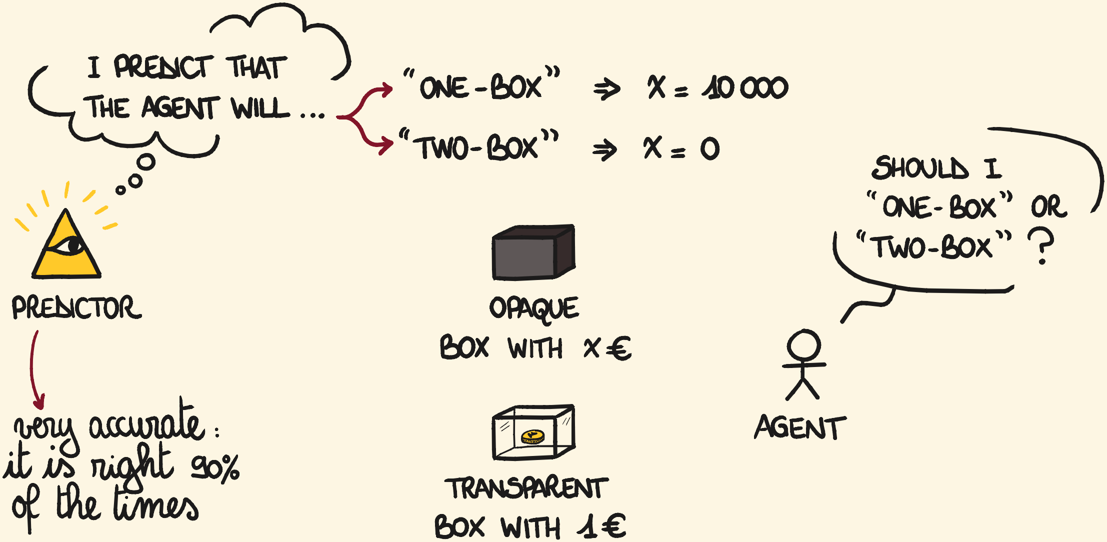
TallCat Seminar 9 March 2023

EVIDENTIAL DECISION THEORY VIA PARTIAL MARKOV CATEGORIES

Elena Di Lavore and Mario Román

Tallinn University of Technology

NEWCOMB'S PROBLEM



(cute drawing inspired by Mario's slides for NWPT '22)

CAUSAL DECISION THEORY

<u>Causal</u> decision theory answers :

"Which action would <u>cause</u> the best-case scenario ?"

Whatever the predictor did,
I get 1€ extra if I two-box
⇒ I will <u>two-box</u>

BEST !

PREDICTOR \ AGENT	ONE-BOX	TWO-BOX
ONE-BOX	10 000 €	10 001€
TWO-BOX	0 €	1 €

EVIDENTIAL DECISION THEORY

Evidential decision theory answers :

"Which action would be evidence for the best-case scenario ?"

My action is evidence for the prediction :
if I one-box I expect 10 000 €,
if I two-box I expect 1 € .
⇒ I will one-box

PREDICTOR \ AGENT	ONE-BOX	TWO-BOX
ONE-BOX	10 000 €	10 001 €
TWO-BOX	0 €	1 €

MOST LIKELY

EVIDENTIAL VS CAUSAL DECISION THEORY

	 CAUSAL DECISION THEORIST	 EVIDENTIAL DECISION THEORIST
EXPECTED UTILITY	0.9 × 1 € + 0.1 × 10 001 € = 1 001 €	0.9 × 10 000 € + 0.1 × 0 € = 9 000 €

MOTIVATION

- lack of a <u>clear syntax</u> to specify the (implicit) assumptions of decision problems
- lack of a <u>calculus</u> for solving them (according to Evidential Decision Theory)

DESIDERATA

1. express probabilistic processes → Markov
2. express constraints → partiality & discreteness
3. explicitly capture implicit assumptions → category theory

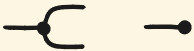
(DISCRETE) PARTIAL MARKOV CATEGORIES

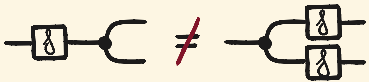
- add partiality to Markov categories

OR

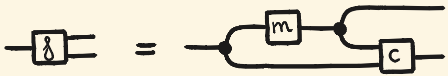
- add probability to (discrete) cartesian restriction categories

COPY - DISCARD STRUCTURE

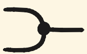


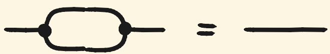


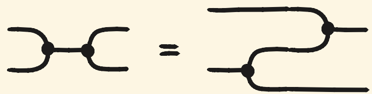
CONDITIONALS



PARTIAL FROBENIUS STRUCTURE







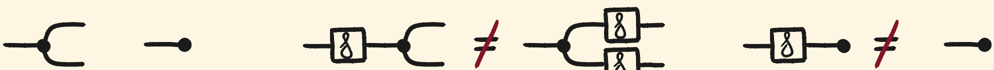
(DISCRETE) PARTIAL MARKOV CATEGORIES

- REMOVE TOTALITY FROM ~~add partiality to~~ Markov categories

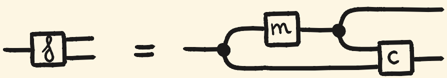
OR

- REMOVE DETERMINISM FROM ~~add probability to~~ (discrete) cartesian restriction categories

COPY-DISCARD STRUCTURE



CONDITIONALS



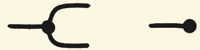
PARTIAL FROBENIUS STRUCTURE

OUTLINE

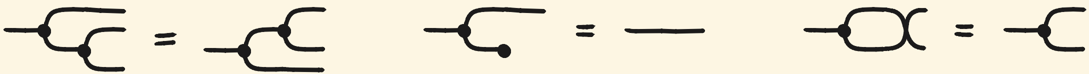
- Partial Markov categories
- Discrete partial Markov categories
- Bayes, Pearl and Jeffrey
- Evidential decision theory
- Exact observations

COPY-DISCARD CATEGORIES

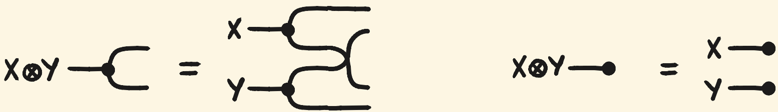
A copy-discard category is a symmetric monoidal category where every object is a uniform cocommutative comonoid.



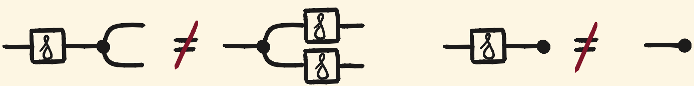
COCOMMUTATIVE COMONOID



UNIFORMITY



NO NATURALITY REQUIRED

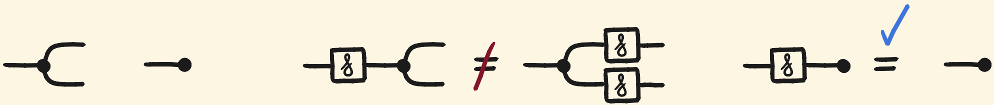


[Corradini & Gadducci 1999]

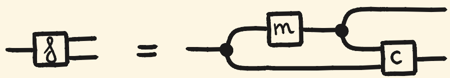
MARKOV CATEGORIES & CONDITIONALS

A Markov category (with conditionals) is a copy-discard category with conditionals where all morphisms are total.

COPY-DISCARD STRUCTURE



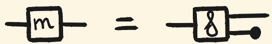
CONDITIONALS



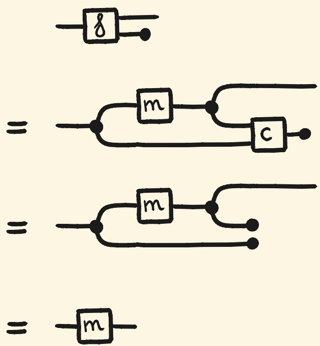
[Fritz 2020]

MARGINALS IN MARKOV CATEGORIES

Marginals in Markov categories are as expected :

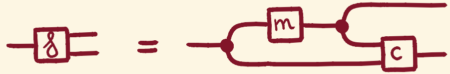


PROOF



conditionals :

$\rightsquigarrow$



$\rightsquigarrow$ totality

□

DROPPING TOTALITY

We want to keep the nice marginals of Markov categories.

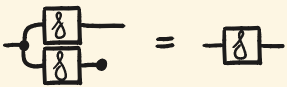
Should we ask conditionals to be total? ✗ NO

⇝ too strong: total conditionals fail to exist in $\mathrm{Kleisli}(\mathcal{D}_{\leq 1})$.

Can we ask conditionals to be quasi-total? ✓ YES

⇝ sweet spot: quasi-total conditionals usually exist and give nice marginals.

QUASI-TOTAL MORPHISM (in a copy-discard category)

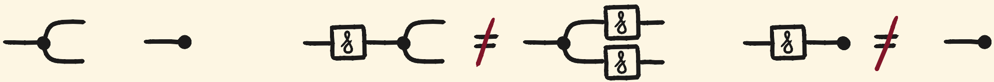


⇝ failure is deterministic

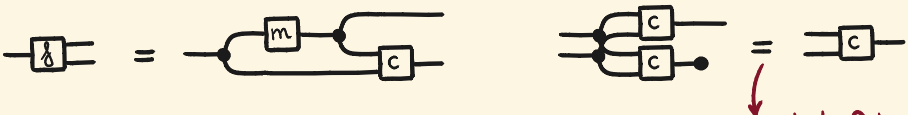
PARTIAL MARKOV CATEGORIES

A partial Markov category is a copy-discard category with quasi-total conditionals.

COPY-DISCARD STRUCTURE



CONDITIONALS



quasi-totality

EXAMPLES : PARTIAL STOCHASTIC PROCESSES

A <u>partial stochastic process</u> is a stochastic process that may fail.

↳ Maybe monad

your favourite Markov category

PROPOSITION

Partial stochastic processes form a partial Markov category.

$\begin{cases} \mathcal{C} \text{ Markov category with coproducts} \\ \text{Some ugly technical conditions} \end{cases}$

$\Rightarrow \mathrm{Kl}(\cdot + 1)$ is a partial Markov category.

EXAMPLES

- $\mathrm{Kl}(\mathcal{D}(\cdot + 1))$ ⇝ finitary subdistributions
- $\mathrm{Kl}(\mathcal{G}iry_B(\cdot + 1))$ ⇝ subdistributions on standard Borel spaces

SUBDISTRIBUTIONS

A <u>subdistribution</u> σ on A is a distribution on $A+1$:

$\sigma \in \mathcal{D}_{\leq 1}(A)$ is a function $\sigma : A \to [0,1]$ such that

- its support, $\mathrm{supp}(\sigma) := \{a \in A \mid \sigma(a) > 0\}$, is finite, and
- its total probability mass is at most 1, $\sum_{a \in A} \sigma(a) \leq 1$.

A morphism $X - \boxed{f} - A$ in $\mathrm{Kl}\mathcal{D}_{\leq 1}$ is a function $X \to \mathcal{D}_{\leq 1}(A)$

$f(a|x)$ = "probability of a given x"

$f(\perp|x)$ = "probability of failure"

Composition is

$$X - \boxed{f} - \boxed{g} - B \; (b|x) := \sum_{a \in A} f(a|x) \cdot g(b|a)$$

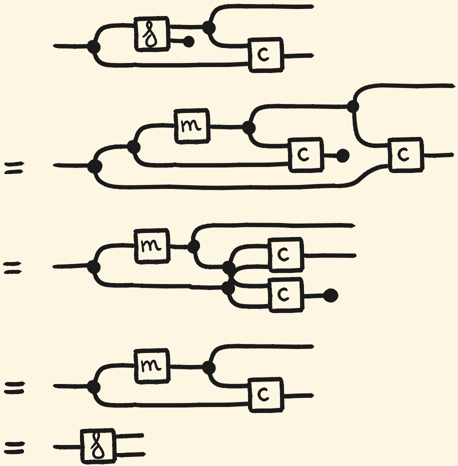
$$X - \boxed{f} - \boxed{g} - B \; (\perp|x) := \sum_{a \in A} f(a|x) \cdot g(\perp|a) + f(\perp|x)$$

MARGINALS OF QUASI-TOTAL CONDITIONALS

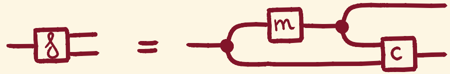
Marginals in partial Markov categories are as expected :

is always a marginal .

PROOF

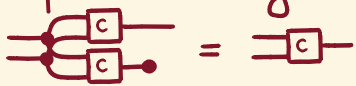


⇝ conditionals



⇝ associativity , commutativity

⇝ quasi-totality



□

CONDITIONALS IN SUBDISTRIBUTIONS

A quasi-total morphism $g : X \to B$ is a function $g : X \to \mathcal{D}B + 1$.

The marginal of $f : X \to A \otimes B$ is

$$X - \boxed{m} - A\,(a|x) = X - \boxed{f} \begin{smallmatrix} - A \\ - \bullet \end{smallmatrix}\,(a|x) = \sum_{b \in B} f(a, b|x)$$

$$X - \boxed{m} - A\,(\perp|x) = X - \boxed{f} \begin{smallmatrix} - A \\ - \bullet \end{smallmatrix}\,(\perp|x) = f(\perp|x)$$

A conditional of f is:

$$\begin{smallmatrix} A \\ X \end{smallmatrix} \boxed{c} - B\,(b|a,x) = \begin{cases} \dfrac{f(a,b|x)}{m(a|x)} & m(a|x) \neq 0 \\ 0 & m(a|x) = 0 \end{cases}$$

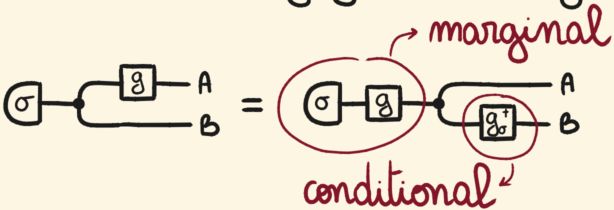
$$\begin{smallmatrix} A \\ X \end{smallmatrix} \boxed{c} - B\,(\perp|a,x) = \begin{cases} 0 & m(a|x) \neq 0 \\ 1 & m(a|x) = 0 \end{cases}$$

BAYES INVERSION

The Bayes inversion of a channel $g : B \to A$ with respect to a distribution $\sigma : I \to B$ is classically defined as

$$g^{\dagger}_{\sigma}(b|a) := \frac{g(a|b)\,\sigma(b)}{\sum_{b' \in B} g(a|b')\,\sigma(b')}$$

In a partial Markov category, it is a $g^{\dagger}_{\sigma} : A \to B$ such that



Bayes inversions are instances of quasi-total conditionals.

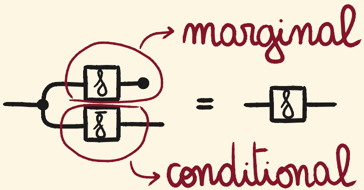
[Cho & Jacobs 2019]

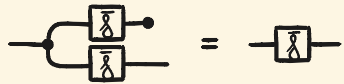
NORMALISATION

The normalisation of a partial channel $f : X \to A$ is classically defined as

$$\bar{f}(a|x) := \frac{f(x|a)}{1 - f(\perp|a)}$$

In a partial Markov category, it is a $\bar{f} : X \to A$ such that





Normalisations are instances of quasi-total conditionals.

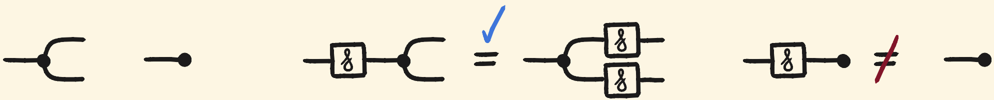
OUTLINE

- Partial Markov categories
- [Discrete partial Markov categories]
- Bayes, Pearl and Jeffrey
- Evidential decision theory
- Exact observations

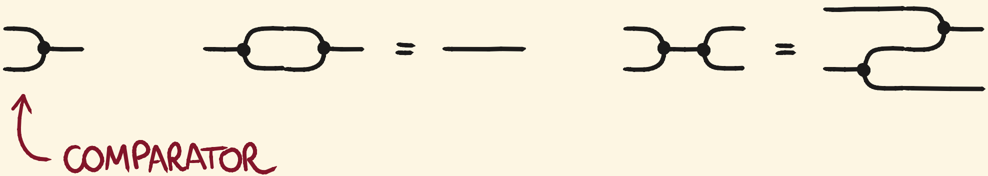
CONSTRAINTS VIA PARTIAL FROBENIUS

A discrete cartesian restriction category is a copy-discard category with comparators where all morphisms are deterministic.

COPY-DISCARD STRUCTURE



PARTIAL FROBENIUS STRUCTURE

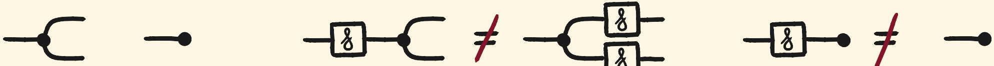


[Cockett & Lack 2003, Cockett, Guo & Hofstra 2012]

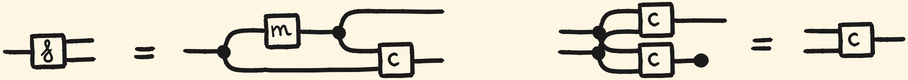
DISCRETE PARTIAL MARKOV CATEGORIES

A discrete partial Markov category is a copy-discard category with quasi-total conditionals and comparators.

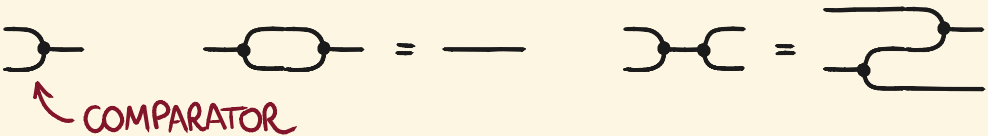
COPY-DISCARD STRUCTURE



CONDITIONALS



PARTIAL FROBENIUS STRUCTURE



EXAMPLES

FINITARY SUBDISTRIBUTIONS

$$\supset\!\!\bullet\!\!-\,(x \mid x_1, x_2) := \begin{cases} 1 & x = x_1 = x_2 \\ 0 & \text{otherwise} \end{cases}$$

$$\supset\!\!\bullet\!\!-\,(\perp \mid x_1, x_2) := \begin{cases} 0 & x_1 = x_2 \\ 1 & \text{otherwise} \end{cases}$$

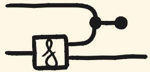
STANDARD BOREL SUBDISTRIBUTIONS

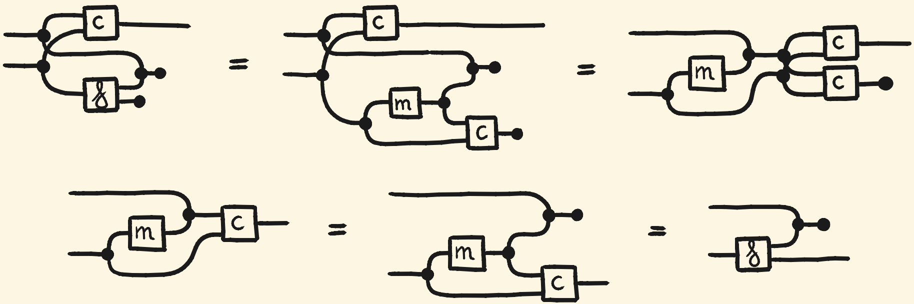
$$\supset\!\!\bullet\!\!-\,(x \mid x_1, x_2) := \begin{cases} 1 & x = x_1 = x_2 \\ 0 & \text{otherwise} \end{cases}$$

⇝ not computable?
bad behaviour?

$$\supset\!\!\bullet\!\!-\,(\perp \mid x_1, x_2) := \begin{cases} 0 & x_1 = x_2 \\ 1 & \text{otherwise} \end{cases}$$

CONDITIONALS BY BENDING WIRES

A quasi-total conditional of  is a normalisation of  .

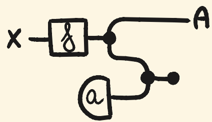


OUTLINE

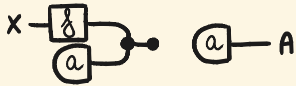
- Partial Markov categories
- Discrete partial Markov categories
- [Bayes, Pearl and Jeffrey]
- Evidential decision theory
- Exact observations

OBSERVATIONS AS CONSTRAINTS

Observing that $f : X \to A$ produced $a \in A$ means keeping only the scenarios in which f produces a.



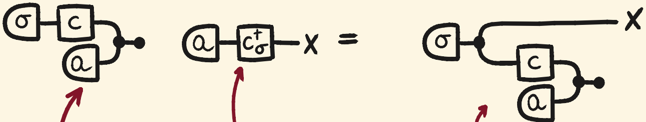
When the observation is deterministic,

$=$ 

.

SYNTHETIC BAYES THEOREM

A deterministic observation $a : I \to A$ from a prior $\sigma : I \to X$ through a channel $c : X \to A$ determines an update proportional to the Bayes inversion $c^{\dagger}_{\sigma}$ evaluated on a.

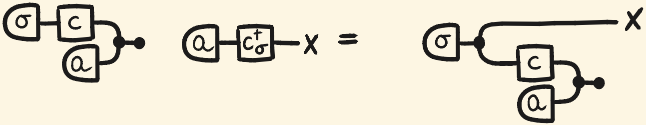


$$P(X = x \mid A = a) = \frac{P(A = a \mid X = x) \cdot P(X = x)}{\sum_{y \in X} P(A = a \mid X = y) \cdot P(X = y)}$$

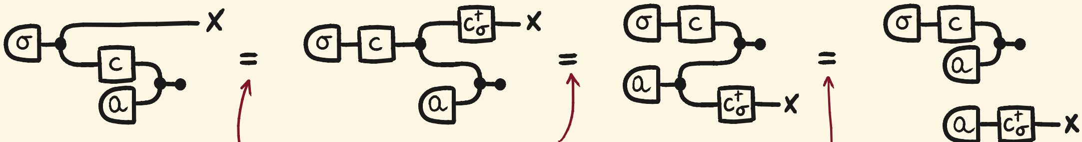
classical formula
for Bayes theorem

SYNTHETIC BAYES THEOREM

A deterministic observation $a : I \to A$ from a prior $\sigma : I \to X$ through a channel $c : X \to A$ determines an update proportional to the Bayes inversion $c^{\dagger}_{\sigma}$ evaluated on a.



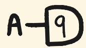
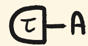
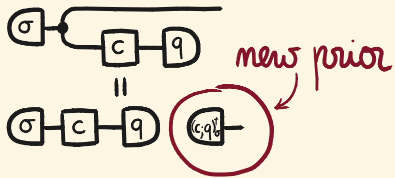
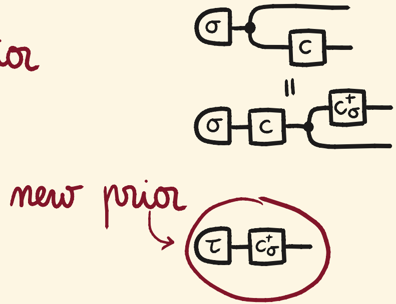
PROOF



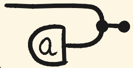
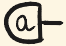
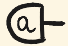
conditionals Frobenius determinism

$\square$

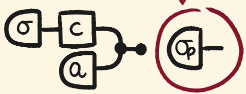
PEARL'S VS JEFFREY'S UPDATES

	PEARL	JEFFREY
PRIOR		
CHANNEL		
EVIDENCE		
UPDATE RULE		

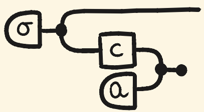
PEARL'S VS JEFFREY'S UPDATES

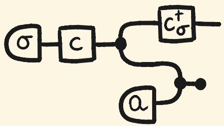
Pearl's update on  coincides with Jeffrey's update on , whenever  is deterministic.

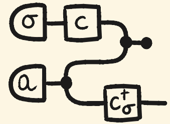
PROOF

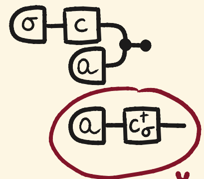


Pearl's updated prior

=  (Pearl's rule)

=  (conditionals)

=  (Frobenius)

=  (determinism)

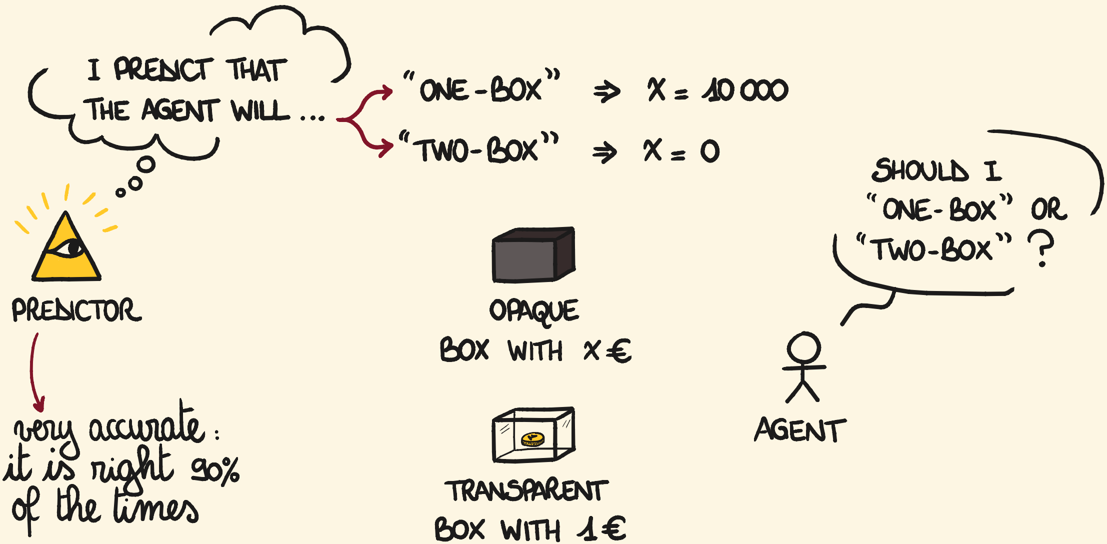
Jeffrey's updated prior

□

OUTLINE

- Partial Markov categories
- Discrete partial Markov categories
- Bayes, Pearl and Jeffrey
- [Evidential decision theory]
- Exact observations

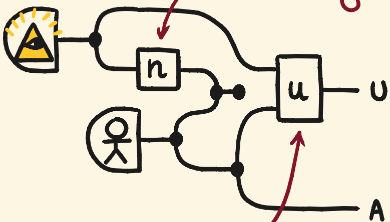
NEWCOMB'S PROBLEM



(cute drawing inspired by Mario's slides for NWPT '22)

NEWCOMB'S PROBLEM CATEGORICALLY

noise: the predictor
is right 90% of the times

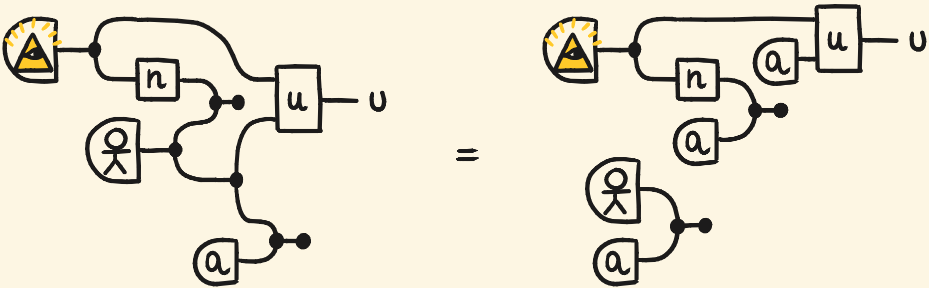


utility function

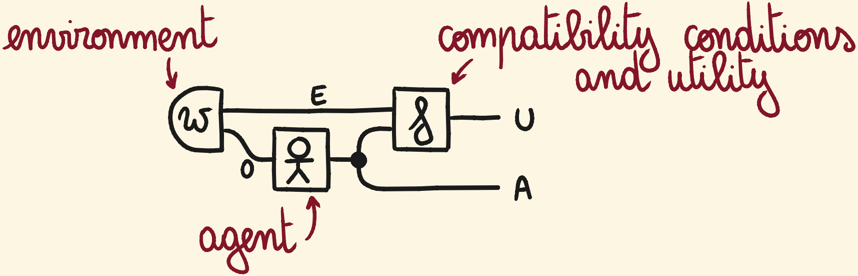
AGENT / PREDICTOR	ONE-BOX	TWO-BOX
ONE-BOX	10 000 €	10 001 €
TWO-BOX	0 €	1 €

SOLVING NEWCOMB'S PROBLEM

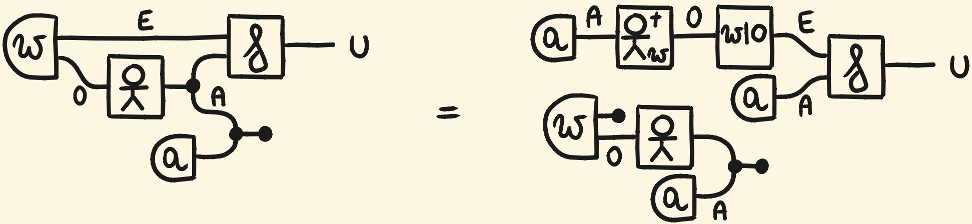
Evidential decision theory asks:
"Which action would be evidence for the best-case scenario?"
i.e. "Which action maximises the average of the state below?"



A TEMPLATE FOR DECISION PROBLEMS



We want to maximise the average of



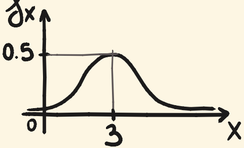
OUTLINE

- Partial Markov categories
- Discrete partial Markov categories
- Bayes, Pearl and Jeffrey
- Evidential decision theory
- [Exact observations]

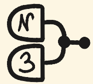
FREELY ADDING EXACT OBSERVATIONS

Comparators may

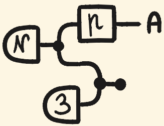
- not exist ⇝ equality need not be computable
- behave badly ⇝ probability 0 events may occur



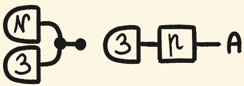
$P(X = 3) = 0$



$= 0$



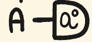
$(a) =$

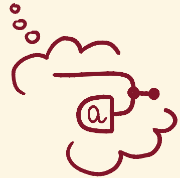


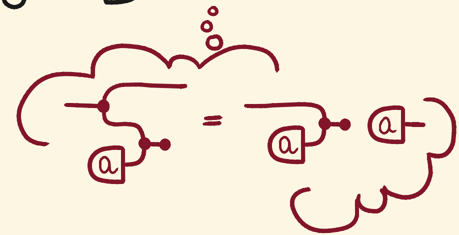
$(a) = 0 \cdot p(a|3) = 0$

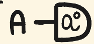
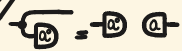
⇒ we add exact observations syntactically

PROCESSES WITH EXACT OBSERVATIONS

We construct a partial Markov category exOb($\mathcal{C}$) on top of a Markov category $\mathcal{C}$ by freely adding, for every deterministic state  in $\mathcal{C}$, a costate  and quotienting by  .



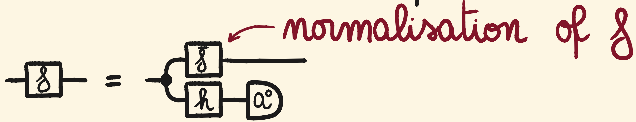


exOb($\mathcal{C}$) = ($\mathcal{C}$ + {  |  deterministic}) / 

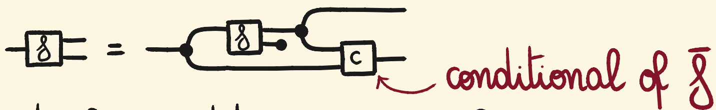
embeds faithfully into ($\mathcal{C}$ + ) / partial Frobenius

COMPUTING PROCESSES WITH EXACT OBSERVATIONS

Morphisms in $\mathrm{exOb}(\mathcal{C})$ have a normal form



that can be computed by conditioning in $\mathcal{C}$,
and they have conditionals



that can be computed by conditioning in $\mathcal{C}$.

SUMMARY

- Partial Markov = Markov − totality
- Discrete partial Markov = partial Markov + equality check
- Decision problems are easier with string diagrams (and also Bayes, Pearl & Jeffrey)
- Problems with equality checks? Add them freely

THANKS FOR LISTENING!