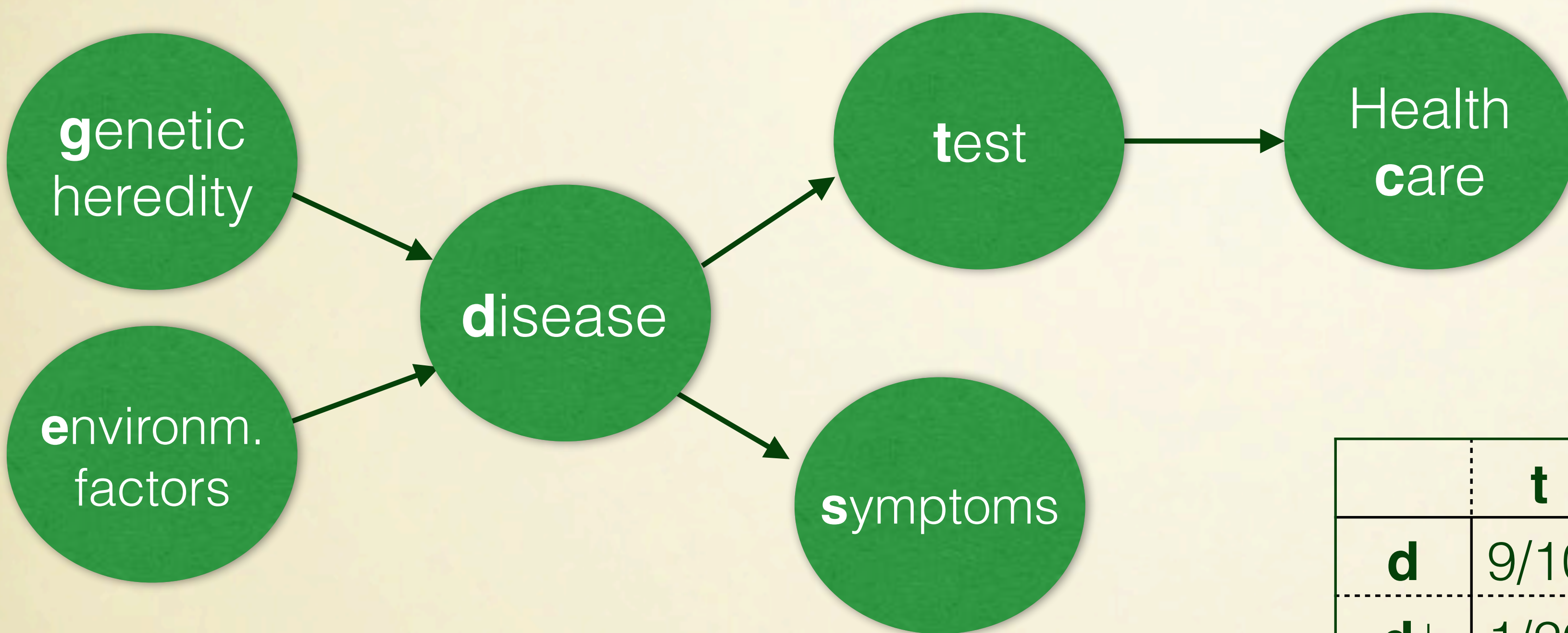
If genetic heredity is excluded, what is the likelihood of a positive test?

Backward inference

P(D|T)

If a patient tests positive, what is the likelihood that she had the disease?

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

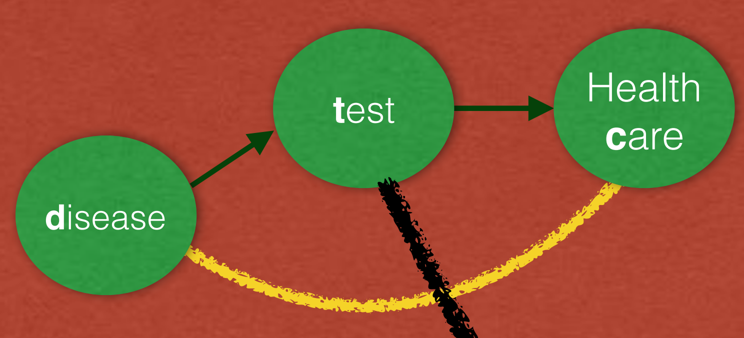
	t
d	9/10
$d^{\perp}$	1/20

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

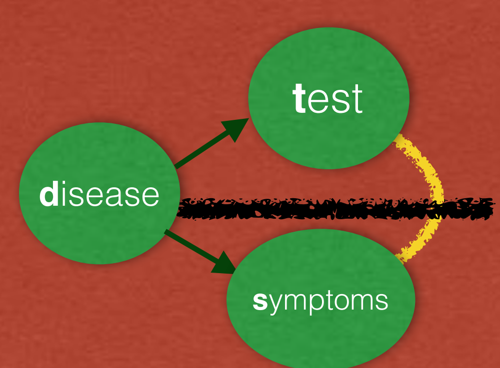
	c
t	4/5
$t^{\perp}$	1/10

Influence and blocking

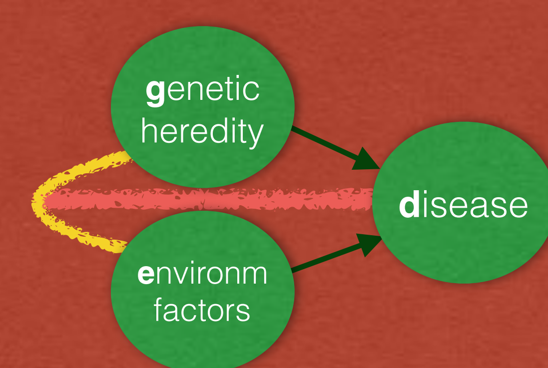
Serial connection



Fork



Collider



Plan

Categorical semantics based on new primitives (from *Effectus Theory*)

State	Predicate	Validity	Conditioning
$\omega \in \mathcal{D}(X)$	$p : X \to [0,1]$	$\omega \models p \in [0,1]$	$\omega\|_p \in \mathcal{D}(X)$

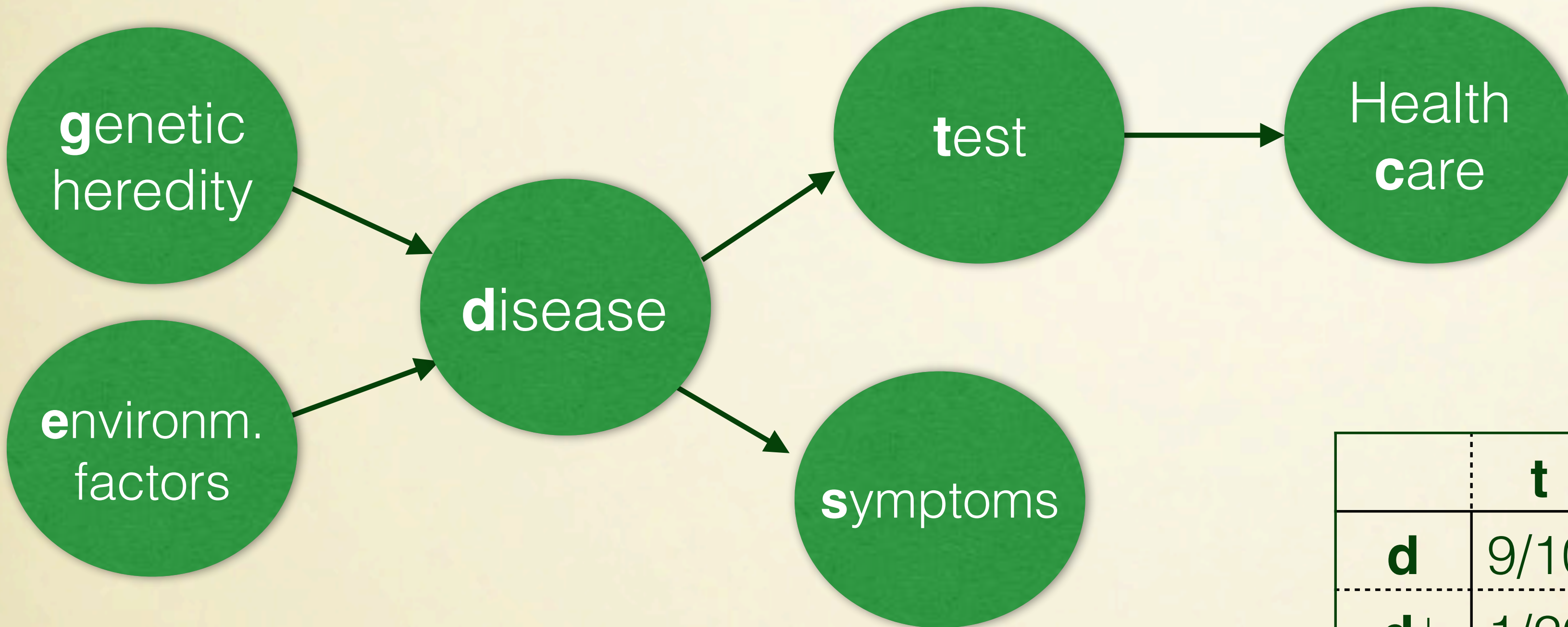
Backward inference = Predicate transformer + conditioning

Forward inference = Conditioning + state transformer

Influence = State entwinedness (+ metric)

2. Categorical Semantics

Bayesian Networks are Kleisli arrows



g
1/50

e
1/10

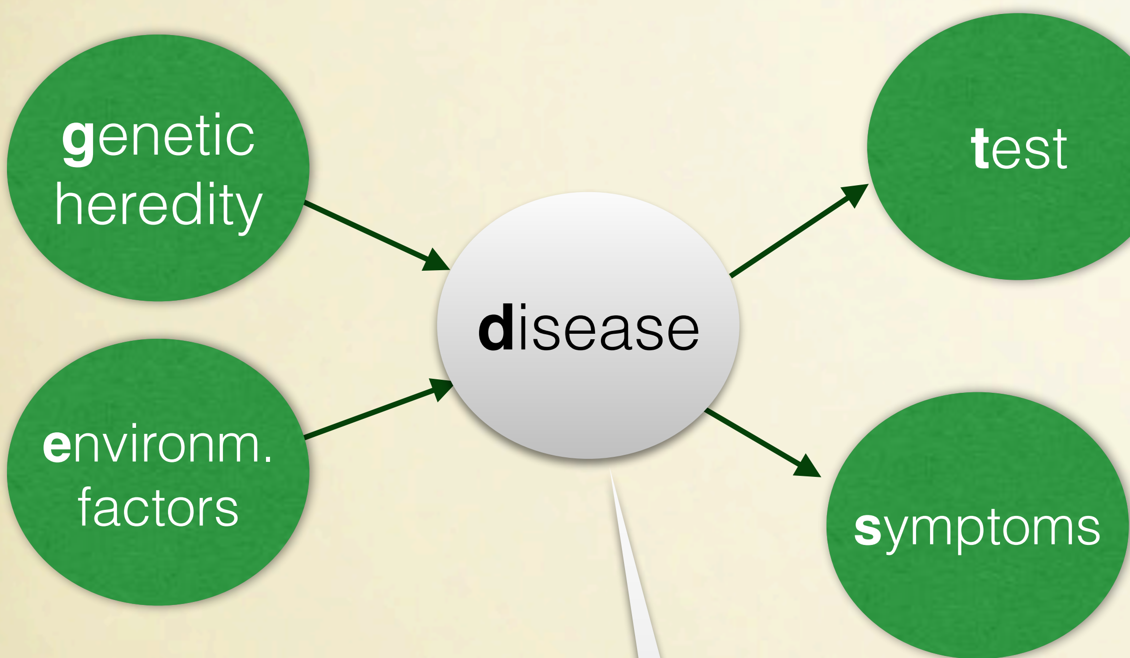
	d
g e	9/10
g $e^\perp$	8/10
$g^\perp$ e	4/10
$g^\perp$ $e^\perp$	0

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

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

	c
t	4/5
$t^\perp$	1/10

Bayesian Networks are Kleisli arrows



g
1/50

e
1/10

	d
g e	9/10
g e⊥	8/10
g⊥ e	4/10
g⊥ e⊥	0

	t
d	9/10
d⊥	1/20

	s
d	9/10
d⊥	1/15

	c
t	4/5
t⊥	1/10

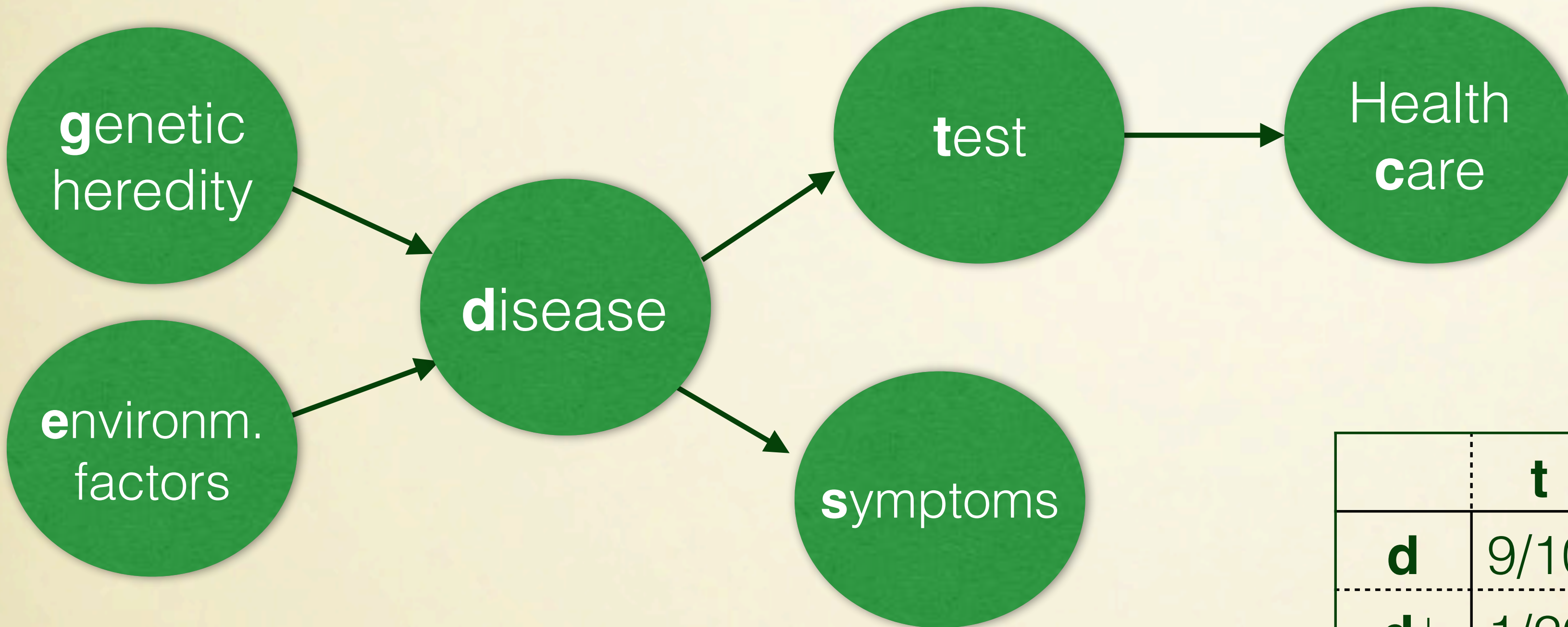
Kl($\mathcal{D}$)

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

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

$$\begin{aligned}
(g, e) &\mapsto 9/10|d\rangle + 1/10|d^\perp\rangle \\
(g, e^\perp) &\mapsto 8/10|d\rangle + 2/10|d^\perp\rangle \\
(g^\perp, e) &\mapsto 4/10|d\rangle + 6/10|d^\perp\rangle \\
(g^\perp, e^\perp) &\mapsto 1|d^\perp\rangle
\end{aligned}$$

Bayesian Networks are Kleisli arrows



g
1/50

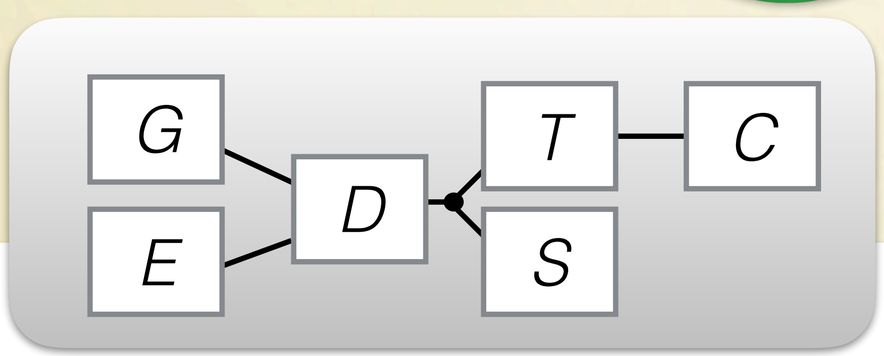
e
1/10

	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

	c
t	4/5
$t^{\perp}$	1/10



$\mathrm{Kl}(\mathcal{D})$

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

Toolbox

Validity

$$\omega \models p \;=\; p \circ \omega \;= 0,36$$

State $\quad 1 \circ\!\!\rightarrow X$

$$\omega = \frac{1}{10}|d\rangle + \frac{9}{10}|d^{\perp}\rangle$$

Conditioning

$$\omega|_p = \sum_{a \in 2_d} \frac{\omega(a)\cdot p(a)}{\omega \models p}|a\rangle = \tfrac{1}{4}|d\rangle + \tfrac{3}{4}|d^{\perp}\rangle$$

Predicate $\quad X \circ\!\!\rightarrow 2$

$$p\colon d \mapsto 0.9 \quad d^{\perp} \mapsto 0.3$$

$$\begin{tikzcd}
(\mathrm{Pred}_{\mathcal{K}\ell(\mathcal{D})})^{\mathrm{op}} \arrow[rr, bend left] \arrow[rr, phantom, "\top"] & & \mathrm{Stat}_{\mathcal{K}\ell(\mathcal{D})} \arrow[ll, bend left] \\
& \mathcal{K}\ell(\mathcal{D}) \arrow[ul, "\mathrm{Pred}=\mathrm{Hom}(-,2)"] \arrow[ur, "\mathrm{Stat}=\mathrm{Hom}(1,-)"'] &
\end{tikzcd}$$

State transformer

$$\mathrm{Stat}(X) \xrightarrow{f_*(\cdot) \,=\, f\circ(\cdot)} \mathrm{Stat}(Y)$$

Predicate transformer

$$\mathrm{Pred}(X) \xleftarrow{f^*(\cdot) \,=\, (\cdot)\circ f} \mathrm{Pred}(Y)$$

3. Learning

Backward Inference

If a patient tests positive, what is the likelihood that she had the disease?

$$\begin{tikzcd}
1 \arrow[r, "G\otimes E"] \arrow[rr, bend right, "(D\circ(G\otimes E))\big|_{T^*(\delta_t)}"'] & 2_g \otimes 2_e \arrow[r, "D"] & 2_d \arrow[r, "T"] \arrow[dr, bend right, "T^*(\delta_t)"'] & 2_t \arrow[d, "\delta_t"] \\
& & & 2
\end{tikzcd}$$

$t \mapsto 1$

$t^\perp \mapsto 0$

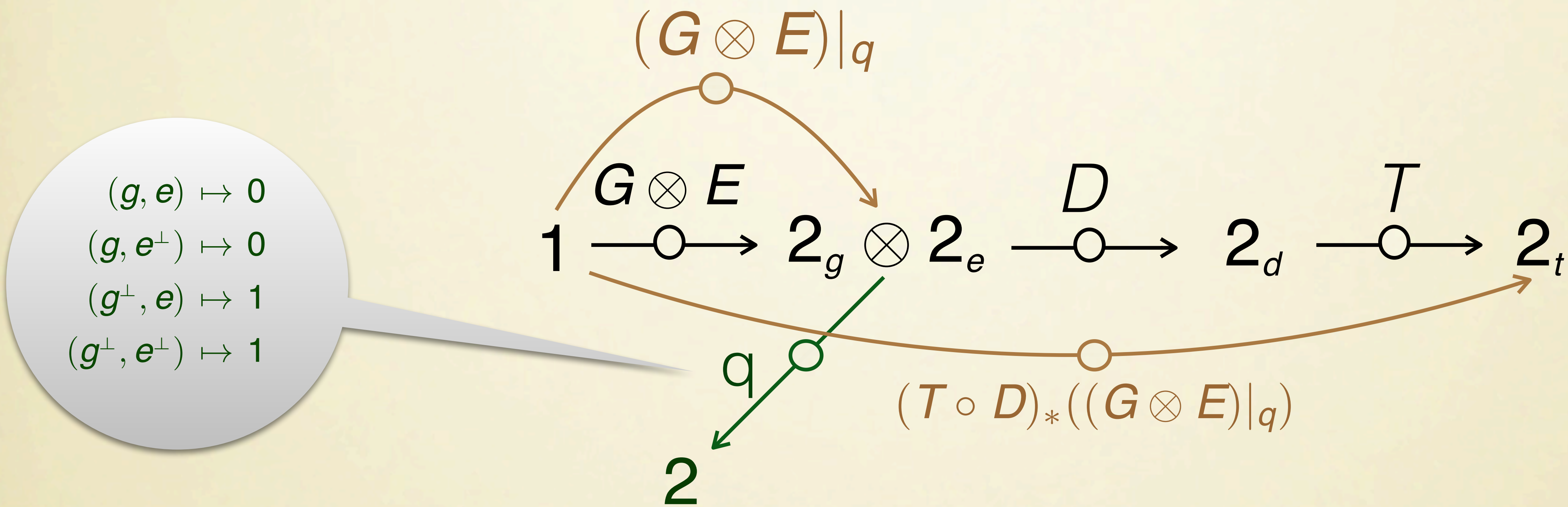
$$(D \circ (G \otimes E)) = 0{,}06\big|\, d\,\big\rangle + 0{,}94\big|\, d^\perp\,\big\rangle$$

$$(D \circ (G \otimes E))\big|_{T^*(\delta_t)} = 0{,}51\big|\, d\,\big\rangle + 0{,}49\big|\, d^\perp\,\big\rangle$$

Backward learning = Predicate transformer + conditioning

Forward Inference

If genetic heredity is excluded, what is the likelihood of a positive test?

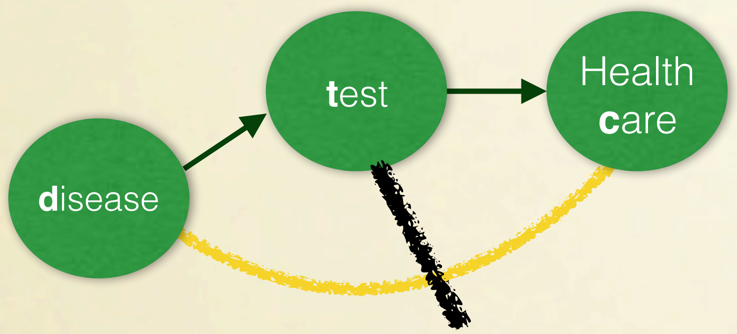


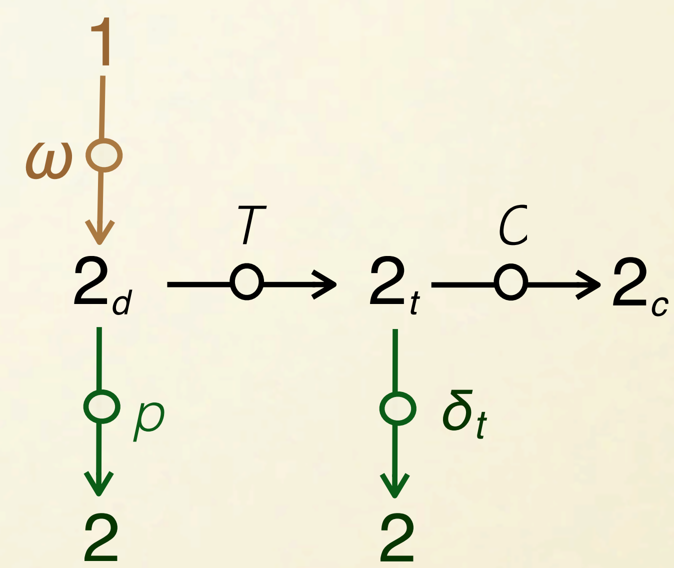
$$(T \circ D)_*(G \otimes E) = 0.1|t\rangle + 0.9|t^{\perp}\rangle$$
$$(T \circ D)_*((G \otimes E)|_q) = 0.06|t\rangle + 0.94|t^{\perp}\rangle$$

Forward learning = Conditioning + state transformer

4. Influence

Serial Connection



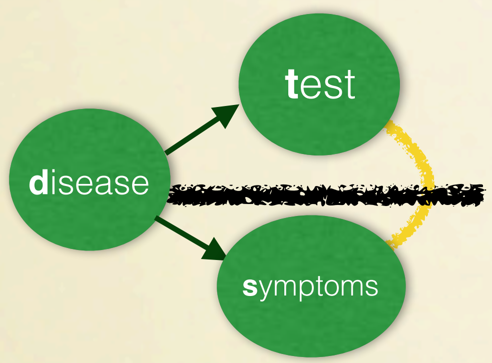


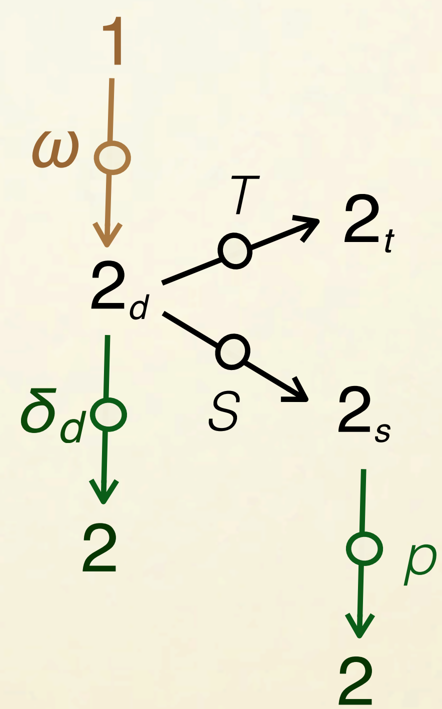
$$H_*(T_*(\omega)) \neq H_*(T_*(\omega|_p))$$

δ_t blocks the influence of *p:*

$$H_*(T_*(\omega)|_{\delta_t}) = H_*(T_*(\omega|_p)|_{\delta_t})$$

Fork



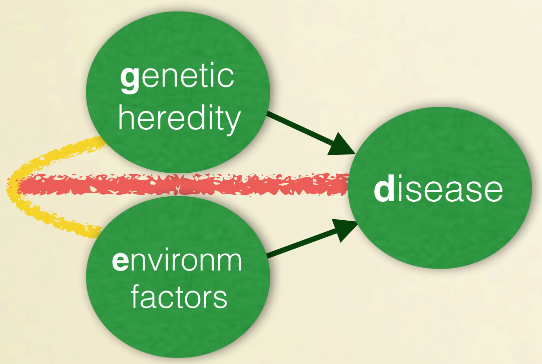


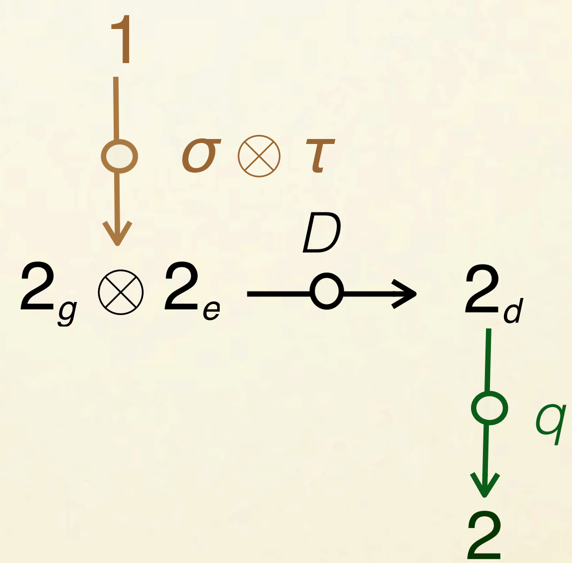
$$T_*(\omega) \neq T_*(\omega|_{S^*(p)})$$

$\boldsymbol{\delta_d}$ blocks the influence of *S*(p)*

$$T_*(\omega|_{\delta_d}) = T_*((\omega|_{S^*(p)})|_{\delta_d})$$

Collider





What we want to express is:

in the conditioned state $(\sigma\otimes\tau)|_{D^*(q)}$, acting on τ may influence σ (and viceversa).

State Entwinedness (non-locality)

Marginal operations M_1 M_2 on joint states

$$\omega : 1 \mathrel{-\!\circ\!\!\rightarrow} 2_g \otimes 2_e$$

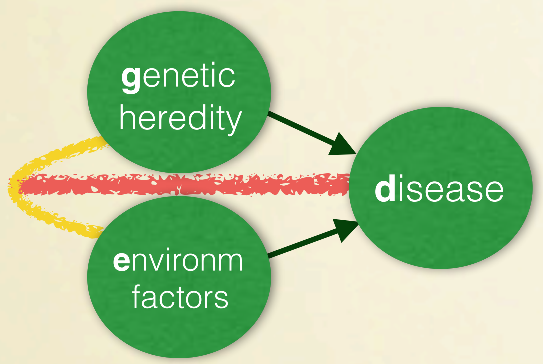
$$M_1(\omega) : 1 \mathrel{-\!\circ\!\!\rightarrow} 2_g \qquad\qquad M_2(\omega) : 1 \mathrel{-\!\circ\!\!\rightarrow} 2_e$$

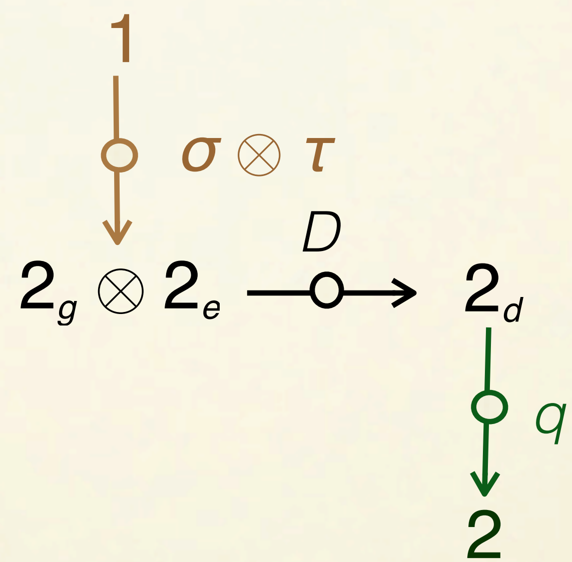
A joint state is **entwined** if it is the product of its marginals

$$M_1(\omega) \otimes M_2(\omega) \;=\; \omega \qquad \text{non-entwined}$$

$$M_1(\omega) \otimes M_2(\omega) \;\neq\; \omega \qquad \text{entwined}$$

Collider, take two





What we want to express is:

in the conditioned state $(\sigma\otimes\tau)|_{D^*(q)}$, acting on τ may influence σ (and viceversa).

Conditioning creates entwinedness

$\sigma \otimes \tau = \mathrm{M}_1(\sigma \otimes \tau) \otimes \mathrm{M}_2(\sigma \otimes \tau)$ | $(\sigma \otimes \tau)|_{D^* q}$

non-entwined | **entwined**

Quantifying Influence

Conditioning a component of an entwined state has ``crossover influence over the other component. *How much?*

Crossover influence of $p\colon 2_g \rightsquigarrow 2$ on $\omega\colon 1 \rightsquigarrow 2_g \otimes 2_e$

$$\mathcal{I}_c(p,\omega) = \mathrm{d}\Big(\mathsf{M}_2(\omega), \mathsf{M}_2\big(\omega|_{\mathsf{W}_1(p)}\big)\Big)$$

total variation distance

$$\mathrm{d}(\omega_1,\omega_2) = \frac{1}{2}\sum_{a\in A}\big|\omega_1(a) - \omega_2(a)\big|.$$

yields a functor $\mathrm{Kl}(\mathcal{D}) \to \mathrm{Met}_1$

Weakening

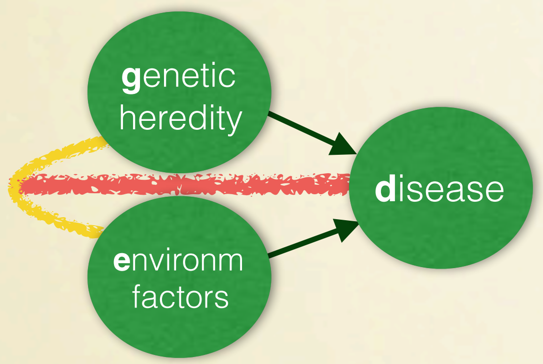
$p\colon 2_g \rightsquigarrow 2$

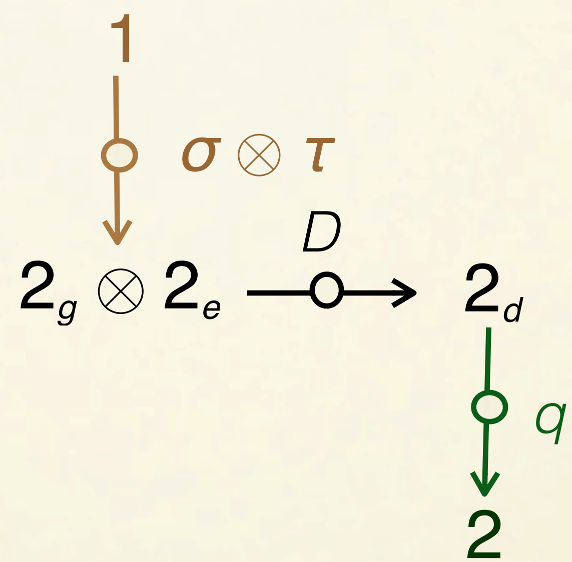
$\mathsf{W}_1(p)\colon 2_g \otimes 2_e \rightsquigarrow 2$

Direct influence of $p\colon 2_g \rightsquigarrow 2$ on $\tau\colon 1 \rightsquigarrow 2_g$

$$\mathcal{I}_d(p,\tau) = \mathrm{d}(\tau, \tau|_p)$$

Collider, take three





What we want to express is:

in the conditioned state $(\sigma \otimes \tau)|_{D^*(q)}$, acting on τ may influence σ (and viceversa).

Conditioning creates entwinedness

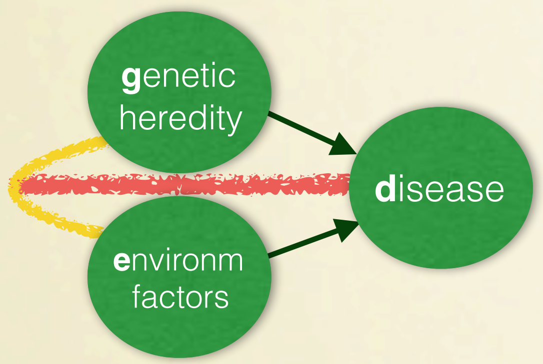
$\sigma \otimes \tau = \mathrm{M}_1(\sigma \otimes \tau) \otimes \mathrm{M}_2(\sigma \otimes \tau)$

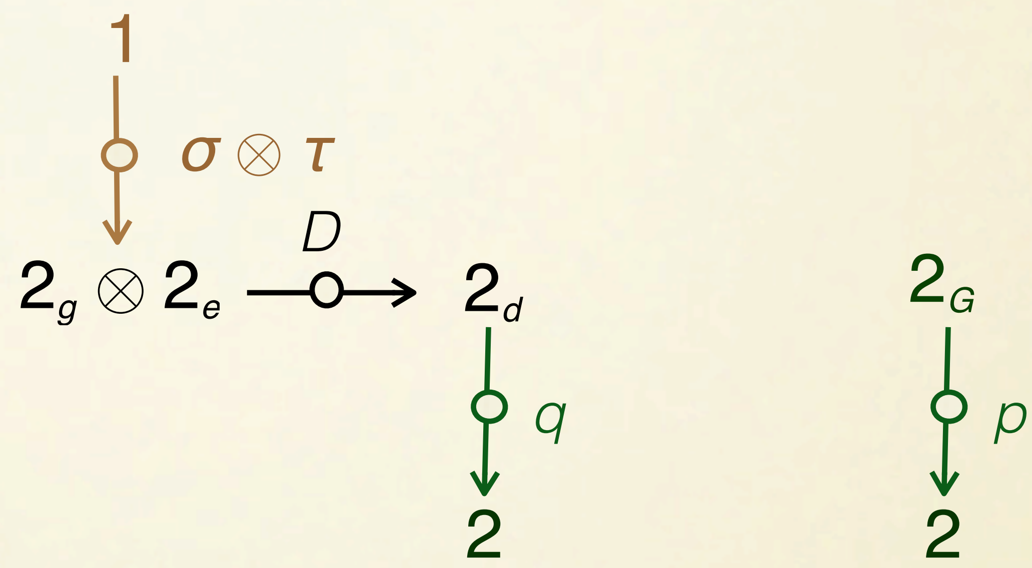
non-entwined

$(\sigma \otimes \tau)|_{D^* q}$

entwined

Collider, take three





What we want to express is:

in the conditioned state $(\sigma\otimes\tau)|_{D^*(q)}$, acting on τ may influence σ (and viceversa).

Conditioning creates entwinedness

$\sigma \otimes \tau = \mathrm{M}_1(\sigma \otimes \tau) \otimes \mathrm{M}_2(\sigma \otimes \tau)$	$(\sigma \otimes \tau)\|_{D^* q}$
non-entwined	**entwined**
$\mathcal{I}_c(p, \sigma \otimes \tau) = 0$	$\mathcal{I}_c(p, (\sigma \otimes \tau)\|_{D^*(q)}) > 0$

Conclusions

- Grounding concepts of Bayesian reasoning (learning, evidence, influence, dependence, blocking, ...) are given a transparent, completely formal meaning in terms of standard constructs of programming language semantics.
- (First?) use of metrics to measure influence.
- Proof of concept: the three d-separation scenarios (serial, fork, collider).
- What's next:
  - other Bayesian phenomena (countefactuals)
  - other metrics (Kantorovich-Wasserstein)
  - other domains (quantum, via Effectus theory).