Cyber Kittens

or Some First Steps Towards Categorical Cybernetics

Toby St Clere Smithe
University of Oxford

ACT 2020, 7 July 2020

UNIVERSITY OF OXFORD

Experimental Psychology

Background story

Presently a theoretical neuroscientist,
interested in how neurons composed
together generate intelligent behaviour

How can we construct a system that plays the games that we study?

Heuristic definition of cybernetic system

"If it perceives and acts, then it is a cybernetic system"

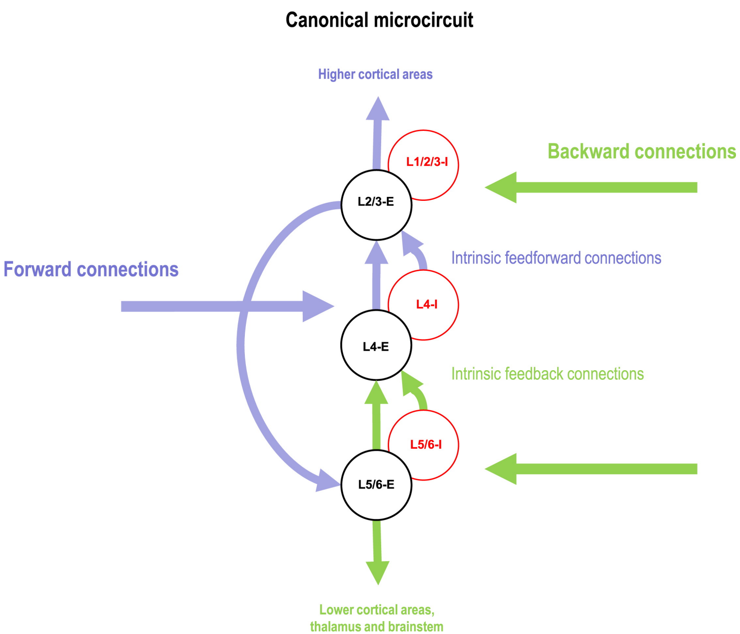
Typically no access to external state
→ must **infer what's going on**,
and **what should be done**

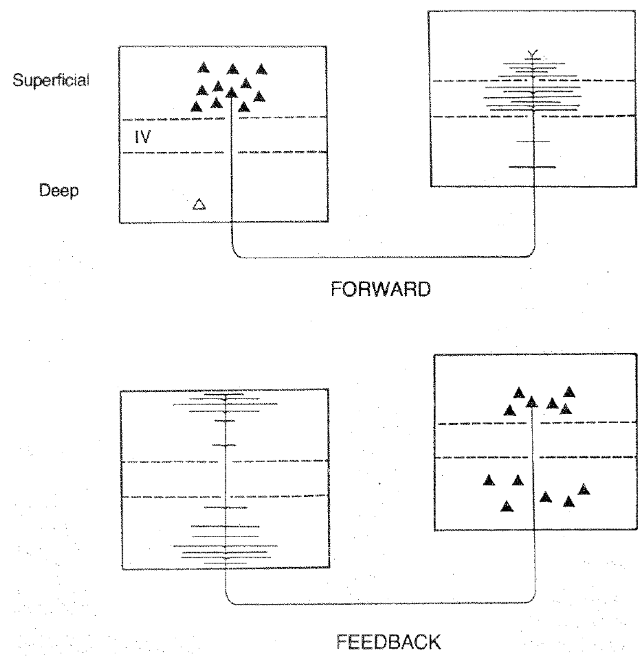
Inference: on the basis of imperfect signals

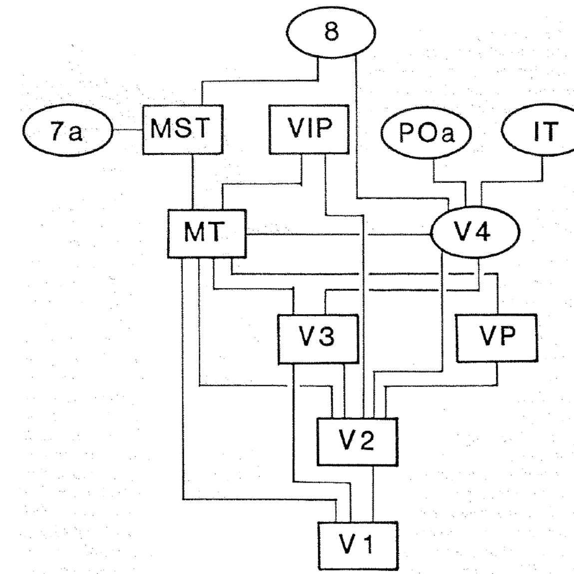
Brain as archetypal cybernetic system

Pervasive cortical structure:
bidirectional circuits

And 'hierarchically' organized
– like a traced monoidal cat. !







Bastos *et al* (2012)

Van Essen & Maunsell (1983)

Can explain both of these features abstractly:

- perceiving and acting mean doing Bayesian inference
- which in turn means embodying a model of the world to be inverted
- the inverse of a composite channel is the composite of the inverses
- so we can invert each factor of the model locally
  $\longrightarrow$ ‘hierarchical’ structure
- and the ‘bidirectional’ structure is precisely the *lens* pattern

Plan:

- <u>Introduce</u>: categorical probability, Bayesian inversion (very briefly)

- <u>Prove</u>: Bayesian updates compose according to the *lens* pattern

- <u>Define</u>: a class of *statistical games* using compositional game theory

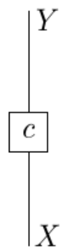
- <u>Suggest</u>: cybernetic systems are dynamical realisations of statistical games

- <u>Exemplify</u>: variational autoencoders, cortical circuits

- <u>Conclude</u>: interacting / nested systems? non-stationary contexts?

Basic setting: categorical probability

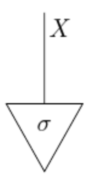
We work in a *Markov* or *copy-delete* category – **canonical example**: $\mathcal{K}\ell(\mathcal{D})$



Objects: spaces X, Y — sets X, Y

Morphisms: "stochastic channels" — *ie.* functions from points to 'beliefs'

$$X \,{\bullet\!\!\rightarrow}\, Y \qquad\qquad X \to \mathcal{D}Y \;\cong\; X \times Y \to [0,1]$$



States: channels out of the monoidal unit — *ie.* probability distributions (formal convex sums)

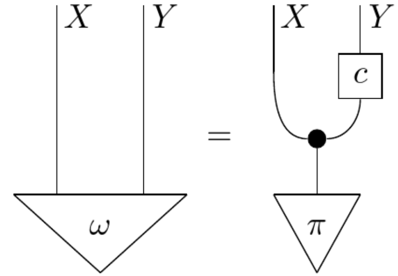
$$I \,{\bullet\!\!\rightarrow}\, X \qquad\qquad X \to [0,1] \qquad\qquad \textstyle\sum_{x:X} \boxed{p(x)}\, |x\rangle$$

so general channels are like 'conditional' probability distributions,
and we adopt the standard notation $p(y|x) := p(x)(y)$

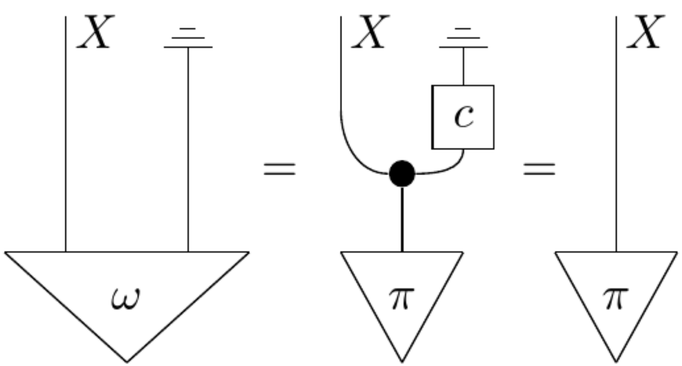
Composition: given $p : X \,{\bullet\!\!\rightarrow}\, Y$ and $q : Y \,{\bullet\!\!\rightarrow}\, Z$, "average over" Y — for example:

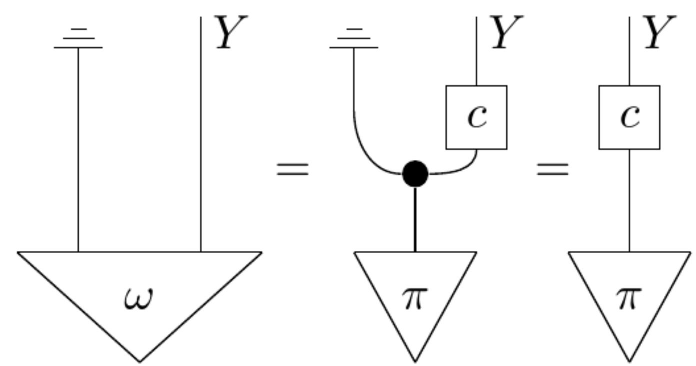
$$q \bullet p : X \to \mathcal{D}Z := x \mapsto \sum_{z:Z} \boxed{\sum_{y:Y} q(z|y) \cdot p(y|x)}\; |z\rangle$$

Joint states



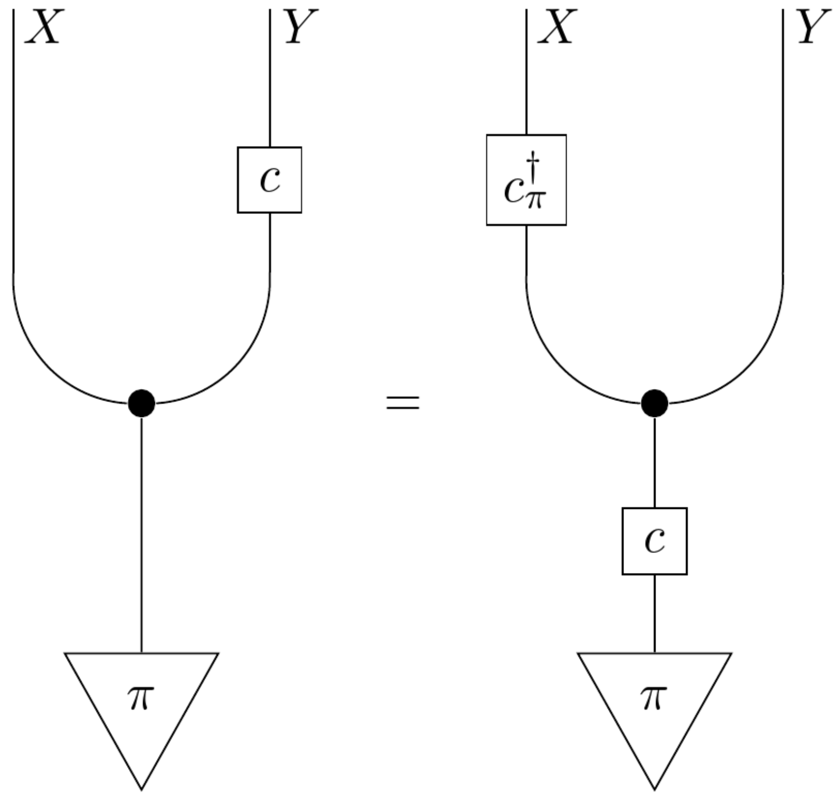
$$P_\omega(A, B) = P_c(B|A) \cdot P_\pi(A)$$





Bayesian inversion

$$\mathrm{P}_c(B|A) \cdot \mathrm{P}_\pi(A) = \mathrm{P}_{c_\pi^\dagger}(A|B) \cdot \mathrm{P}_{c \bullet \pi}(B)$$



NB: The Bayesian inverse of a channel is always defined *with respect to* some "prior" state !

What is $c^\dagger$?

An indexed category of state-dependent channels

$$\mathsf{Stat} \;:\; \mathcal{K}\ell(\mathcal{P})^{\text{op}} \to \mathbf{V\text{-}Cat}$$

$$X \mapsto \mathsf{Stat}(X) := \begin{pmatrix} & \mathsf{Stat}(X)_0 & := & \mathbf{Meas}_0 \\ & \mathsf{Stat}(X)(A,B) & := & \mathbf{Meas}(\mathcal{P}X, \mathbf{Meas}(A, \mathcal{P}B)) \\ \mathsf{id}_A : & \mathsf{Stat}(X)(A,A) & := & \begin{cases} \mathsf{id}_A : \mathcal{P}X \to \mathbf{Meas}(A, \mathcal{P}A) \\ \quad \rho \;\mapsto\; \eta_A \end{cases} \end{pmatrix}$$

$\mathsf{Stat}(X)$ is a category of stochastic channels with respect to states on X

Morphisms $d^\dagger : \mathcal{P}X \to \mathcal{K}\ell(\mathcal{P})(A,B)$ in $\mathsf{Stat}(X)$
are generalized Bayesian inversions:

given a state π on X, obtain a
channel $d^\dagger_\pi : A \to\!\!\!\!\bullet\; B$ with respect to π

An indexed category of state-dependent channels

$$\mathsf{Stat} \;:\; \mathcal{K}\ell(\mathcal{P})^{\mathsf{op}} \to \mathbf{V\text{-}Cat}$$

$$X \mapsto \mathsf{Stat}(X) := \begin{pmatrix} & \mathsf{Stat}(X)_0 & := & \mathbf{Meas}_0 \\ & \mathsf{Stat}(X)(A,B) & := & \mathbf{Meas}(\mathcal{P}X, \mathbf{Meas}(A, \mathcal{P}B)) \\ \mathsf{id}_A : & \mathsf{Stat}(X)(A,A) & := & \begin{cases} \mathsf{id}_A : \mathcal{P}X \to \mathbf{Meas}(A, \mathcal{P}A) \\ \quad \rho \mapsto \eta_A \end{cases} \end{pmatrix}$$

$$c : \mathcal{K}\ell(\mathcal{P})(Y,X) \mapsto \begin{pmatrix} \mathsf{Stat}(c) : & \mathsf{Stat}(X) & \to & \mathsf{Stat}(Y) \\ & \mathsf{Stat}(X)_0 & = & \mathsf{Stat}(Y)_0 \\ & \begin{pmatrix} d^\dagger : & \mathcal{P}X & \to \mathcal{K}\ell(\mathcal{P})(A,B) \\ & \pi & \mapsto d^\dagger_\pi \end{pmatrix} & \mapsto & \begin{pmatrix} c^* d^\dagger : \mathcal{P}Y \to \mathcal{K}\ell(\mathcal{P})(A,B) \\ \rho \mapsto d^\dagger_{c \bullet \rho} \end{pmatrix} \end{pmatrix}$$

$\mathsf{Stat}(X)$ is a category of stochastic channels with respect to states on X

Morphisms $d^\dagger : \mathcal{P}X \to \mathcal{K}\ell(\mathcal{P})(A,B)$ in $\mathsf{Stat}(X)$ are generalized Bayesian inversions:

given a state π on X, obtain a channel $d^\dagger_\pi : A \to\!\!\!\!\!\bullet\; B$ with respect to π

Given $c : Y \to\!\!\!\!\!\bullet\; X$ in the base, can pull $d^\dagger$ back along c, obtaining $c^* d^\dagger : \mathcal{P}Y \to \mathcal{K}\ell(\mathcal{P})(A,B)$

This takes $\rho : \mathcal{P}Y$ to $d^\dagger_{c\bullet\rho} : A \to\!\!\!\!\!\bullet\; B$ defined by pushing ρ through c then applying $d^\dagger$.

But: given $d \bullet c$, what is $(d \bullet c)^\dagger$?

Given $d \bullet c$, what is $(d \bullet c)^\dagger$?

If $\mathbf{Meas}$ is Cartesian closed (*e.g.*, quasi-Borel spaces), then $d^\dagger : \mathcal{P}A \to \mathcal{K}\ell(\mathcal{P})(B, A)$ is equivalently $\mathcal{P}A \times B \to \mathcal{P}A$.

Paired with a map $d : B \nrightarrow A$, this looks like a **simple lens**: classically, a pair of type $\mathbf{Set}(A, B) \times \mathbf{Set}(A \times B, A)$.

Here, we have $\mathcal{K}\ell(\mathcal{P})(A, B) \times \mathbf{Meas}(\mathcal{P}A \times B, \mathcal{P}A)$.

But this is just a hom-set in the Grothendieck construction of the pointwise opposite of Stat!

Let's check this ... and then see how these things compose.

Grothendieck lenses

Definition ($\mathbf{GrLens}_F$). Let $F : \mathcal{C}^{\text{op}} \to \mathbf{Cat}$.
Objects $(\mathbf{GrLens}_F)_0$: pairs (C, X) of objects C in $\mathcal{C}$ and X in $F(C)$.

Hom-sets $\mathbf{GrLens}_F\big((C,X),(C',X')\big)$: dependent sums

$$\mathbf{GrLens}_F\big((C,X),(C',X')\big) = \sum_{f\,:\,\mathcal{C}(C,C')} F(C)\big(F(f)(X'), X\big)$$

so $(C,X) \nrightarrow (C',X')$ is a pair $(f, f^\dagger)$ of $f : \mathcal{C}(C,C')$ and $f^\dagger : F(C)\big(F(f)(X'), X\big)$.

Identities: $\mathsf{id}_{(C,X)} = (\mathsf{id}_C, \mathsf{id}_X)$

Composition: suppose $(f, f^\dagger) : (C,X) \nrightarrow (C',X')$ and $(g, g^\dagger) : (C',X') \nrightarrow (D,Y)$.
Then $(g, g^\dagger) \mathbin{⦶} (f, f^\dagger) = \big(g \bullet f, F(f)(g^\dagger)\big) : (C,X) \nrightarrow (D,Y)$.

When $F = \mathsf{Stat} : \mathcal{K}\ell(\mathcal{P})^{\text{op}} \to \mathbf{Cat}$: $\mathbf{GrLens}_{\mathsf{Stat}}\big((X,A),(Y,B)\big) \cong \mathcal{K}\ell(\mathcal{P})(X,Y) \times \mathbf{Meas}\big(\mathcal{P}X, \mathcal{K}\ell(\mathcal{P})(B,A)\big)$

Given $(c, c^\dagger) : (X,A) \nrightarrow (Y,B)$ and $(d, d^\dagger) : (Y,B) \nrightarrow (Z,C)$,
$(d, d^\dagger) \mathbin{⦶} (c, c^\dagger) = \big((d \bullet c), (c^\dagger \circ c^* d^\dagger)\big) : (X,A) \nrightarrow (Z,C)$
where $(d \bullet c) : \mathcal{K}\ell(\mathcal{P})(X,Z)$ and
where $(c^\dagger \circ c^* d^\dagger) : \mathbf{Meas}\big(\mathcal{P}X, \mathcal{K}\ell(\mathcal{P})(C,A)\big)$ takes $\pi : \mathcal{P}X$ to $c^\dagger_\pi \bullet d^\dagger_{c\bullet\pi}$.

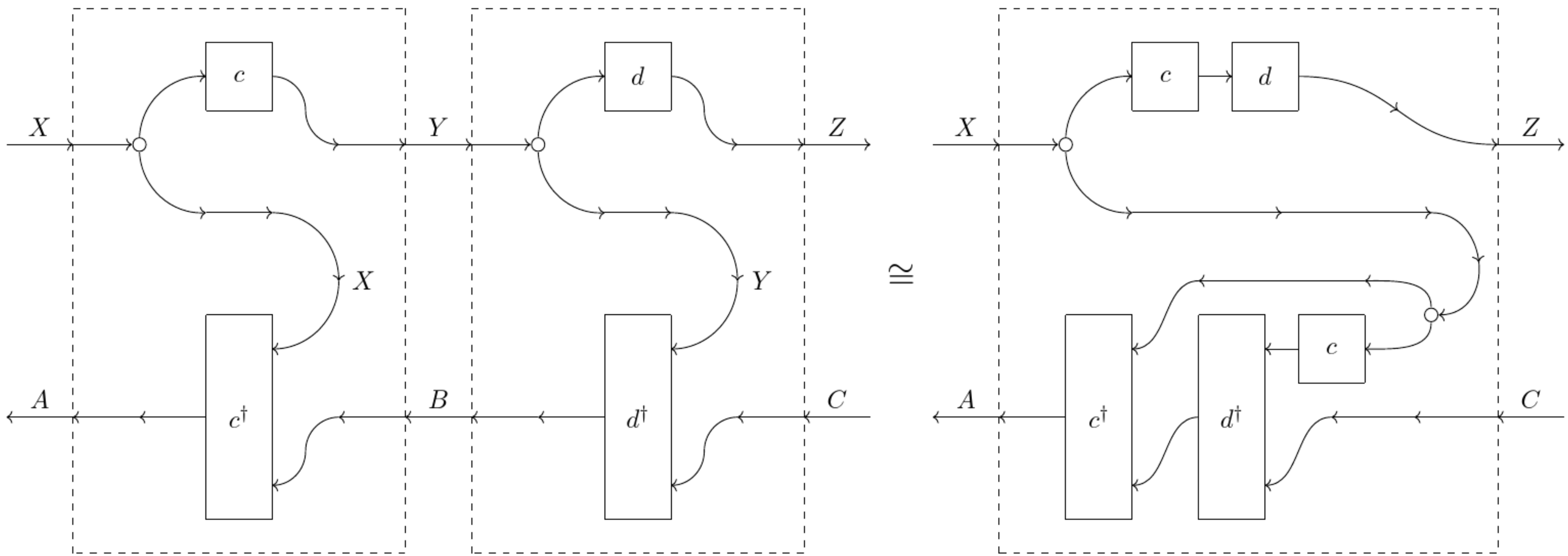
So we seek to show
$$(d \bullet c)^\dagger_\pi \simeq c^\dagger_\pi \bullet d^\dagger_{c\bullet\pi}$$

But first ...

An optical interlude

Optics are the contemporary home of compositional game theory

Plus, if our lenses are *optics*, then they acquire suggestive formal depictions:



And, indeed, Bayesian lenses *are* optics ...

An optical interlude

Proposition. $\mathbf{Optic}_{\times,\odot}\Big((\hat{X},\check{A}),(\hat{Y},\check{B})\Big) \cong \mathbf{GrLens}_{\mathsf{Stat}}\Big((X,A),(Y,B)\Big)$

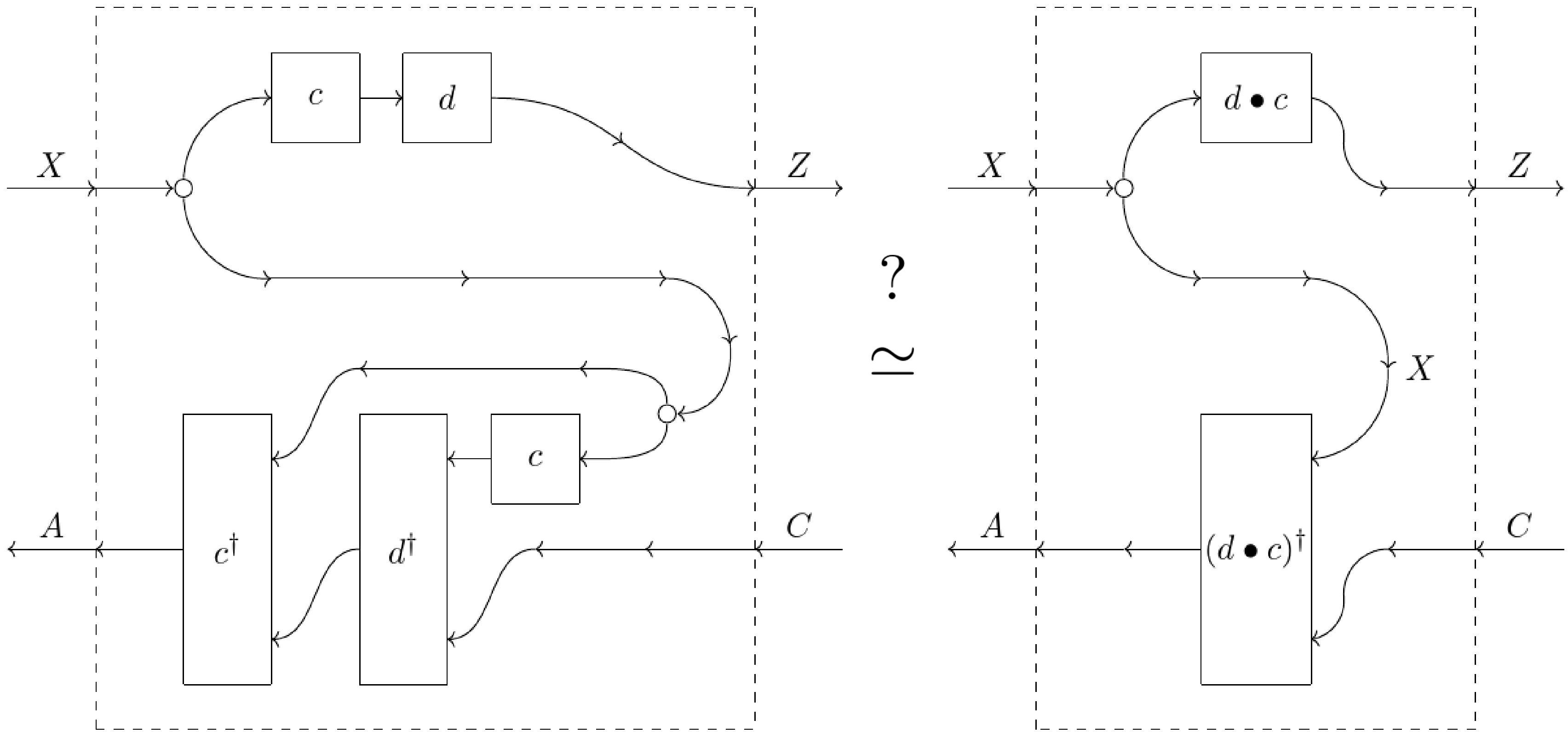
Proof: $\mathbf{Optic}_{\times,\odot}\Big((\hat{X},\check{A}),(\hat{Y},\check{B})\Big) = \int^{\hat{M}:\hat{\mathcal{C}}} \hat{\mathcal{C}}(\hat{X},\hat{M}\times\hat{Y}) \times \check{\mathcal{C}}(\hat{M}\odot\check{B},\check{A})$

$$\odot : \hat{\mathcal{C}} \to \mathbf{V\text{-}Cat}(\check{\mathcal{C}},\check{\mathcal{C}})$$
$$\hat{M} \mapsto \begin{pmatrix} \hat{M}\odot - & : & \check{\mathcal{C}} & \to & \check{\mathcal{C}} \\ & & P & \mapsto & \mathbf{V}\big(\hat{M}(I),P\big) \end{pmatrix}$$

$$\begin{aligned}
\mathbf{Optic}_{\times,\odot}\Big((\hat{X},\check{A}),(\hat{Y},\check{B})\Big) &\cong \int^{\hat{M}:\hat{\mathcal{C}}} \hat{\mathcal{C}}(\hat{X},\hat{Y}) \times \hat{\mathcal{C}}(\hat{X},\hat{M}) \times \check{\mathcal{C}}(\hat{M}\odot\check{B},\check{A}) \\
&\cong \int^{\hat{M}:\hat{\mathcal{C}}} \hat{\mathcal{C}}(\hat{X},\hat{Y}) \times \hat{\mathcal{C}}(\hat{X},\hat{M}) \times \check{\mathcal{C}}\left(\mathbf{V}(\hat{M}(I),\check{B}),\check{A}\right) \\
&\cong \int^{\hat{M}:\hat{\mathcal{C}}} \mathcal{C}(X,Y) \times \hat{M}(X) \times \mathbf{V}\left(\hat{M}(I),\mathcal{C}(B,A)\right) \\
&\cong \mathbf{GrLens}_{\mathsf{Stat}}\Big((X,A),(Y,B)\Big)
\end{aligned}$$

(And we can define 'mixed' Bayesian optics, too!)

Does Bayesian inversion commute with lens composition?



where $c^\dagger \circ \left(\mathsf{id}_{\check{X}} \odot \mathrm{d}^\dagger\right) \circ a^{\odot}_{\hat{X},\hat{Y},\check{X}} \circ \left(\mathsf{id}_{\hat{X}} \times c\right) \circ \curlyvee$

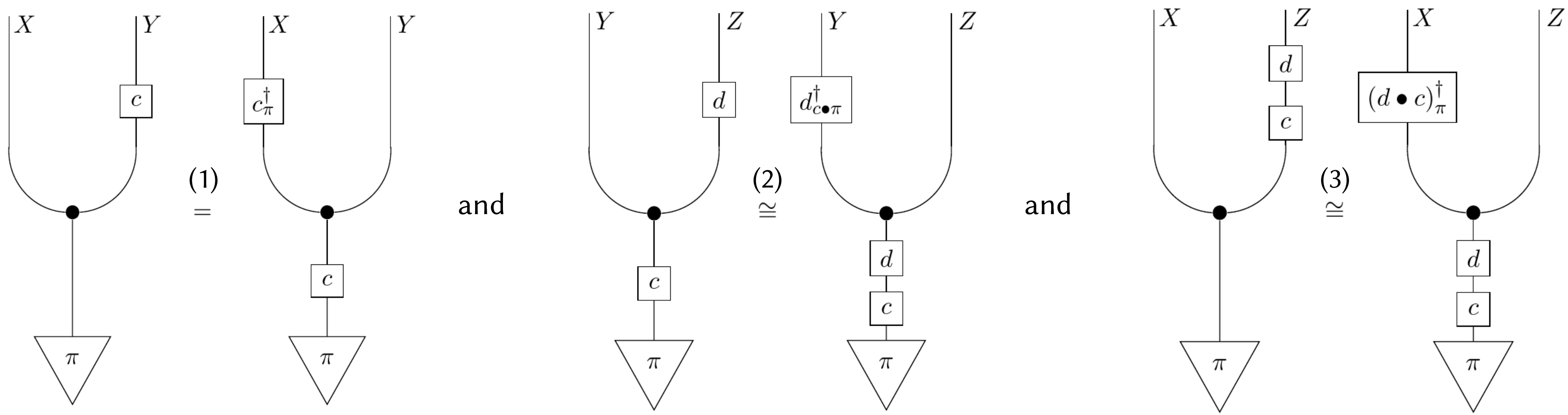
$$\cong \left(c^\dagger_{(-)} \bullet \mathrm{d}^\dagger_{c\bullet(-)}\right) \circ \curlyvee$$

$$\cong c^\dagger \circ c^* d^\dagger \cong c^\dagger_\pi \bullet d^\dagger_{c\bullet\pi}$$

Does Bayesian inversion commute with lens composition?

<u>Yes</u>! **Lemma** *(Bayesian updates compose optically).* $(d \bullet c)^\dagger_\pi \simeq c^\dagger_\pi \bullet d^\dagger_{c \bullet \pi}$

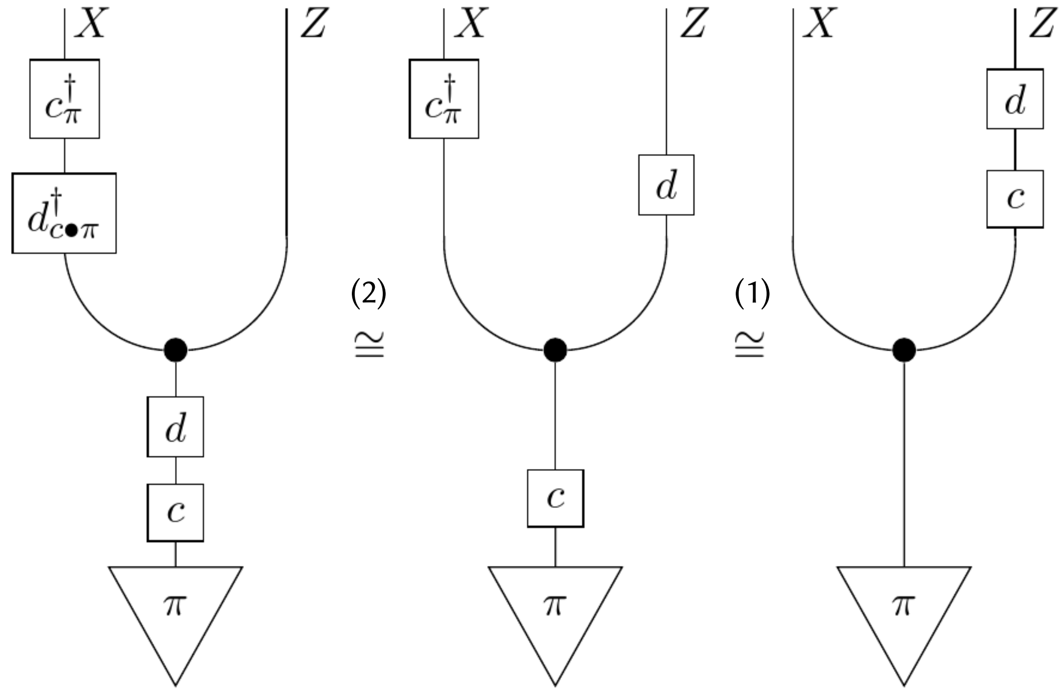
Suppose:



(These relations just define the relevant Bayesian inversions.)

Lemma *(Bayesian updates compose optically).* $(d \bullet c)^\dagger_\pi \simeq c^\dagger_\pi \bullet d^\dagger_{c\bullet\pi}$

Proof:



So $(d \bullet c)^\dagger_\pi$ and $c^\dagger_\pi \bullet d^\dagger_{c\bullet\pi}$ are both Bayesian inversions for $d \bullet c$ with respect to π.

But Bayesian inversions are almost-equal. Hence $(d \bullet c)^\dagger_\pi \simeq c^\dagger_\pi \bullet d^\dagger_{c\bullet\pi}$ □

Back to cybernetics

We will see: *inference problems are games over Bayesian lenses*

Recall: cybernetic system trying to estimate external state, given complex "generative model"

"In the wild": system will try to *improve* its estimation

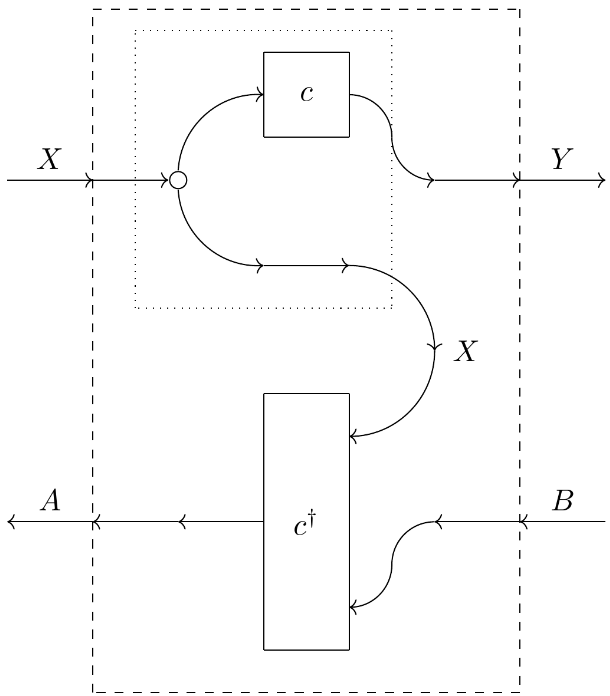
Note: all interactions of a cybernetic system are mediated through an interface (~ boundary)
– this is all the system has access to

Context := representation of boundary behaviour

First: yet another graphical calculus ...
(luckily, one we saw this morning!)

(Cartesian) lenses are optics

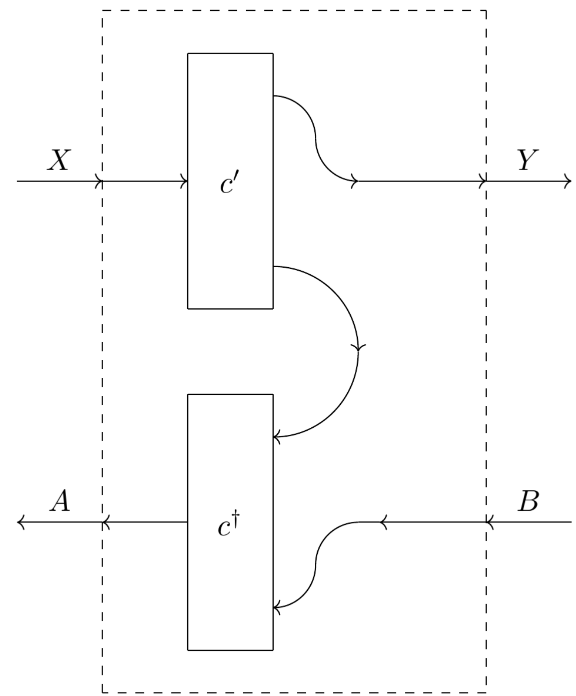
Lens



$$\mathcal{C}(X, Y) \times \mathcal{D}(X \odot B, A)$$

$\cong$

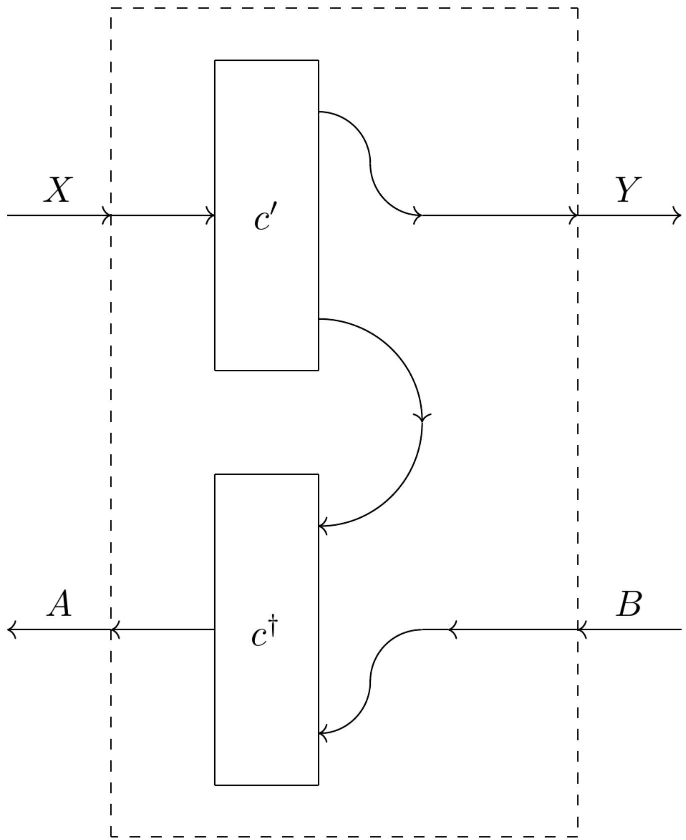
Optic



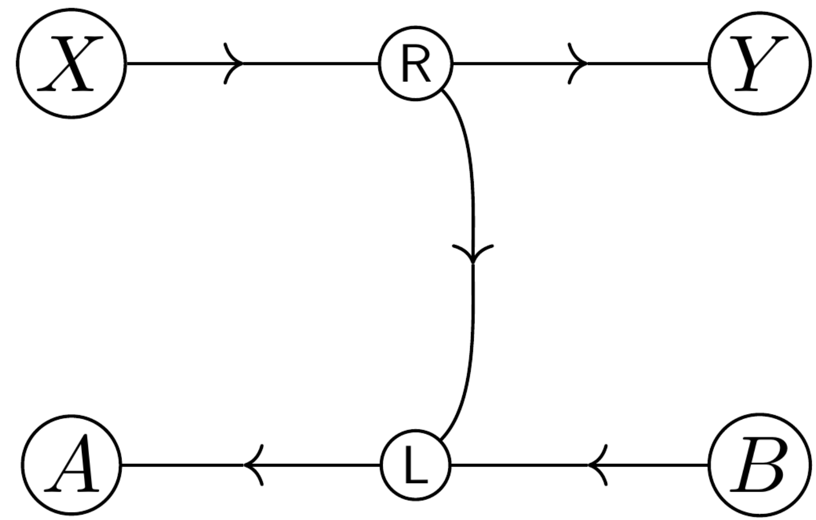
$$\mathbf{Optic}_{\times,\odot}\Big((X, A), (Y, B)\Big)$$

$$\cong \int^{M\,:\,\mathcal{C}} \mathcal{C}(X, M \times Y) \times \mathcal{D}(M \odot B, A)$$

Elements of objects, graphically



$$\in$$



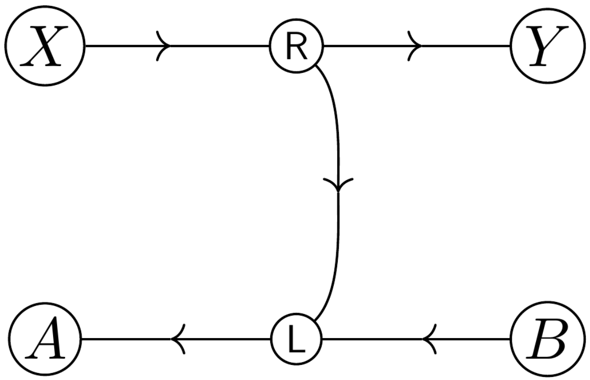
$$\left\langle c' \,\middle|\, c^{\dagger} \right\rangle \in \mathbf{Optic}_{®,Ⓛ}\Big((X, A), (Y, B)\Big)$$

$$\cong \int^{M\,:\,\mathcal{M}} \mathcal{C}(X, M ® Y) \times \mathcal{D}(M Ⓛ B, A)$$

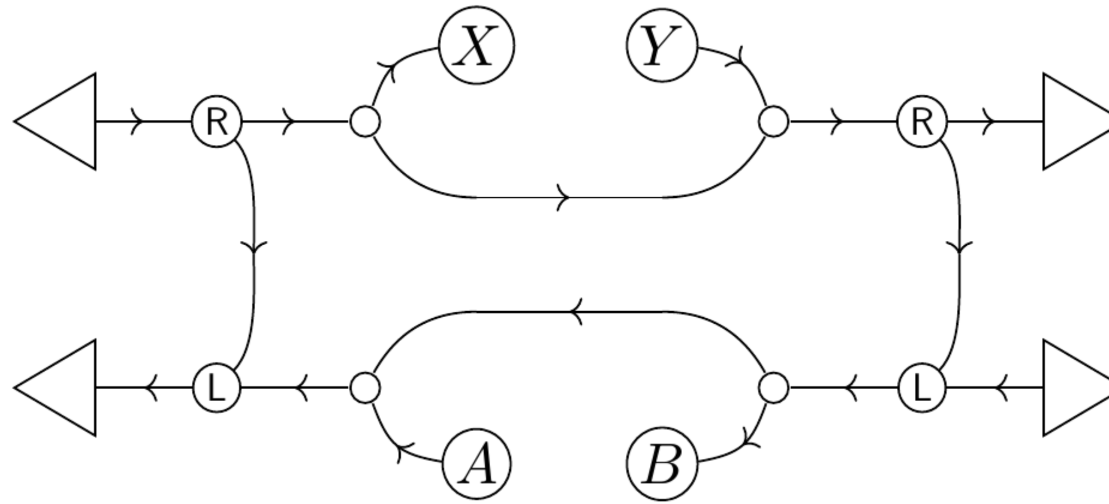
"inflate the tubes"

Contexts: closed environments "with a hole in them"

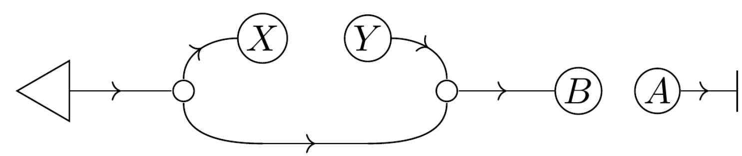
<u>Optic</u>:



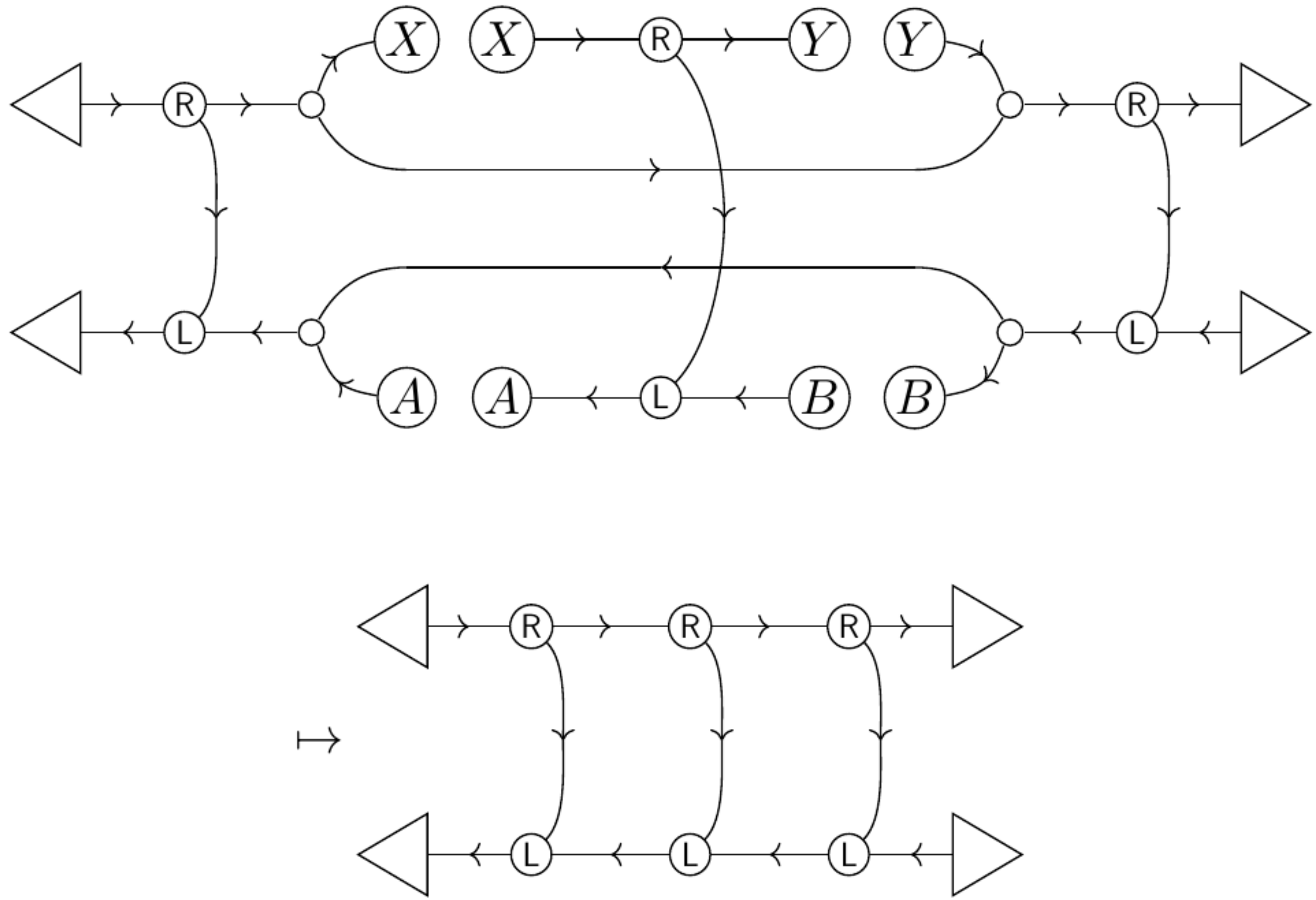
<u>Context</u>:



When monoidal units are terminal, this simplifies to:

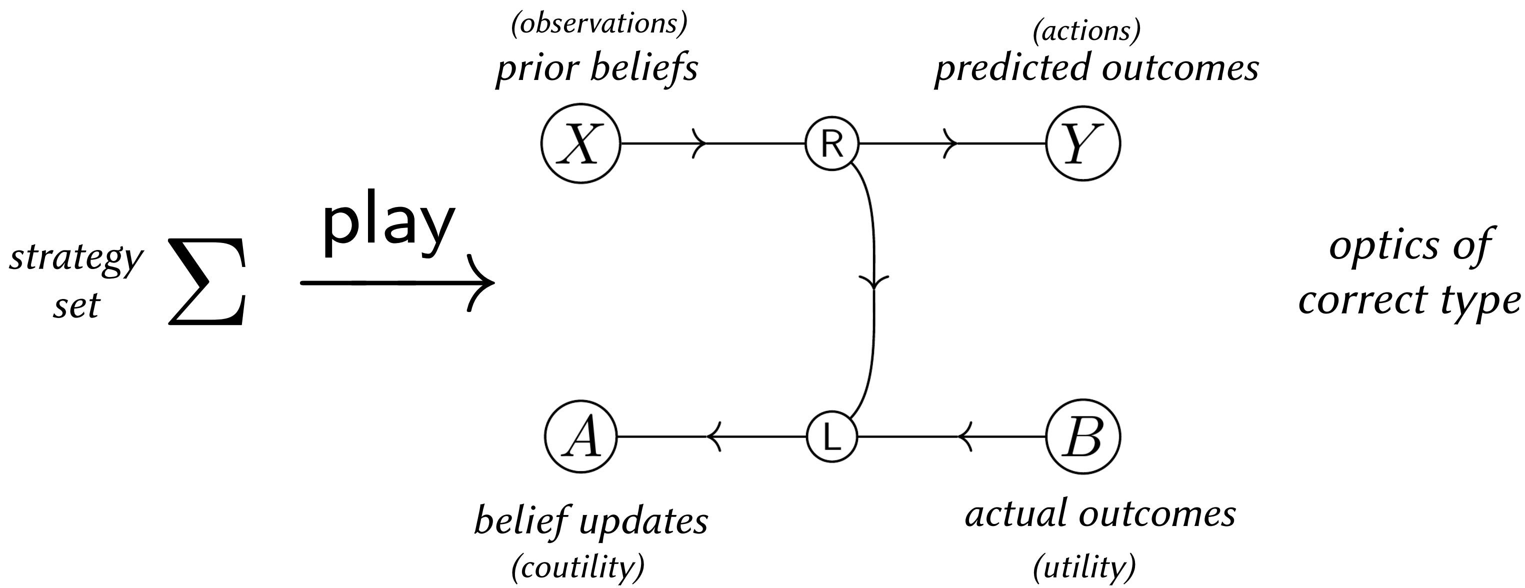


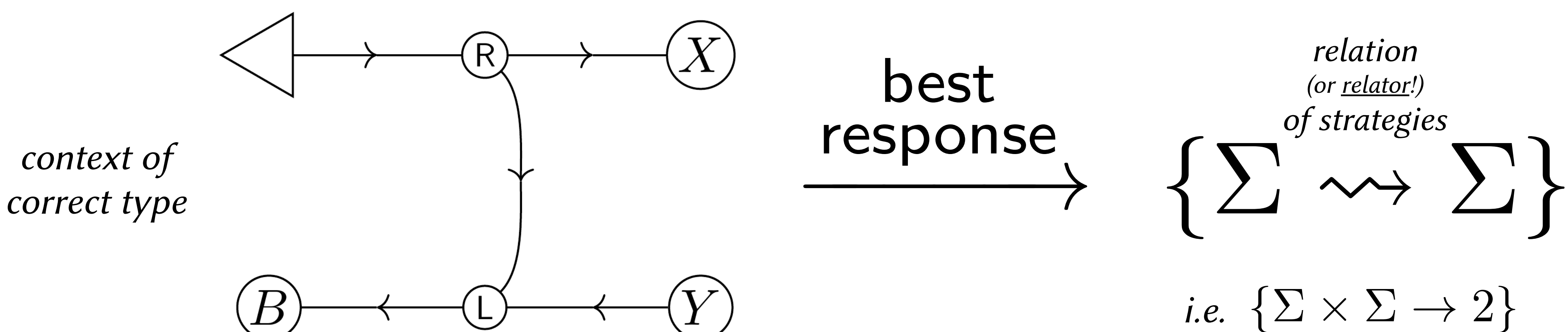
Open system in context is closed



<u>Now</u>: primer on open games ...

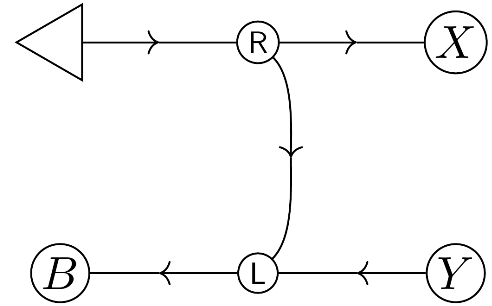
A game $G : (X, A) \xrightarrow{\Sigma} (Y, B)$ constitutes :





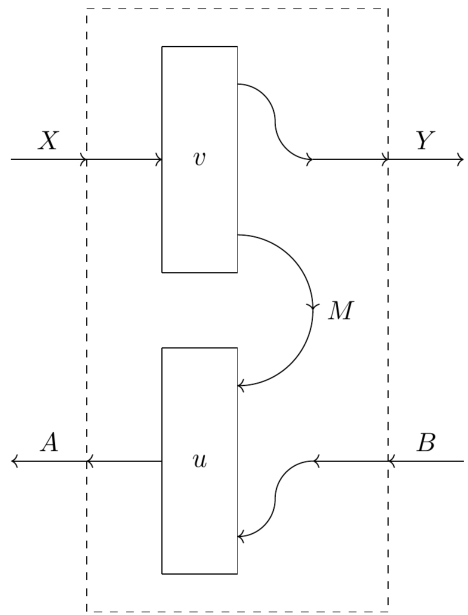
Best response, demystified

suppose context $\langle \pi \mid k \rangle \in$

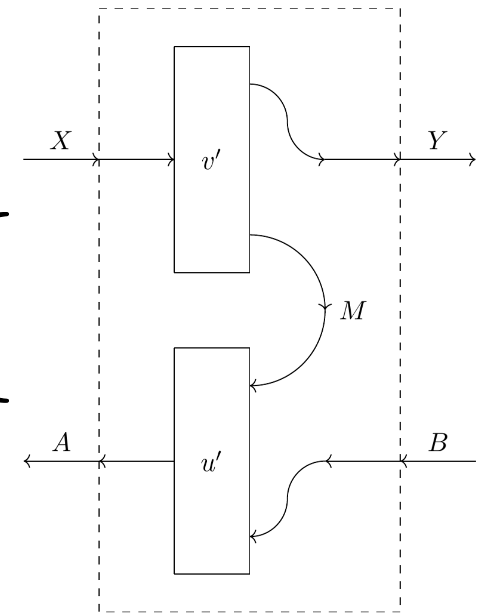


Current strategy | **Better strategies**

best response$\big(\langle \pi \mid k \rangle\big) =$



$\mapsto \Bigg\{$



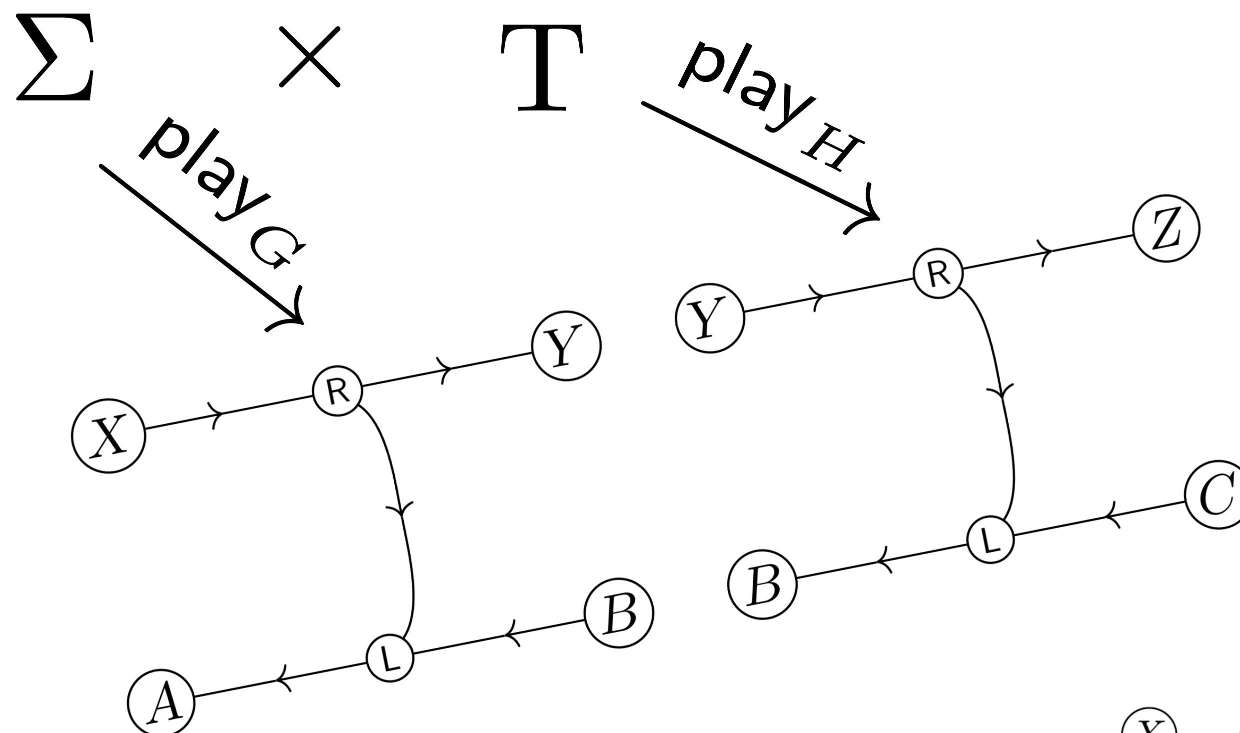
$\Bigg\}$

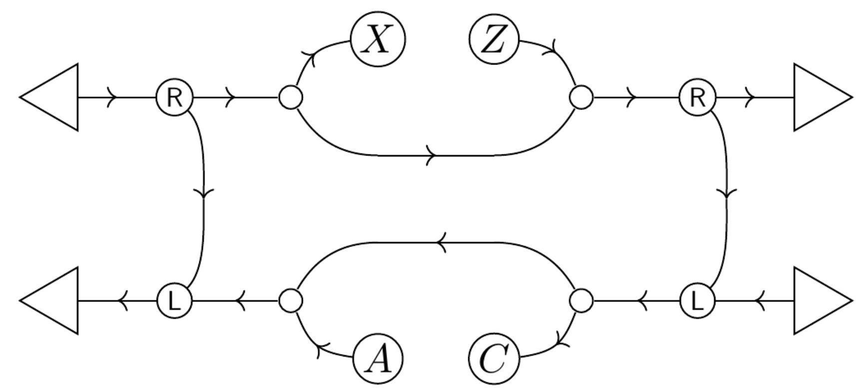
$\langle v \mid u \rangle_\sigma$ $\qquad$ $\langle v' \mid u' \rangle_\tau$

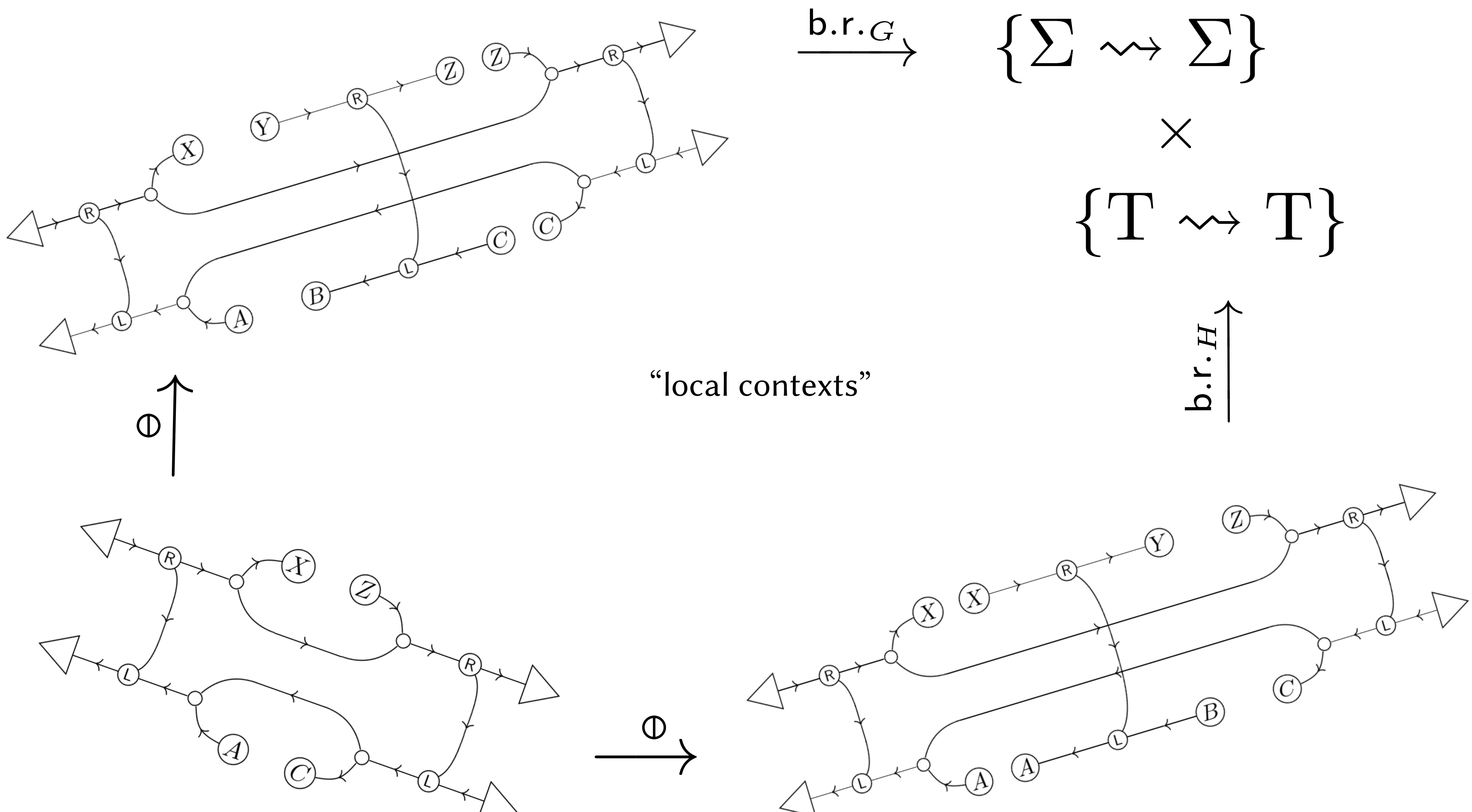
$\sigma \in \Sigma \quad \mapsto \quad \{\tau\} \subseteq \Sigma$

(but how do agents learn to deviate? ...)

$$H \mathbin{\obar} G : (X, A) \xrightarrow{\Sigma} (Y, B) \xrightarrow{\mathrm{T}} (Z, C)$$



$\mathsf{ctx}(H \mathbin{\obar} G) =$ 


b.r.$_G$
$\{\Sigma \rightsquigarrow \Sigma\}$
$\times$
$\{\mathrm{T} \rightsquigarrow \mathrm{T}\}$
b.r.$_H$
“local contexts”
Φ
Φ
X
Y
Z
A
B
C
R
L

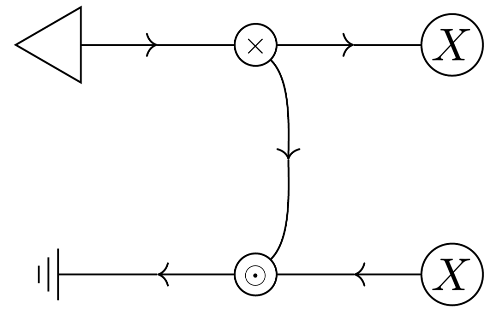
Now we can start to construct some
“atomic” cybernetic systems !

Maximum likelihood game $(I, I) \to (X, X)$

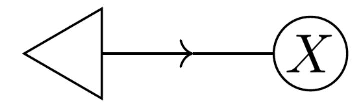
Aim find state π that 'best explains' the data observed through k

Play

$$\{\langle \rho \mid ! \rangle_\sigma , \langle \pi \mid ! \rangle_\tau , \ldots \} \subset$$



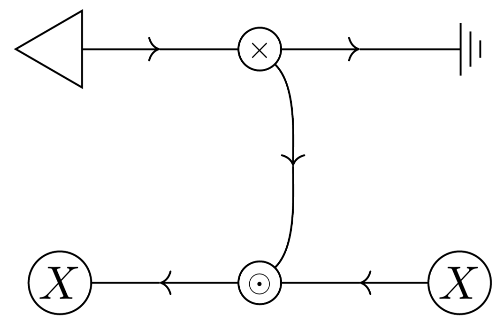
$\cong$



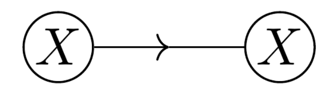
$\cong \mathcal{P}X$

Context

$$\langle ! \mid k \rangle \in$$



$\cong$



$\cong \mathcal{K}\ell(\mathcal{P})(X, X)$

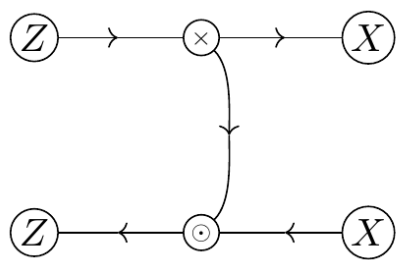
Best response

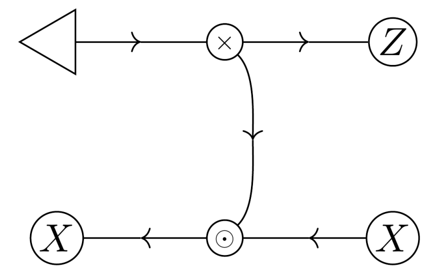
$$B(\langle ! \mid k \rangle) = \langle \rho \mid ! \rangle_\sigma \mapsto \left\{ \langle \pi \mid ! \rangle_\tau \middle| \pi \in \arg\max_{\pi : I \bullet\!\!\rightarrow X} \mathop{\mathbb{E}}_{k \bullet \pi} [\pi] \right\}$$

Bayesian inference game $(Z, Z) \to (X, X)$

Fix a channel $c : Z \to\!\!\!\!\!\bullet\; X$

Aim: find state-dependent channel $c' : Z \odot X \to Z$
closest to exact inversion of c (in the context)





$$B(\langle \pi \mid k \rangle)$$

$$= \langle d \mid d' \rangle_\sigma \mapsto \left\{ \langle c \mid c' \rangle_\tau \,\middle|\, c' \in \underset{c' : \mathbf{Meas}(\mathcal{P}Z, \mathcal{K}\ell(\mathcal{P})(X,Z))}{\arg\min} \underset{x \sim k \bullet c \bullet \pi}{\mathbb{E}} \left[D_{KL}\big(c'_\pi(x), c^\dagger_\pi(x)\big) \right] \right\}$$

$$= \langle d \mid d' \rangle_\sigma \mapsto \left\{ \langle c \mid c' \rangle_\tau \,\middle|\, c' \in \underset{c' : \mathbf{Meas}(\mathcal{P}Z, \mathcal{K}\ell(\mathcal{P})(X,Z))}{\arg\min} \underset{x \sim k \bullet c \bullet \pi}{\mathbb{E}} \left[\underset{z \sim c'_\pi(x)}{\mathbb{E}} [-\log p_c(x|z)] + D_{KL}(c'_\pi(x), \pi) \right] \right\}$$

Proposition: Bayesian inference games are closed under composition

Proof: Bayesian updates compose optically

Autoencoder game $(Z, Z) \to (X, X)$

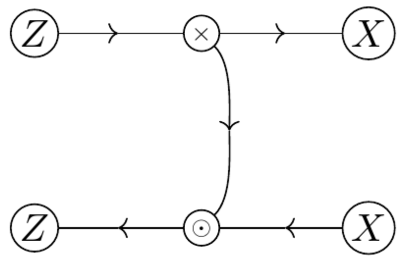
Fix:

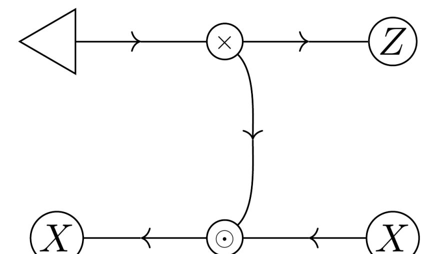
“Generative” models: $\Gamma \hookrightarrow \mathcal{K}\ell(\mathcal{P})(Z, X)$

“Recognition” models: $\mathsf{P} \hookrightarrow \mathcal{K}\ell(\mathcal{P})(X, Z)$

Aim:

find pair (c, c') such that $c \bullet \pi$ maximizes the likelihood of data from k, and c' best approximates the exact inverse of c in the context



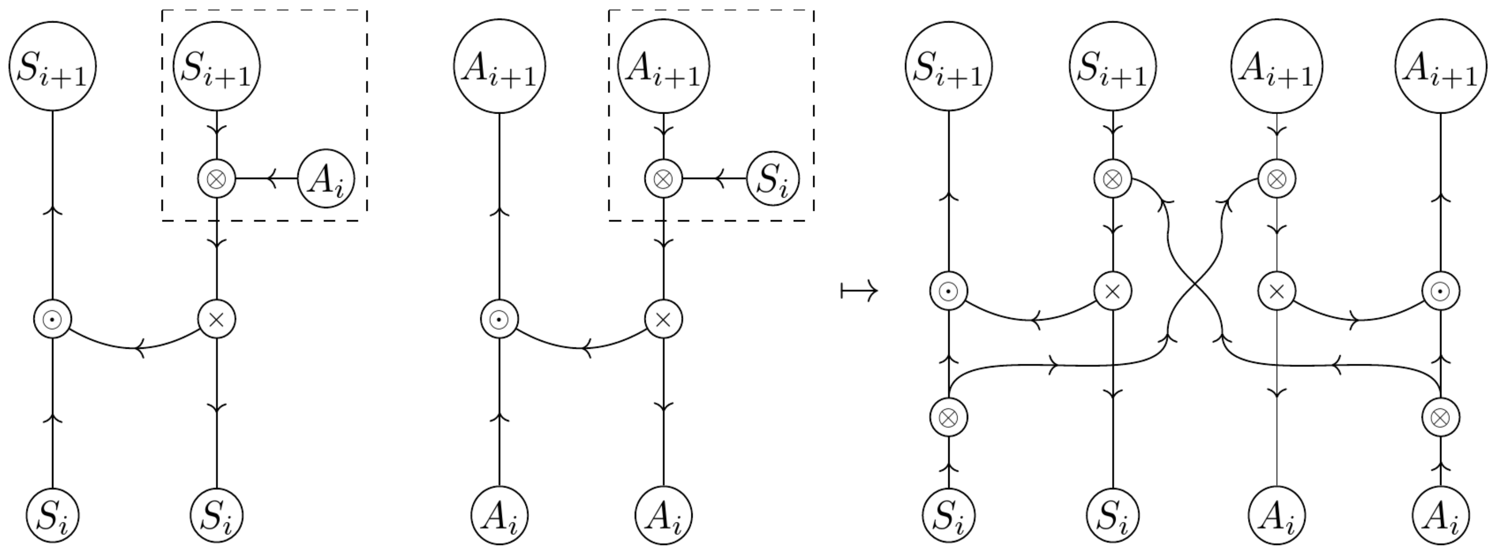


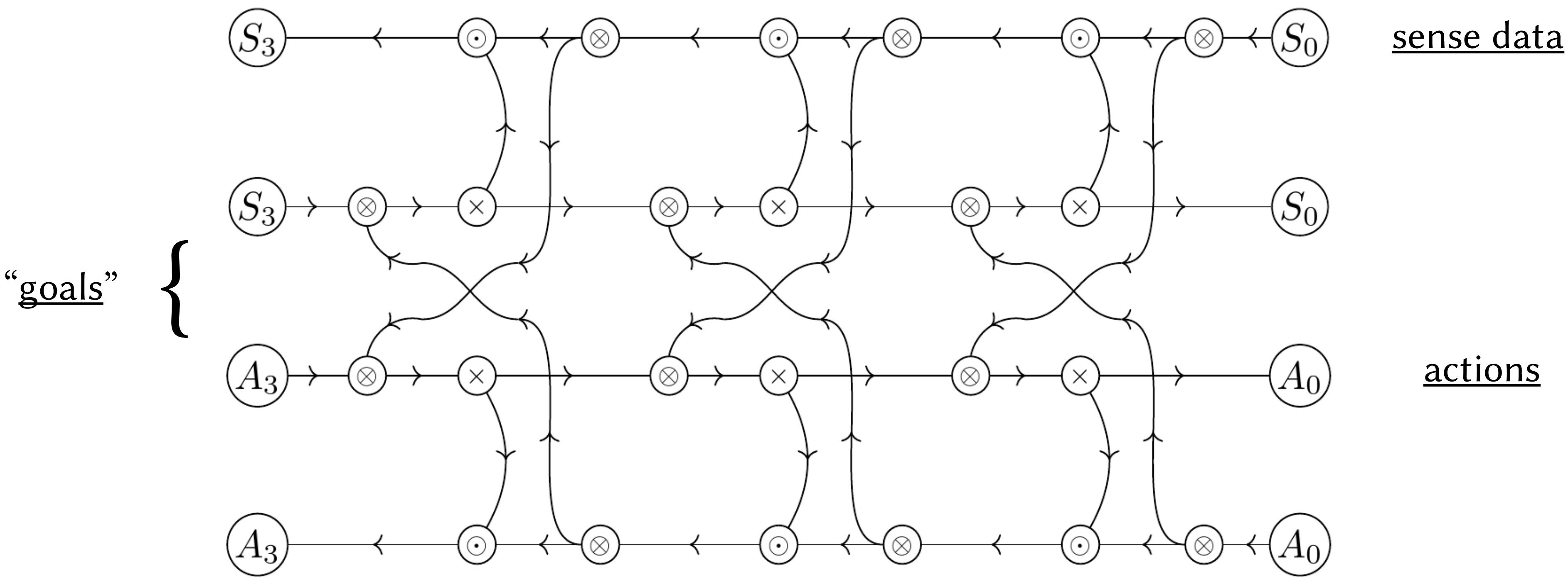
$$B(\langle \pi \mid k \rangle) =$$

$$\langle d \mid d' \rangle_\sigma \mapsto \left\{ \langle c \mid c' \rangle_\tau \,\middle|\, (c, c') \in \underset{\substack{c \in \Gamma, \\ c' \in \mathbf{Meas}(\mathcal{P}Z, \mathrm{P})}}{\arg\min} \; \underset{x \sim k \bullet c \bullet \pi}{\mathbb{E}} \left[\underset{z \sim c'_\pi(x)}{\mathbb{E}} [-\log p_c(x|z)] + D(c'_\pi(x), \pi) \right] \right\}$$

– this objective captures many such models in the ML literature (Knoblauch *et al*, 2019)

“Active inference” game





S_3
S_0
sense data
S_3
S_0
"goals"
A_3
A_0
actions
A_3
A_0

Optimization games

MLE:

$$B(\langle ! \mid k \rangle) = \langle \rho \mid ! \rangle_\sigma \mapsto \left\{ \langle \pi \mid ! \rangle_\tau \middle| \pi \in \underset{\pi : I \to X}{\arg\max} \underset{k \bullet \pi}{\mathbb{E}} [\pi] \right\}$$

Inference:

$$B(\langle \pi \mid k \rangle) = \langle d \mid d' \rangle_\sigma \mapsto \left\{ \langle c \mid c' \rangle_\tau \middle| c' \in \underset{c' : \mathbf{Meas}(\mathcal{P}Z, \mathcal{K}\ell(\mathcal{P})(X,Z))}{\arg\min} \underset{x \sim k \bullet c \bullet \pi}{\mathbb{E}} \left[D_{KL}\big(c'_\pi(x), c^\dagger_\pi(x)\big) \right] \right\}$$

$$= \langle d \mid d' \rangle_\sigma \mapsto \left\{ \langle c \mid c' \rangle_\tau \middle| c' \in \underset{c' : \mathbf{Meas}(\mathcal{P}Z, \mathcal{K}\ell(\mathcal{P})(X,Z))}{\arg\min} \underset{x \sim k \bullet c \bullet \pi}{\mathbb{E}} \left[\underset{z \sim c'_\pi(x)}{\mathbb{E}} [-\log p_c(x|z)] + D_{KL}(c'_\pi(x), \pi) \right] \right\}$$

Autoencoder:

$$B(\langle \pi \mid k \rangle) = \langle d \mid d' \rangle_\sigma \mapsto \left\{ \langle c \mid c' \rangle_\tau \middle| (c, c') \in \underset{\substack{c \in \Gamma, \\ c' \in \mathbf{Meas}(\mathcal{P}Z, \mathrm{P})}}{\arg\min} \underset{x \sim k \bullet c \bullet \pi}{\mathbb{E}} \left[\underset{z \sim c'_\pi(x)}{\mathbb{E}} [-\log p_c(x|z)] + D(c'_\pi(x), \pi) \right] \right\}$$

All of the form:

$$\varphi_G : \mathsf{ctx} \times \Sigma \to \mathbb{R}$$

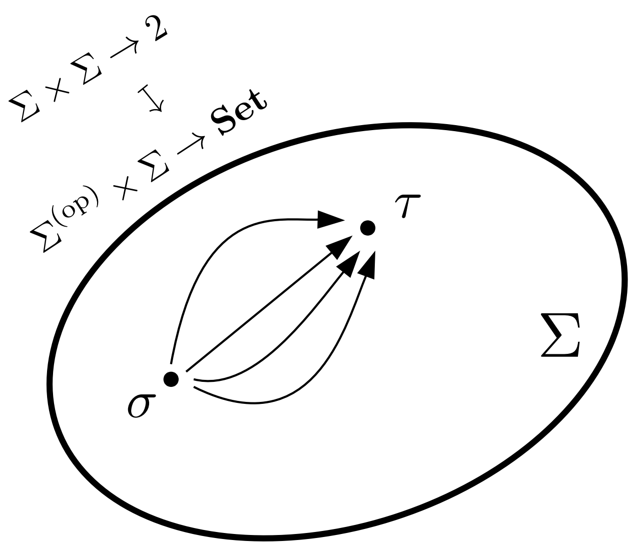
$$B(\langle \pi \mid k \rangle) = \langle d \mid d' \rangle_\sigma \mapsto \left\{ \langle c \mid c' \rangle_\tau \middle| (c, c') \in \arg\max \varphi_G\big(\langle \pi \mid k \rangle, \tau \big) \right\}$$

But how to "get better"?..

Note: $\varphi^{\sharp}_{G} : \mathsf{ctx} \to \Sigma \to \mathbb{R}$ *Given a context, obtain a "fitness landscape" or "potential field" over the strategy space*

Can we categorify best-response relations, to make them *proof-relevant* ?

Then: strategic deviation (improvement) witnessed by trajectory / process



Can we characterize this process compositionally?

And: don't we act on "story snippets"?

Discrete-time dynamical systems

$$\mathbf{Dyn}_{\mathcal{C}}(A, B) = \sum_{S:\mathcal{C}} \mathbf{Comon}(\mathcal{C})(S, B) \times \mathcal{C}(S \otimes A, S)$$

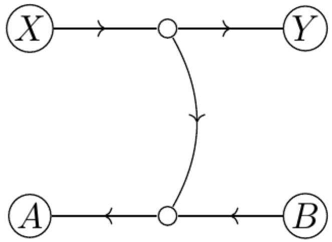
$$f : A \xrightarrow{S} B \;=\; (S, f^{out} : S \to B, f^{upd} : S \otimes A \to\!\!\!\!\bullet\; S)$$

$$\mathsf{id}_A : A \xrightarrow{A} A \;=\; (A, \mathsf{id}_A : A \to A, \pi_2 : A \otimes A \to\!\!\!\!\bullet\; A)$$

Composition: "wire" outputs to inputs, using lenses

Dynamical lenses & contexts

Since dynamical systems form a monoidal category, we can construct lenses over them

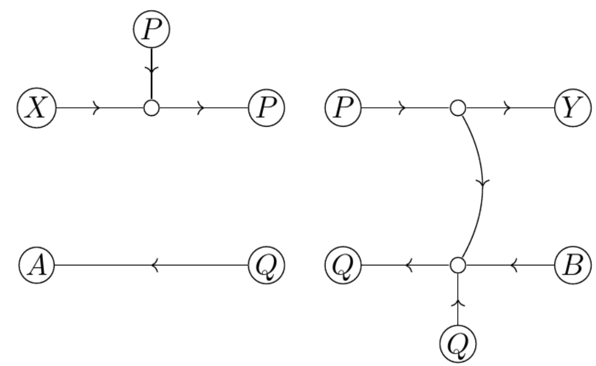


$$\int^{M:\mathcal{C}} \mathbf{Dyn}_{\mathcal{C}}(X, M \otimes Y) \times \mathbf{Dyn}_{\mathcal{C}}(M \otimes B, A)$$

$$\cong$$

$$\sum_{P,Q:\mathcal{C}} \int^{M:\mathcal{C}} \mathcal{C}(P \otimes X, P) \times \mathbf{Comon}(\mathcal{C})(P, M \otimes Y) \times \mathcal{C}(Q \otimes M \otimes B, Q) \times \mathbf{Comon}(\mathcal{C})(Q, A)$$

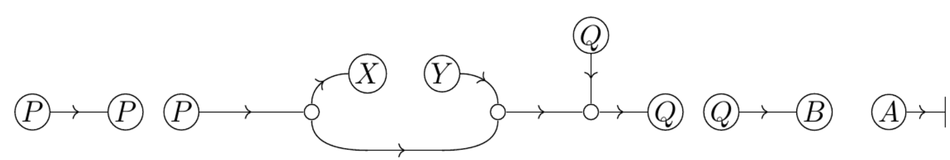
$$\sum_{P,Q:\mathcal{C}}$$



Dynamical lens: just a pair of (coupled) dynamical systems

When the monoidal unit is terminal, dynamical contexts are of the type

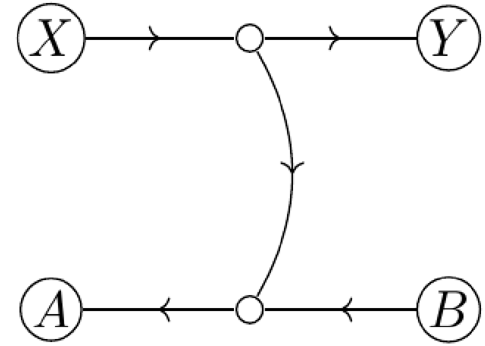
$$\sum_{P,Q:\mathcal{C}}$$



Dynamical games

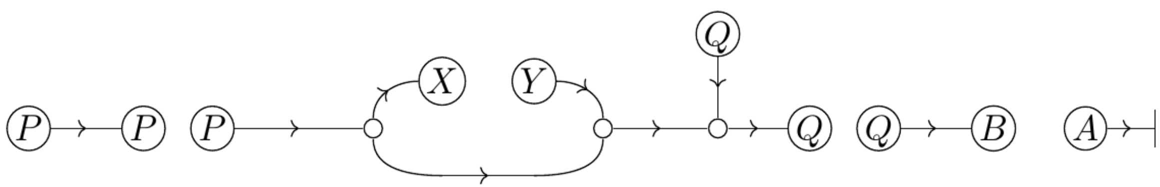
(Just open games over dynamical lenses !)

$\tilde{\Sigma} \xrightarrow{\text{play}}$



$$\int^{M:\mathcal{C}} \mathbf{Dyn}_{\mathcal{C}}(X, M \otimes Y) \times \mathbf{Dyn}_{\mathcal{C}}(M \otimes B, A)$$

$\sum_{P,Q\,:\,\mathcal{C}}$



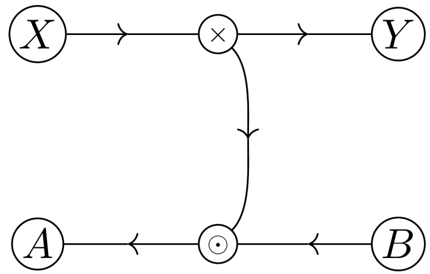
$\xrightarrow{\text{best response}} \{\tilde{\Sigma} \rightsquigarrow \tilde{\Sigma}\}$

Strategies here can be thought of as possible learning algorithms
– won't discuss this more today

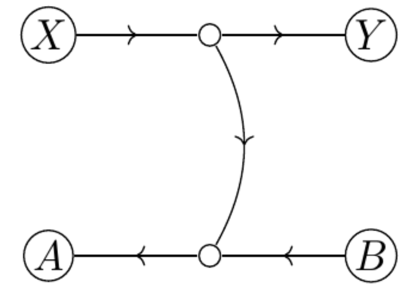
Time to start joining the dots...

Dynamical realisation

On plays:



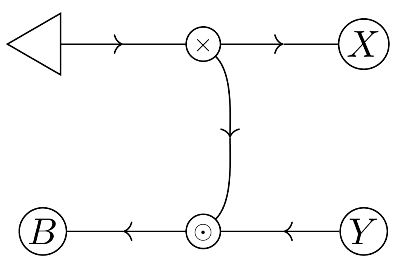
$\xrightarrow{[\![\cdot]\!]}$



$$\int^{\hat{M}:\hat{\mathcal{C}}} \hat{\mathcal{C}}(\hat{X}, \hat{M} \times \hat{Y}) \times \check{\mathcal{C}}(\hat{M} \odot \check{B}, \check{A})$$

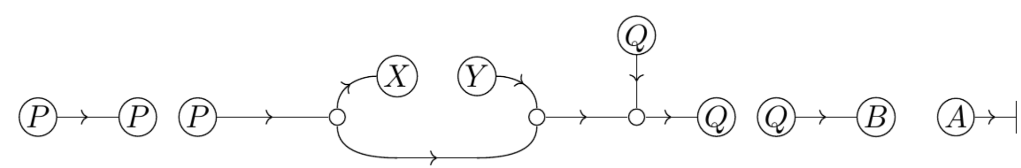
$$\int^{M:\mathcal{C}} \mathbf{Dyn}_{\mathcal{C}}(X, M \otimes Y) \times \mathbf{Dyn}_{\mathcal{C}}(M \otimes B, A)$$

On contexts:



$\xrightarrow{[\![\cdot]\!]}$

$\sum_{P,Q:\mathcal{C}}$



NB: Here, we only have "static" contexts, hence constant realisations

Open cybernetic systems (preliminary definition !)

An **open cybernetic system** G constitutes

- a ('static') optimization game such that
- the fitness function factors through some optimization objective

$$\varphi_G : \mathsf{ctx} \to \Sigma \xrightarrow{\varphi_{(\pi,k)}} \mathbb{R}$$

e.g. $\varphi_{(\pi,k)}(c,c') = - \mathop{\mathbb{E}}_{x \sim k \bullet c \bullet \pi} \left[D_{KL}\left(c'_\pi(x), c^\dagger_\pi(x)\right)\right]$

- subject to a coherence condition – roughly, that

letting the state space of the closed system $[\![\langle \pi \mid k \rangle]\!] \circ \mathsf{play}(\tilde{\sigma})$ be S

we can project from the state space to the optimization space $\mathsf{proj} : S \to \Sigma$

then: $\exists$ fixed point $\zeta^* \in S$ such that $\mathsf{proj}(\zeta^*) \in \arg\max \varphi_{(\pi,k)}$

(Non-stationary contexts? Perhaps: stationary context in a dynamic coordinate system)

Conjecture. Open cybernetic systems form a category

(i.e., fixed point of composite realisation satisfies the cybernetic condition)

Time for some examples ...

Variational autoencoders constitute a category of cybernetic systems

Recall the best-response objective: (here, using Kullback-Leibler divergence)

$$\underset{\substack{c \in \Gamma, \\ c' \in \mathbf{Meas}(\mathcal{P}Z, \mathrm{P})}}{\arg\min} \quad \varphi_{(\pi,k)}(c, c') = \underset{x \sim k \bullet c \bullet \pi}{\mathbb{E}} \; \underset{z \sim c'_\pi(x)}{\mathbb{E}} \left[\log p_{c'_\pi(x)}(z|x) - \log p_c(x|z) - \log p_\pi(z) \right]$$

Define parameterized channels:

"generative"

$$\mathbb{R}^n \cong \Gamma \hookrightarrow \mathbf{Meas}(Z, \mathcal{P}X)$$

$$\vartheta : \mathbb{R}^n \mapsto \gamma^{(\vartheta)} : Z \to \mathcal{P}X$$

"recognition"

$$\mathbb{R}^m \cong \mathrm{P} \hookrightarrow \mathbf{Meas}(X, \mathcal{P}Z)$$

$$\psi : \mathbb{R}^m \mapsto \rho_\pi^{(\psi)} : X \to \mathcal{P}Z$$

Assume no dependence on "action": $k = \kappa \bullet !$

so $$\varphi_{(\pi,k)}(\vartheta, \psi) = \underset{x \sim \kappa}{\mathbb{E}} \; \underset{z \sim \rho_\pi^{(\psi)}(x)}{\mathbb{E}} \left[\log p_{\rho_\pi^{(\psi)}(x)}(z|x) - \log p_{\gamma^{(\vartheta)}}(x|z) - \log p_\pi(z) \right]$$

Then, dynamics realizes gradient descent on the objective...

(but what about those expectations..?)

Assume:

$$z \sim \rho_{\pi}^{(\psi)}(x) \iff z = g\left(\psi, x, r\right)$$

g deterministic, differentiable

$r \sim \sigma_{\pi}(x)$

$r \perp \psi$

So that:

$$\varphi_{(\pi,k)}(\vartheta, \psi) = \mathop{\mathbb{E}}_{x \sim \kappa} \mathop{\mathbb{E}}_{r \sim \sigma_{\pi}(x)} \left[\log p_{\rho_{\pi}^{(\psi)}(x)}\left(g(\psi, x, r) | x\right) - \log p_{\gamma^{(\vartheta)}}\left(x | g(\psi, x, r)\right) - \log p_{\pi}\left(g(\psi, x, r)\right) \right]$$

Then:

$$\begin{aligned}
\nabla_{\psi} \varphi_{(\pi,k)}(\vartheta, \psi) &= \nabla_{\psi} \mathop{\mathbb{E}}_{x \sim \kappa} \mathop{\mathbb{E}}_{r \sim \sigma_{\pi}(x)} \left[\log p_{\rho_{\pi}^{(\psi)}(x)}\left(g(\psi, x, r) | x\right) - \log p_{\gamma^{(\vartheta)}}\left(x | g(\psi, x, r)\right) - \log p_{\pi}\left(g(\psi, x, r)\right) \right] \\
&= \mathop{\mathbb{E}}_{x \sim \kappa} \mathop{\mathbb{E}}_{r \sim \sigma_{\pi}(x)} \left[\nabla_{\psi} \log p_{\rho_{\pi}^{(\psi)}(x)}\left(g(\psi, x, r) | x\right) - \nabla_{\psi} \log p_{\gamma^{(\vartheta)}}\left(x | g(\psi, x, r)\right) - \nabla_{\psi} \log p_{\pi}\left(g(\psi, x, r)\right) \right]
\end{aligned}$$

$$\begin{aligned}
\nabla_{\vartheta} \varphi_{(\pi,k)}(\vartheta, \psi) &= \nabla_{\vartheta} \mathop{\mathbb{E}}_{x \sim \kappa} \mathop{\mathbb{E}}_{r \sim \sigma_{\pi}(x)} \left[\log p_{\rho_{\pi}^{(\psi)}(x)}\left(g(\psi, x, r) | x\right) - \log p_{\gamma^{(\vartheta)}}\left(x | g(\psi, x, r)\right) - \log p_{\pi}\left(g(\psi, x, r)\right) \right] \\
&= \mathop{\mathbb{E}}_{x \sim \kappa} \mathop{\mathbb{E}}_{r \sim \sigma_{\pi}(x)} \left[-\nabla_{\vartheta} \log p_{\gamma^{(\vartheta)}}\left(x | g(\psi, x, r)\right) \right]
\end{aligned}$$

Sketch of the dynamical system

Input $\pi : \mathcal{P}Z; \;\; x \sim \kappa : \mathcal{P}X$

Update $\vartheta(t+1) = \vartheta(t) - \mathbb{E}_{r\sim\sigma_\pi(x)} \left[\nabla_\vartheta \log p_{\gamma^{(\vartheta)}}(x|g(\psi,x,r))\right]$

$$\begin{aligned}\psi(t+1) \; &= \psi(t) - \mathop{\mathbb{E}}_{r\sim\sigma_\pi(x)} \Big[\nabla_\psi \log p_{\rho_\pi^{(\psi)}(x)}\left(g(\psi,x,r)|x\right) \\ &\qquad - \nabla_\psi \log p_{\gamma^{(\vartheta)}}(x|g(\psi,x,r)) - \nabla_\psi \log p_\pi\left(g(\psi,x,r)\right)\Big]\end{aligned}$$

NB: *trajectories over strategy space; fixed point at best response*

Output $x' \sim \gamma^{(\vartheta)} \bullet \pi : \mathcal{P}X; \;\; z' \sim \rho_\pi^{(\psi)}(x) : \mathcal{P}Z$

‘**Theorem**’: VAE games form a category of open cybernetic systems (by BUCO)

Corollary: “*deep active inference*” agents are cybernetic systems realizing active inference games

Friston's "free energy framework" defines a category of cybernetic systems

$$\underset{\substack{\gamma\in\Gamma,\\ \rho\in\mathbf{Meas}(\mathcal{P}Z,\mathrm{P})}}{\arg\min} \quad \varphi_{(\pi,k)}(\gamma,\rho) = \underset{x\sim k\bullet\gamma\bullet\pi}{\mathbb{E}}\left[\underset{z\sim\rho_\pi(x)}{\mathbb{E}}\left[\log p_{\rho_\pi(x)}(z|x) - \log p_\gamma(x|z) - \log p_\pi(z)\right]\right]$$

$$= \underset{x\sim k\bullet\gamma\bullet\pi}{\mathbb{E}}\left[\mathrm{H}\left[\rho_\pi(x)\right] + \underset{z\sim\rho_\pi(x)}{\mathbb{E}}\left[E(z,x)\right]\right]$$

where $E(z,x) = -\log p_\gamma(x|z) - \log p_\pi(z)$

This time the realisation won't just be glorified functions, with dynamics on the parameters – rather, we will have dynamics directly on the system's beliefs (as well as param.s)

Key assumption: all spaces Euclidean, and all states Gaussian

So each $\gamma(z) : \mathcal{P}X$ and $\rho_\pi(x) : \mathcal{P}Z$ is determined by a pair of vectors

$$\gamma(z) \leftrightarrow \big(\mu_\gamma(z),\ \Sigma_\gamma(z)\big) : \mathbb{R}^{|X|} \times \mathbb{R}^{|X|^2}$$

$$\rho_\pi(x) \leftrightarrow \big(\mu_{\rho_\pi}(x),\ \Sigma_{\rho_\pi}(x)\big) : \mathbb{R}^{|Z|} \times \mathbb{R}^{|Z|^2}$$

We define dynamical systems directly on these vectors

Assume:

$k = \kappa \bullet !$

(means: no dependence on 'action' **on the timescale** of the dynamics)

each $\rho_\pi(x)$ is 'tightly peaked'

(means: density function well approximated by 2nd-order Taylor expansion around mean)

better: **least action**

→ minimize time-integral of free-energy

→ 2nd order ODEs

(so, neater in continuous time)

so the 'fitness function' here is really something like an "open Lagrangian"

So that:

$$\nabla_{\mu_{\rho_\pi}} \varphi_{(\pi,k)}(\gamma, \rho) = \mathbb{E}_{x\sim\kappa}\left[\nabla_{\mu_{\rho_\pi}} E(\mu_{\rho_\pi}, x)\right]$$

$$\nabla_{\mu_{\rho_\pi}} E(\mu_{\rho_\pi}, x) = -\nabla_z \mu_\gamma(\mu_{\rho_\pi})^T {\Sigma_\gamma}^{-1} \epsilon_\gamma + {\Sigma_\pi}^{-1} \epsilon_\pi$$

where $\epsilon_\gamma = x - \mu_\gamma(\mu_{\rho_\pi})$ and $\epsilon_\pi = \mu_{\rho_\pi} - \mu_\pi$

(since Gaussian)

Then:

Suppose $\mu_\gamma : \mathbb{R}^{|Z|} \to \mathbb{R}^{|X|}$ is 'neurally computable'.

Then so is the dynamical system $\mu_{\rho_\pi}(t+1) = \mu_{\rho_\pi}(t) - \nabla_{\mu_{\rho_\pi}} E(\mu_{\rho_\pi}, x)$

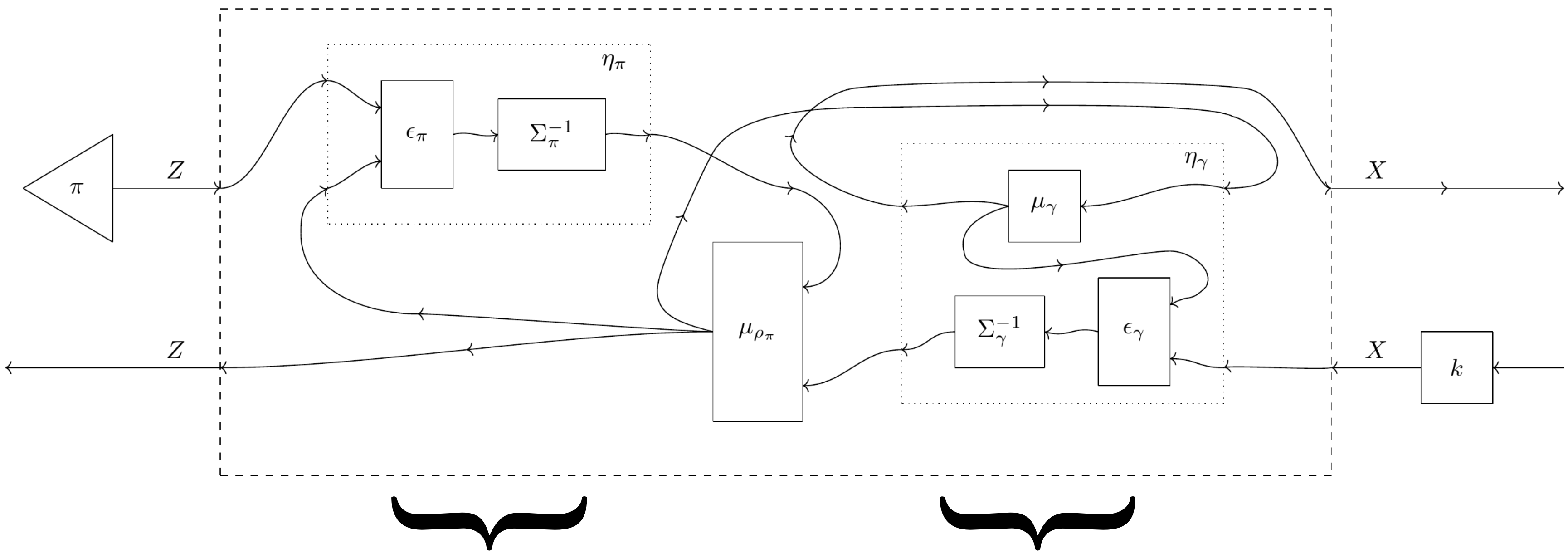
(being a composite of linear and 'neural' maps)

Sketch of the dynamical system

Input $\pi : \mathcal{P}Z; \; x \sim \kappa : \mathcal{P}X$

Update $\mu_{\rho_\pi}(t+1) = \mu_{\rho_\pi}(t) + \nabla_z \mu_\gamma(\mu_{\rho_\pi})^T {\Sigma_\gamma}^{-1} \epsilon_\gamma - {\Sigma_\pi}^{-1} \epsilon_\pi$

Output $z' \sim \rho_\pi(x) : \mathcal{P}Z$



Bidirectional, hierarchical

Cortical communications = precision-weighted prediction errors

adjacent layers are ***local contexts*** *!*

Summary

1. Showed that *Bayesian updates compose optically*
2. Characterized *inference problems* as *open games over Bayesian lenses*
3. *Cybernetic systems* have dynamics governed by best-response objective
4. Example: abstract explanation for the *gross structure of cortical circuits*

On-going work and open problems

- Continuous dynamics:
  more intricate formally, but neater conceptually
  (nice links to classical mechanics!)

- "Truly dynamical" games:
  - trajectories on the interfaces
  - non-stationary contexts
    (ie, dynamics in the base as well as the fibres)
  - nested systems (as in evolution)

- Interacting cybernetic systems:
  - players playing game-theoretic games?
  - link with iterated games?
  - reinforcement learning?

& hopefully end up with a more elegant category of cybernetic systems ...!

Thanks!