

Émergence d'une structure  
causale à partir d'apprentissage  
par observation dans une  
architecture psychologique  
hybride

Sébastien Hélié

# Plan

1. Motivation
2. Problématique
3. Hypothèses
4. Plan de travail
5. Choix du réseau connexionniste
6. Inférence du réseau Bayésien
7. Modèle complet

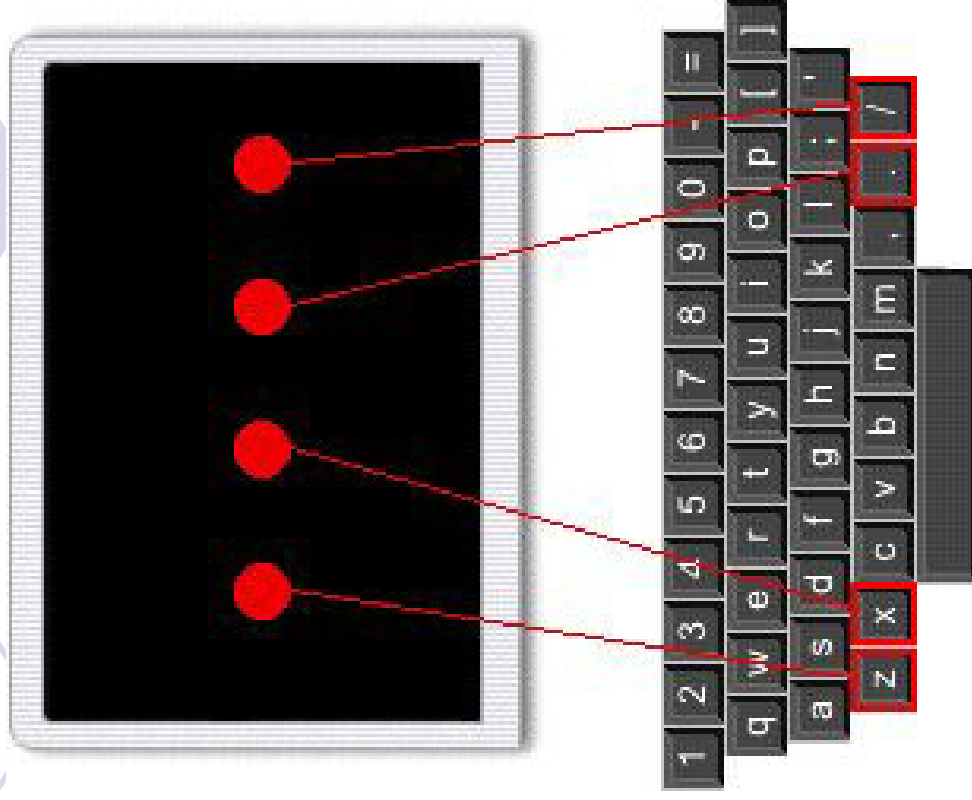
# 1. Motivation

- Connaissances explicites
  - Conscientes (i.e., metacognition, savoir que l'on sait);
  - Verbalisables (i.e., *Voici comment fonctionne ce logiciel.*);
  - Factuelles (i.e., *La tour Eiffel est à Paris.*).
- Avantages: Facilitent le transfert des habiletés, le développement des connaissances procédurales et la communication des connaissances.
- Inconvénients: Lenteur du traitement.
- Exemple d'expérience: Le rappel libre.

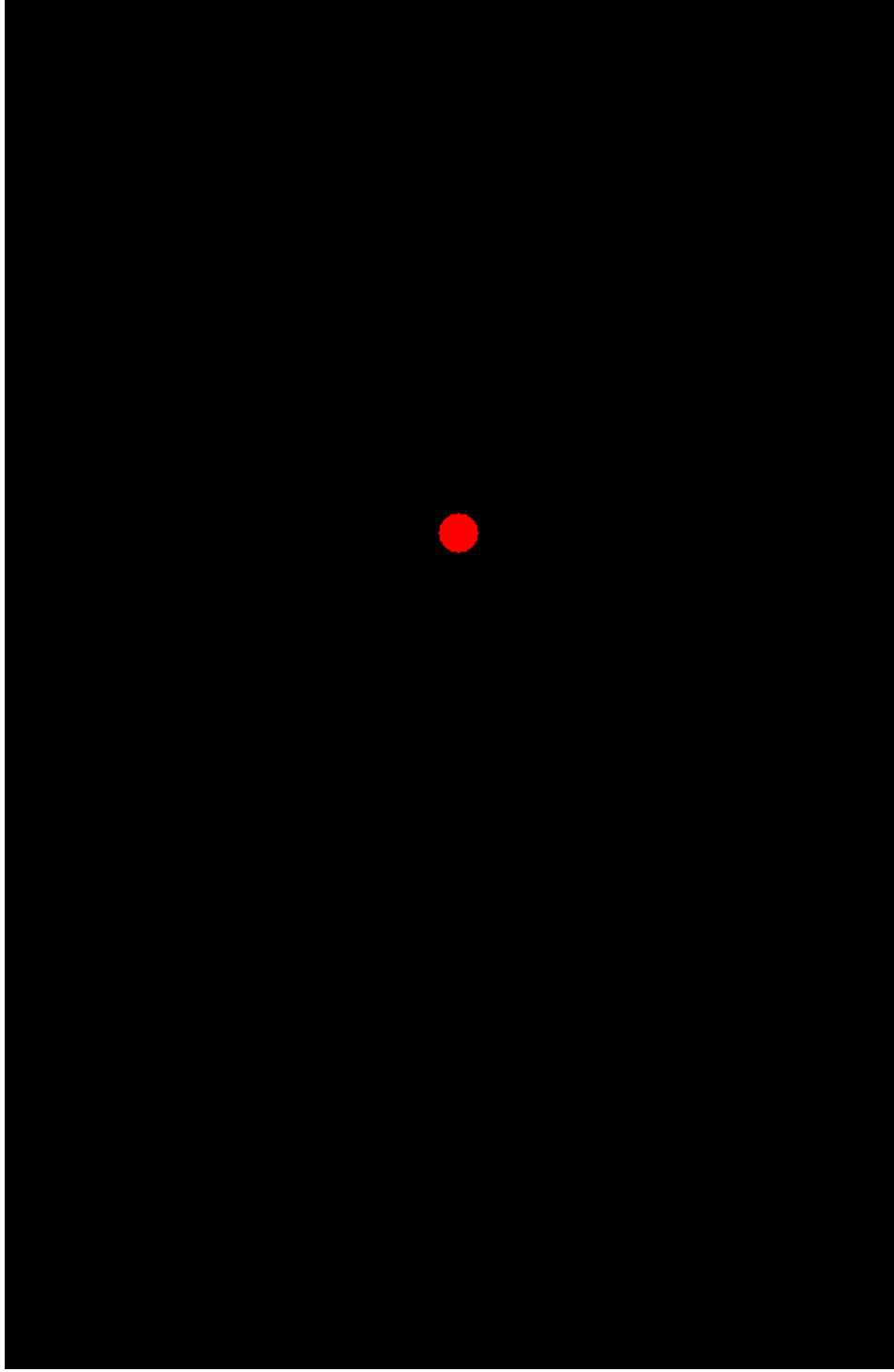
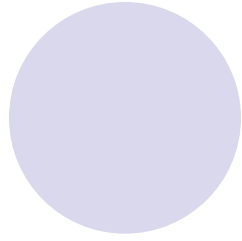
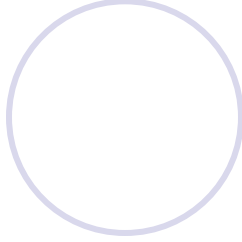
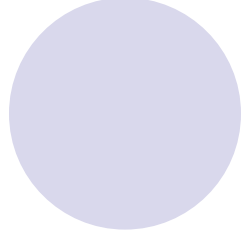
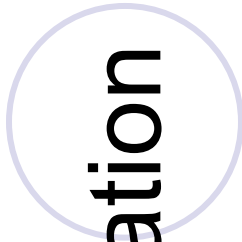
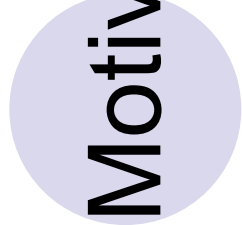
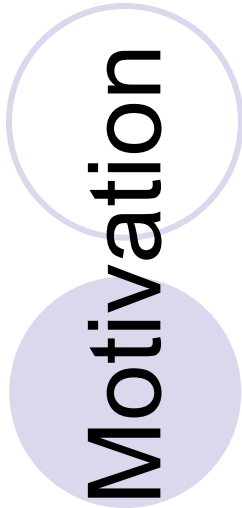
# 1. Motivation

- **Connaissances implicites:**
  - Non-conscientes (i.e, amorçage);
  - Non-verbalisables (i.e, apprentissage de séquences);
  - Procédurales (i.e., série de mouvements).
- **Avantages: Rapides et automatiques.**
- **Inconvénients: Difficiles à communiquer, transfert des habiletés déficient.**
- **Exemple d'expérience: Apprentissage de séquences.**

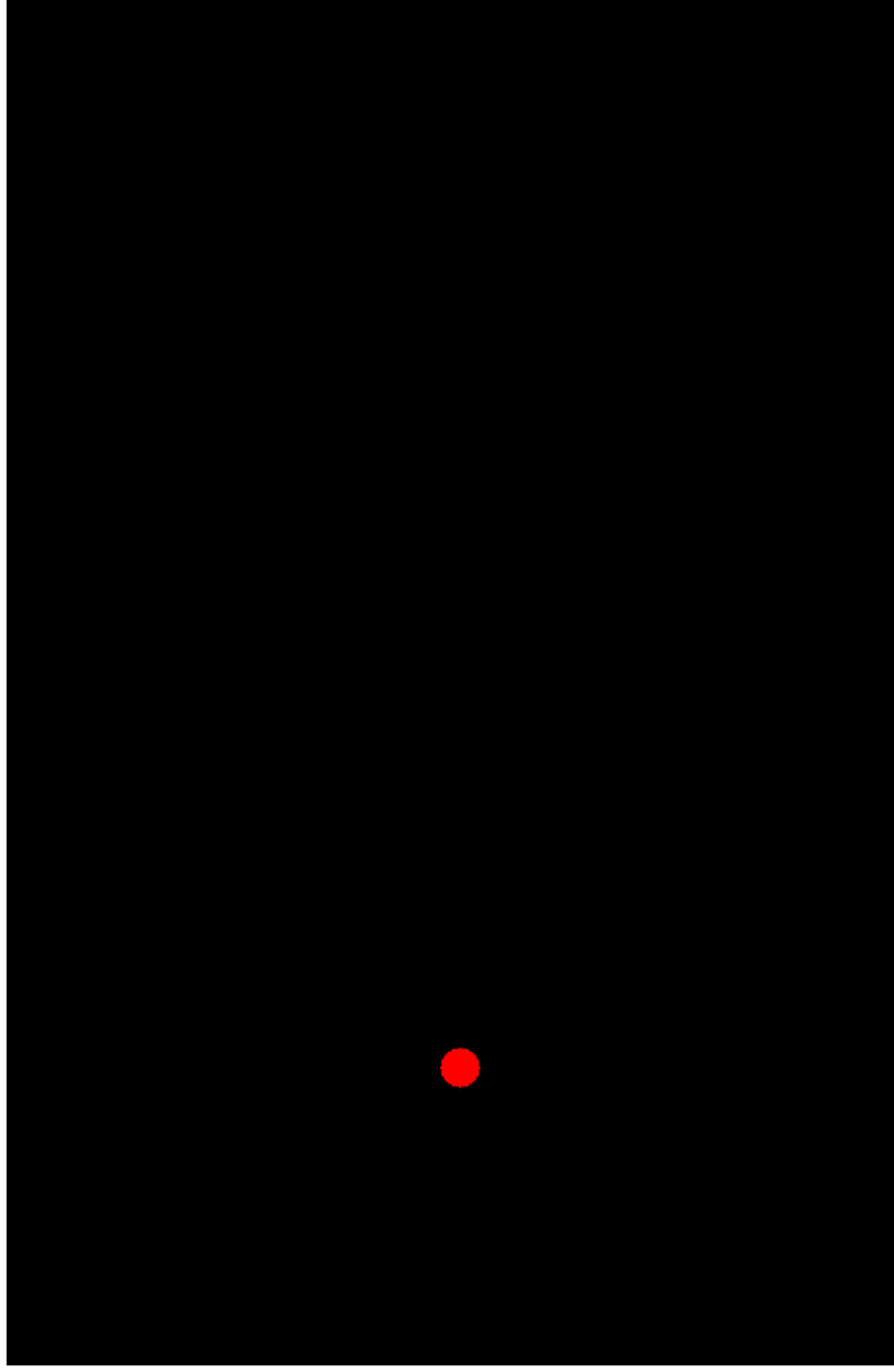
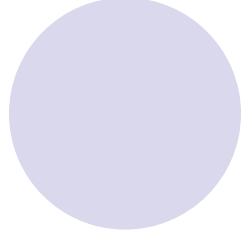
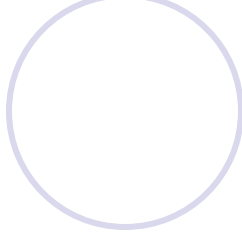
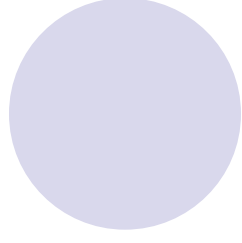
# 1. Motivation



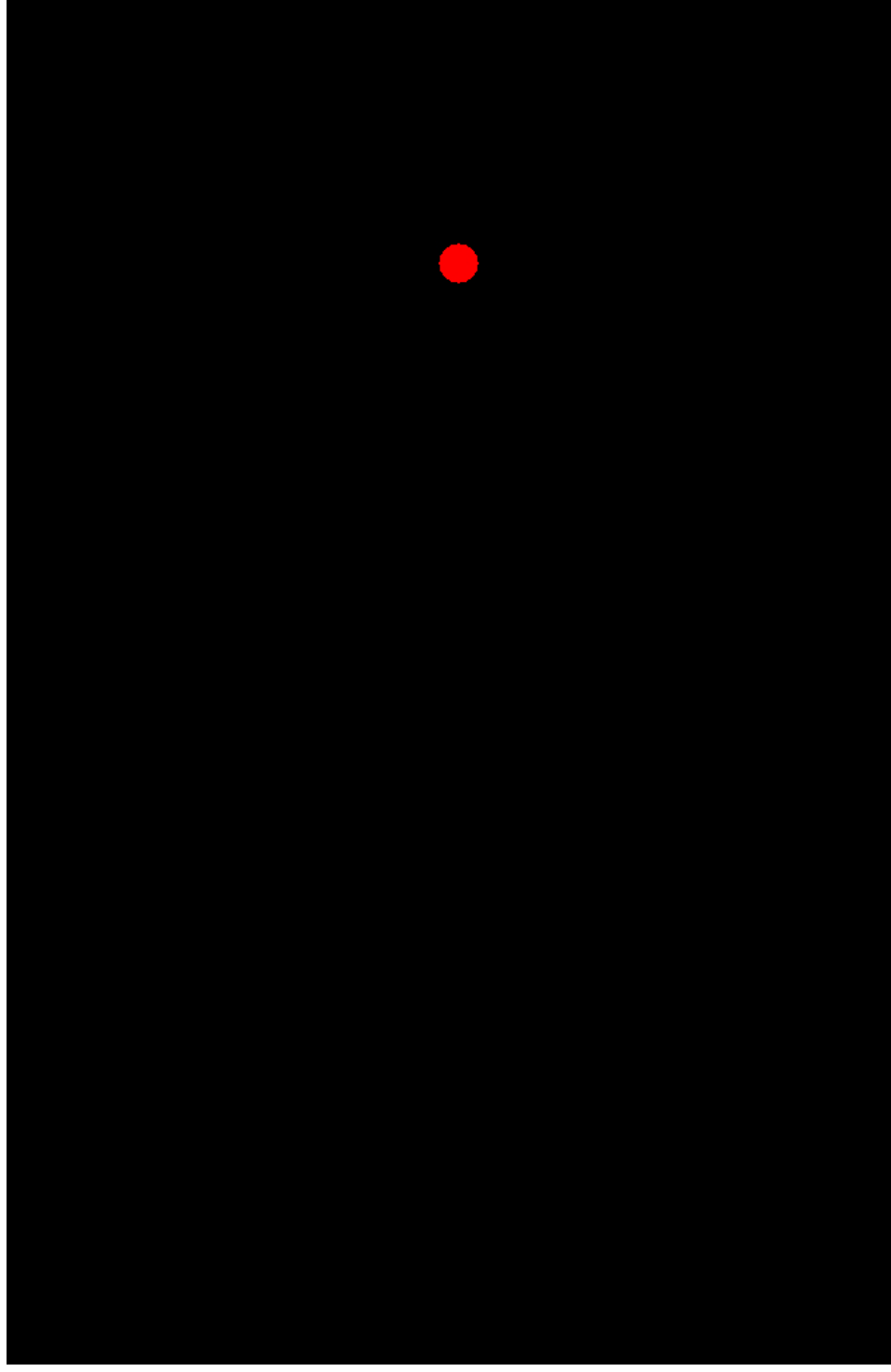
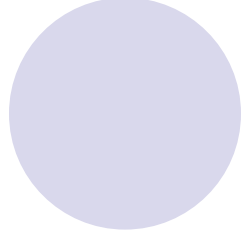
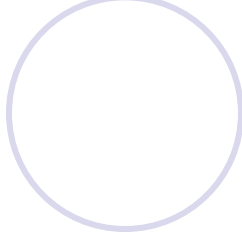
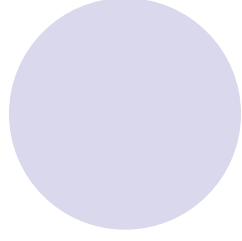
# 1. Motivation



# 1. Motivation

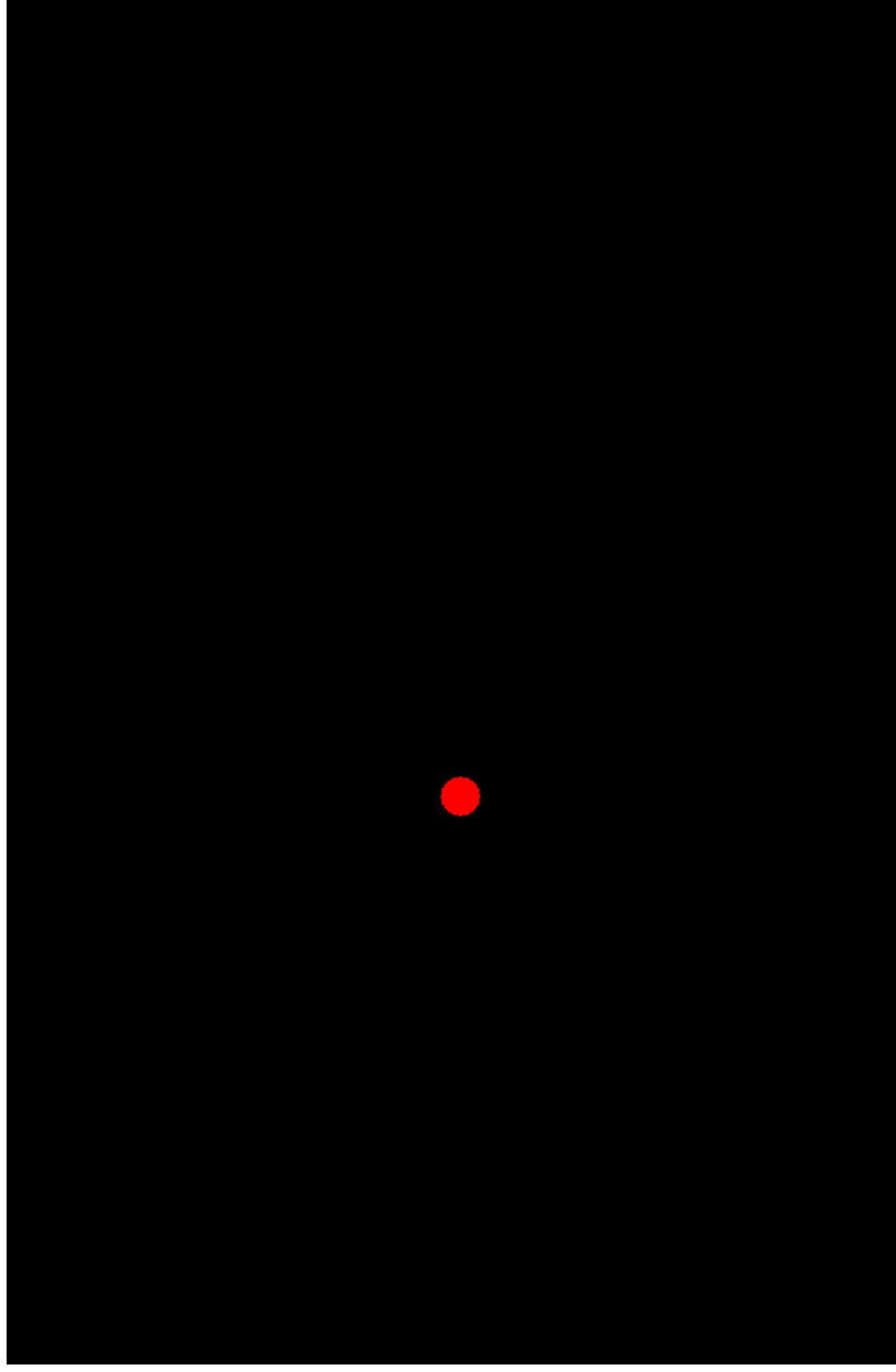
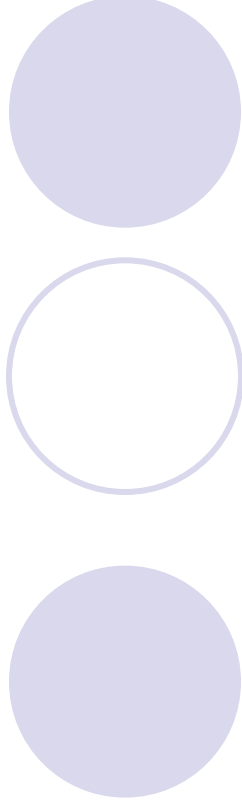


# 1. Motivation





# 1. Motivation



S'il existe une séquence qui se répète, les TR diminuent!



# 1. Motivation

- Les connaissances implicites et explicites ont des avantages différents.
- Avantage des systèmes incluant une redondance représentationnelle.
- Les humains semblent utiliser ces deux types de connaissances.

Un modèle psychologique complet doit inclure ces deux types de connaissances.



# Plan

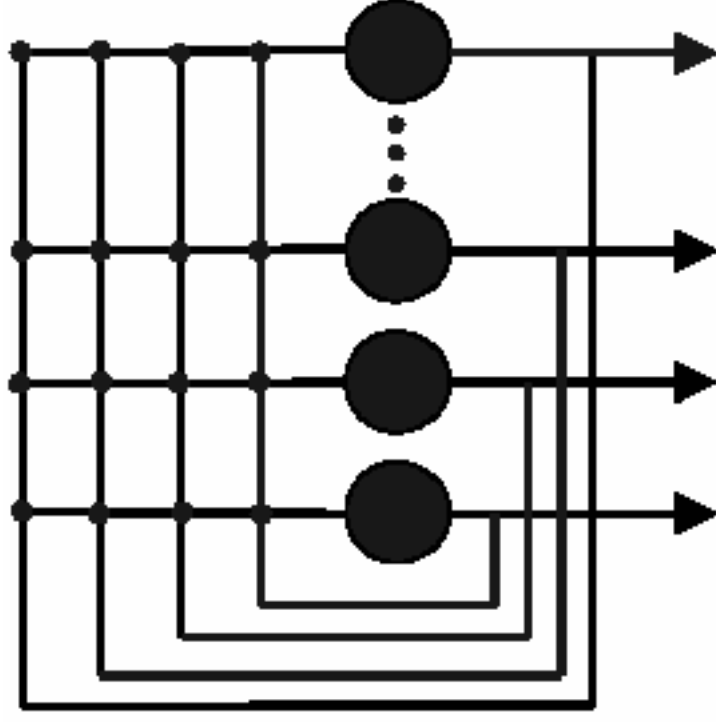
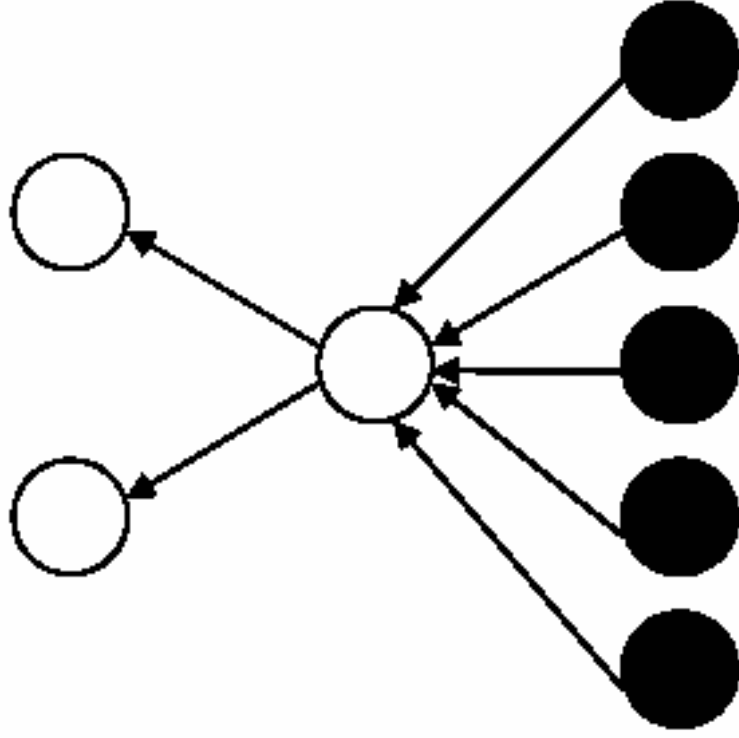
1. Motivation
2. **Problématique**
3. Hypothèses
4. Plan de travail
5. Choix du réseau connexionniste
6. Inférence du réseau Bayésien
7. Modèle complet

## 2. Problématique

- Les modèles habituellement utilisés en psychologie utilisent un seul type de représentations.
- Deux grandes familles de modèles:
  - Les réseaux connexionnistes;
  - Les modèles à base de règles (symboliques).

## 2. Problématique

Réseaux connexionnistes



## 2. Problématique

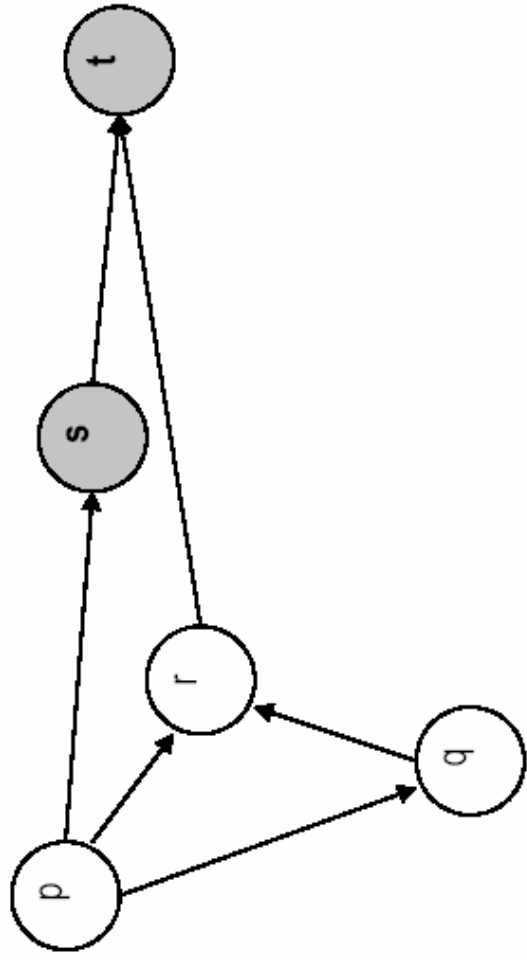
- Apprentissage:
  - Rétropropagation de l'erreur ↔ descente de gradient
  - Règles compétitives ↔ analyse en k moyens
  - Règles associatives ↔ ACP
- Représentations distribuées:
  - Microsémantique;
  - Représentations sous-symboliques.
- Difficulté d'interprétation des poids de connexions, ce qui correspond à l'opacité des connaissances implicites.

## 2. Problématique

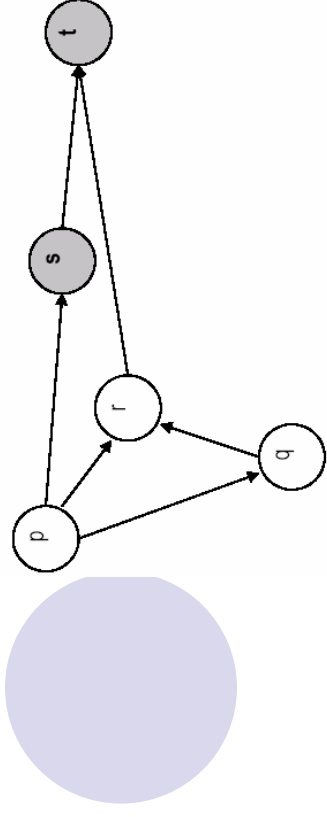
Modèles à base de règles

Système de production

$p \rightarrow q$   
 $p \rightarrow r$   
 $p \rightarrow s$   
 $q \rightarrow r$   
 $r \rightarrow t$   
 $s \rightarrow t$   
 $s$



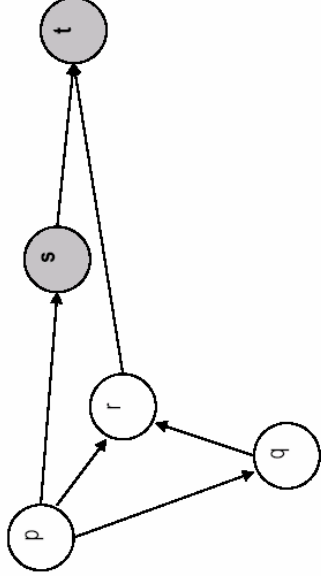
## 2. Problématique



- Graphe directionnel acyclique (*DAG*), dans lequel les liens représentent la causalité.
- Graphe d'indépendance (*IMAP*) représentant une distribution conjointe.
- Chaque sommet représente une variable et, si le graphe contient toutes (et seulement) les indépendances de la distribution, chaque arc représente une dépendance directe.



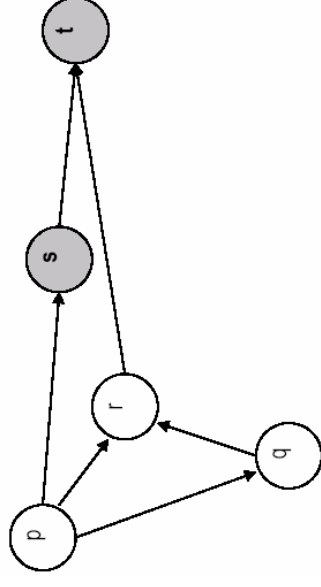
## 2. Problématique



THÉORÈME: Si  $a$  et  $b$  sont conditionnellement indépendants étant donné  $z$  dans la distribution représentée par un graphe  $G$ , alors  $a$  et  $b$  sont d-séparés par  $z$  dans le graphe  $G$ .

Indépendance conditionnelle  $\leftrightarrow$  d-séparation

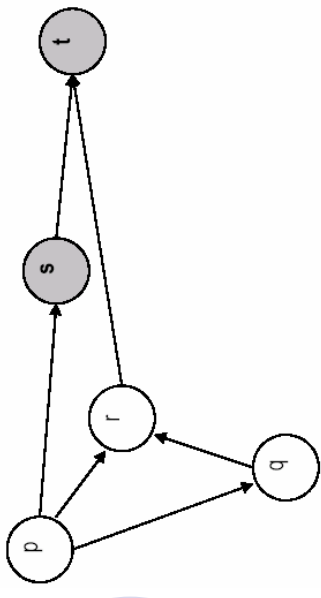
## 2. Problématique



DÉFINITION: Deux sommets  $\{A, B\}$  sont d-  
séparés par l'ensemble  $C$  ssi tous les  
chemins entre  $A$  et  $B$  sont bloqués par  $C$ .

Un chemin entre  $A$  et  $B$  est bloqué par  $C$  ssi:

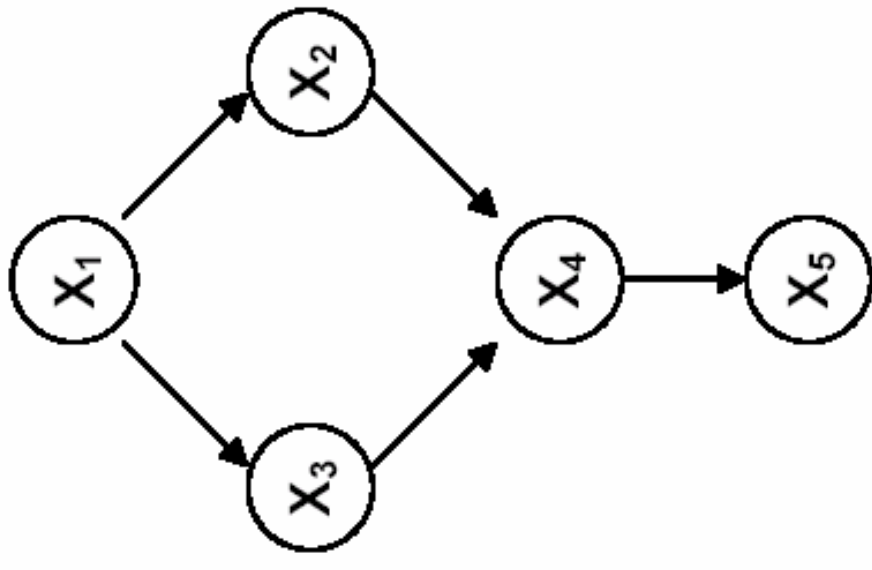
## 2. Problématique



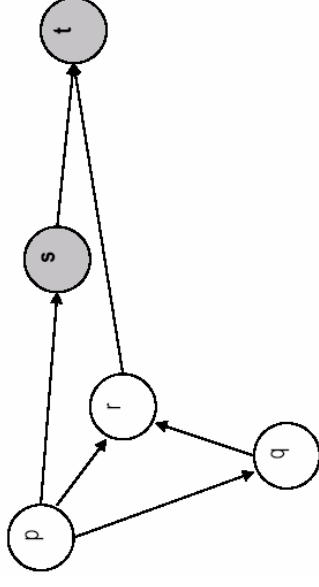
1. il existe un élément  $i$  inclus dans  $C$  et  $i$  n'est pas un face-à-face ( $\rightarrow \leftarrow$ )

OU

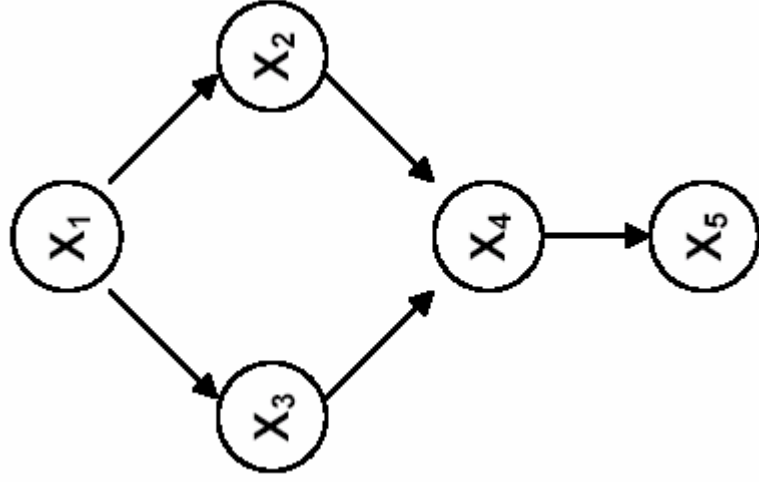
2. il existe un face-à-face  $i$  tel que  $i$  ainsi que tous ses descendants ne sont pas inclus dans  $C$



## 2. Problématique



1.  $\{X_3\}$  et  $\{X_2\}$  sont d-séparés par  $z = \{X_1\}$ 
  1.  $X_1$  n'est pas un face-à-face et est inclus dans  $z$  (bloqué par 1).
  2.  $X_4$  est un face-à-face ne faisant pas partie de  $z$  (ni ses descendants: bloqué par 2).
2.  $\{X_3\}$  et  $\{X_2\}$  ne sont pas d-séparés par  $z = \{X_1, X_5\}$ 
  1. Idem à 1.1
  2.  $X_5$  est le descendant d'un face-à-face inclus dans  $z$ : ce chemin est ouvert.



## 2. Problématique

- Revenons aux modèles à base de règles...
- Représentations symboliques;
  - Un réseau Bayésien est équivalent à un système de production codé en utilisant une logique propositionnelle.
- Puissance de calcul (thèse Church-Turing);
- Compositionnalité ( $a \wedge b$  est formé à partir de  $\{a, b\}$ );
- Systématicité (si le système comprend  $a \wedge b$ , il comprend  $b \wedge a$ );
- Interprétation directe des symboles et règles, ce qui rappelle la transparence des connaissances explicites.



## 2. Problématique

Revenons au problème de l'unicité représentationnelle des modèles...

Une solution évidente est d'utiliser un modèle psychologique hybride, utilisant un réseau connexionniste pour représenter les connaissances implicites et une composante à base de règles pour les connaissances explicites.

Cependant, l'utilisation d'une composante symbolique engendre un nouveau problème, celui de l'ancrage des symboles.

## 2. Problématique

- Le dictionnaire chinois – chinois:

班

烏

Même problématique que la chambre chinoise de Searle.

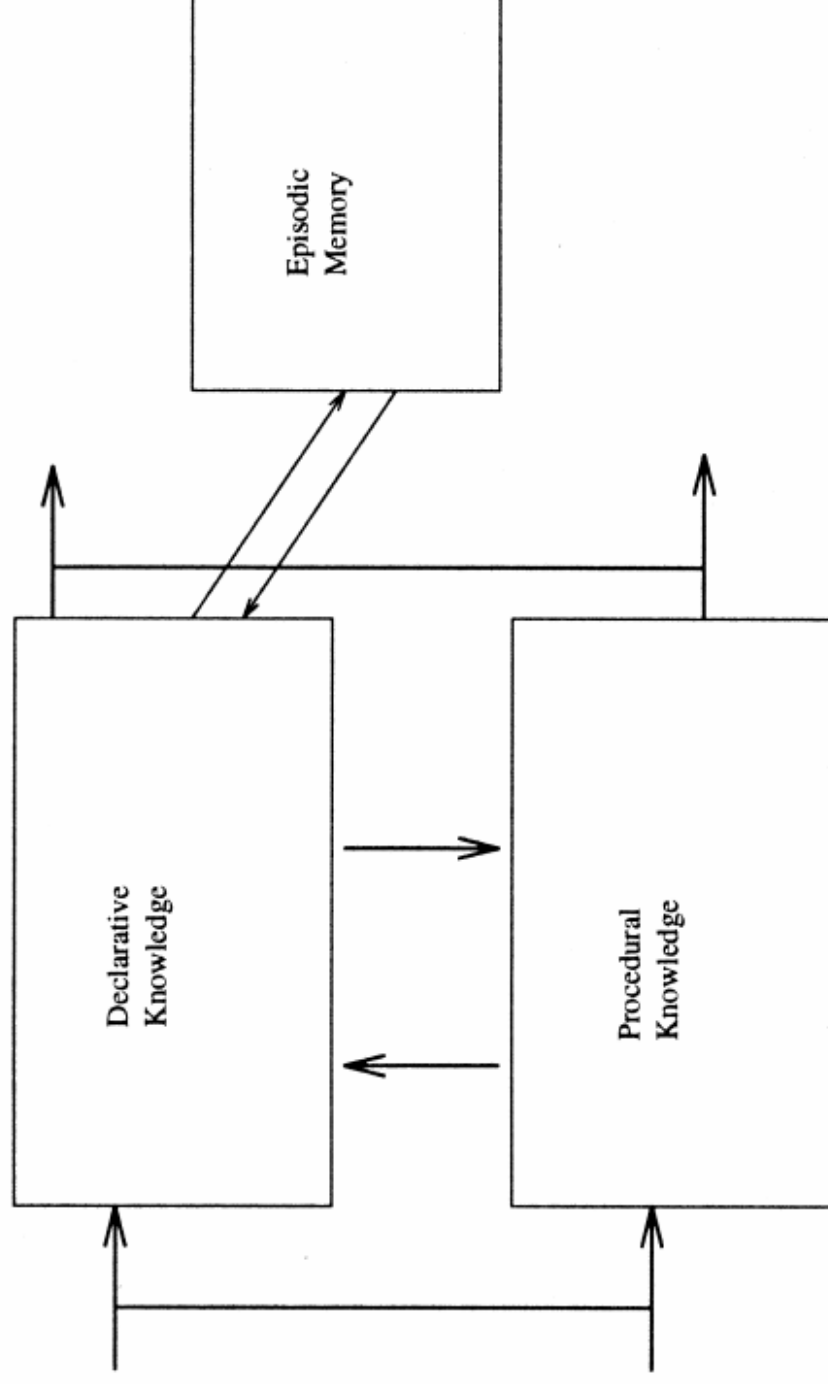
## 2. Problématique

- L'ancrage de certains symboles (*primitifs*) se transmet aux symboles formés par composition;
- Des symboles *actifs* peuvent être utilisés comme primitifs (ancrés de façon significative dans le réseau connexionniste);
- Une façon de créer des symboles actifs est de laisser le réseau connexionniste générer ses propres symboles (un retour à l'idée de la redondance représentationnelle).



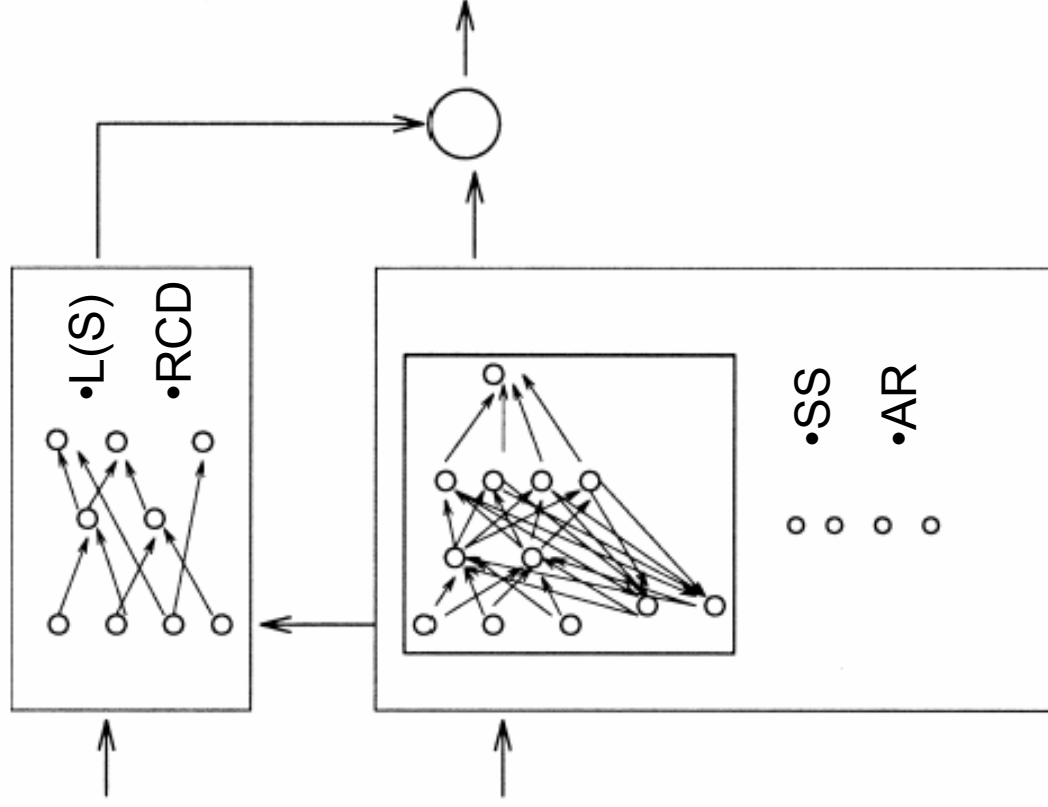
## 2. Problématique

Connectionnist Learning with Adaptive Rule Induction ON-line (CLARION)



Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science*, 25, 203-244.

## 2. Problématique



### Mémoire épisodique

{stimulus<sub>1</sub>, réponse<sub>1</sub>, résultat<sub>1</sub>}

{stimulus<sub>2</sub>, réponse<sub>2</sub>, résultat<sub>2</sub>}

{stimulus<sub>3</sub>, réponse<sub>3</sub>, résultat<sub>3</sub>}

...

{stimulus<sub>n</sub>, réponse<sub>n</sub>, résultat<sub>n</sub>}

- Règles apprises en utilisant un algorithme de programmation logique inductive;
- À chaque essai, une règle peut être:
  - Créée;
  - Spécifiée;
  - Généralisée.

## 2. Problématique

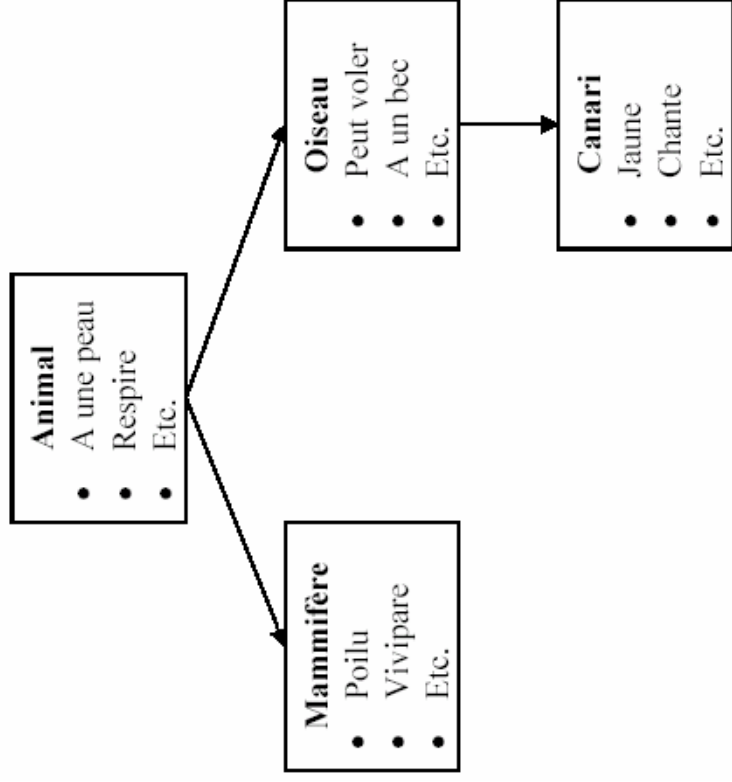
CLARION est le meilleur modèle psychologique hybride et ancré. Il a été utilisé pour modéliser des tâches de navigation, ainsi que l'apprentissage de séquences.

4 problèmes majeurs:

1. Représentations sous-symboliques ou microsémantique?
2. Supervision du niveau procédural.
3. Formation des symboles.
4. Déterminisme des règles déclaratives.

## 2. Problématique

### Modèle de mémoire sémantique



$$\left( \begin{array}{l} \text{A une peau} \\ \text{Respire} \\ \text{Poilu} \\ \text{Vivipare} \\ \text{Peut voler} \\ \text{A un bec} \end{array} \right) \Rightarrow \text{Oiseau} = \left( \begin{array}{l} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{array} \right)$$

Cette représentation n'est pas sous-symbolique!



## 2. Problématique

- Dans une tâche de navigation, CLARION voit:

$\{0,1\}^{43} \Rightarrow$

(  
Essence  
Portée  
Très à gauche  
Gauche  
Devant  
Droite  
Très à droite  
Derrière  
S<sub>1</sub> Très loin  
S<sub>1</sub> Loin  
S<sub>1</sub> Moyen  
S<sub>1</sub> Proche  
S<sub>1</sub> Très proche  
...  
S<sub>7</sub> Très loin  
S<sub>7</sub> Loin  
S<sub>7</sub> Moyen  
S<sub>7</sub> Proche  
S<sub>7</sub> Très proche  
)

- Ce qui est clairement une microsémantique!

## 2. Problématique

- Dans CLARION, le niveau procédural (implicite) est supervisé (apprentissage par renforcement).
- Or, les expériences portant sur la mémoire implicite n'incluent généralement pas de rétroaction.

→ Le module implicite devrait être non-supervisé.

## 2. Problématique

- Dans CLARION, seules les règles émergent des connaissances procédurales (les symboles sont déjà présents).
- Ainsi, seules les règles sont ancrées.
- Cette limitation résulte de l'algorithme d'inférence utilisé, qui nécessite des connaissances *a priori*.



## 2. Problématique

- Dans CLARION, l'utilisation des règles déclaratives est déterministe.
- Or, le débat déterministe vs. stochastique est toujours ouvert en psychologie expérimentale.
- Comme certains résultats soutiennent chacun de ces processus de décisions, un modèle général de la cognition doit être en mesure d'effectuer les deux types de décisions.



# Plan

1. Motivation
2. Problématique
3. **Hypothèses**
4. Plan de travail
5. Choix du réseau connexionniste
6. Inférence du réseau Bayésien
7. Modèle complet

### 3. Hypothèses

Hypothèses générales:

1. Il existe deux types de connaissances chez l'humain (implicites et explicites);
2. Un modèle psychologique général doit être hybride afin de représenter séparément ces deux types de connaissances;
3. La partie explicite du modèle doit être entièrement générée par la partie implicite afin de s'assurer de son ancrage (3);
4. Les deux composantes du modèle sont autonomes, mais co-existent et coopèrent afin de solutionner chaque tâche.

### 3. Hypothèses



Hypothèses spécifiques:

1. Le niveau implicite peut être modélisé par un réseau connexionniste non-supervisé utilisant des représentations sous-symboliques (1 et 2);
2. Le niveau explicite peut être modélisé par un réseau Bayésien (règles causales stochastiques: 4).



# Plan

1. Motivation
2. Problématique
3. Hypothèses
4. **Plan de travail**
5. Choix du réseau connexionniste
6. Émergence du réseau de connaissance Bayésien
7. Modèle complet

## 4. Plan de travail

1. Choix d'un réseau connexionniste non-supervisé pour modéliser le niveau implicite (mémoires associatives vs. réseaux compétitifs).

**Validation: Capacité des modèles à effectuer une estimation de densité.**

2. Inférence d'un réseau Bayésien (méthodes Bayésiennes vs. méthodes par contraintes).

**Validation: Capacité de reproduire le réseau ALARM.**

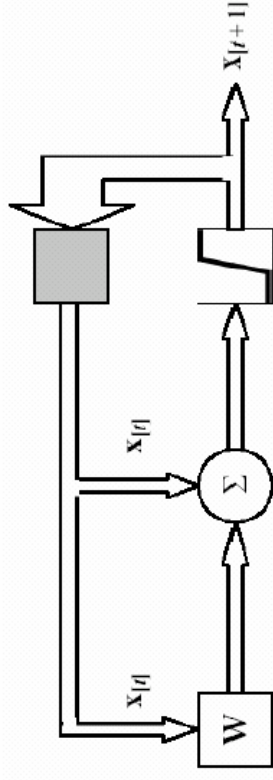
3. Modèle complet: fonctionnement autonome de chaque module (inférence et apprentissage), ainsi que l'interaction entre les deux modules.

**Validation: Capacité de reproduction des données humaines (inférence de structures causales et apprentissage de séquences).**

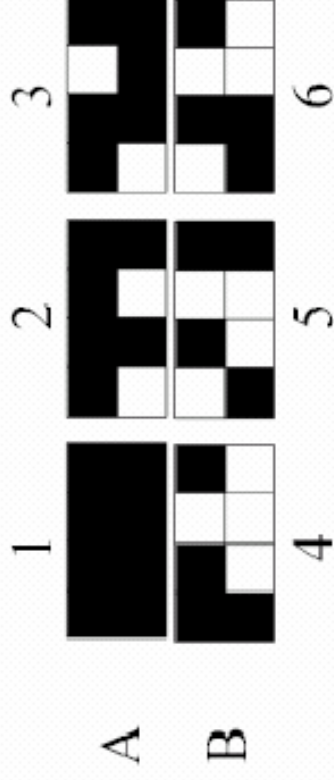
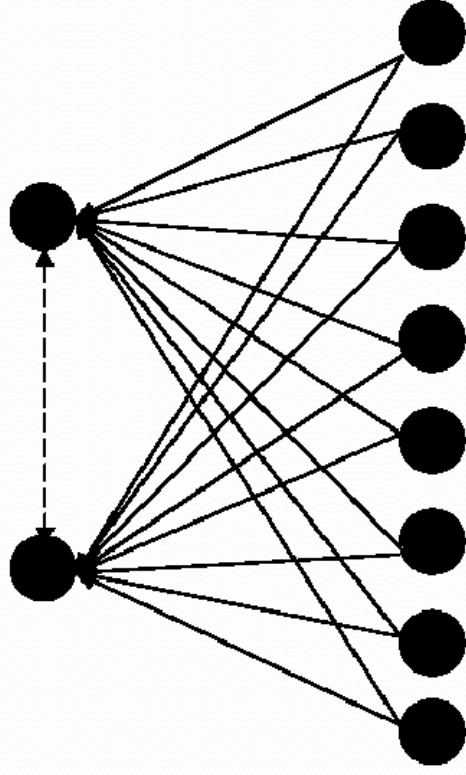
# Plan

1. Motivation
2. Problématique
3. Hypothèses
4. Plan de travail
5. **Choix du réseau connexionniste**
6. Inférence du réseau Bayésien
7. Modèle complet

# 5. Choix du réseau connexionniste



(a)

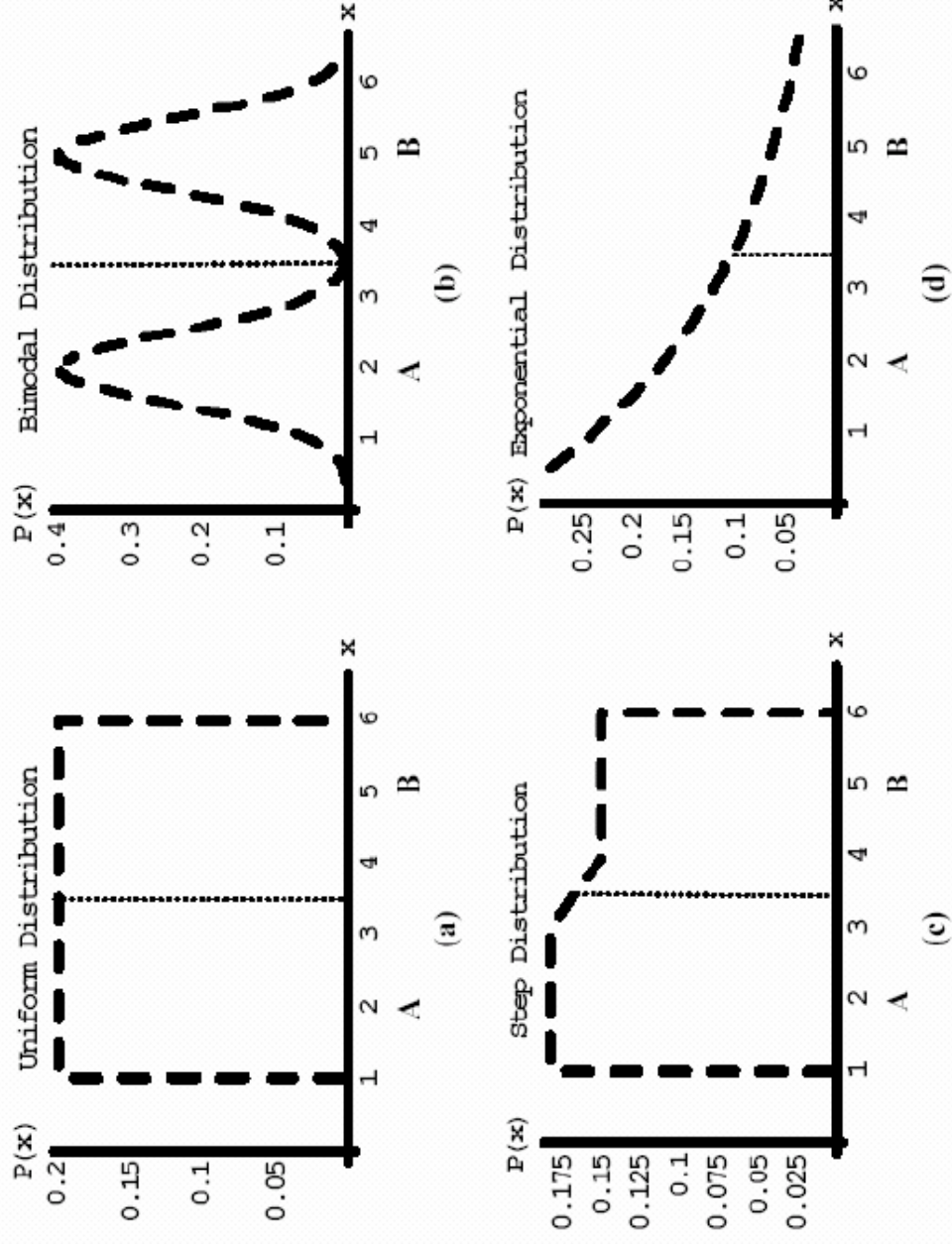


$$\begin{pmatrix} 1. & 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 1. & 0.5 & 0 & 0 & 0 \\ 0.5 & 0.5 & 1. & 0 & 0 & 0 \\ 0 & 0 & 0 & 1. & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 1. & 0.5 \\ 0 & 0 & 0 & 0.5 & 0.5 & 1. \end{pmatrix}$$

Hélie, S., Chartier, S., & Proulx, R. (accepted). Are unsupervised neural networks ignorant? Sizing the effect of environmental distributions on unsupervised learning. *Cognitive Systems Research*.



# 5. Choix du réseau connexionniste



Hélie, S., Chartier, S., & Proulx, R. (accepté). Are unsupervised neural networks ignorant? Sizing the effect of environmental distributions on unsupervised learning. *Cognitive Systems Research*.

# 5. Choix du réseau connexionniste

Table 1.

Classification of the random vectors

	RAMs			Competitive Networks		
	"A"s	"B"s	$\chi^2$	"A"s	"B"s	$\chi^2$
Uniform	153 (0.46)	178 (0.54)	0.593	237 (0.47)	263 (0.53)	0.512
Bimodal	220 (0.44)	276 (0.56)	2.915	246 (0.49)	254 (0.51)	0.200
Step	195 (0.64)	110 (0.36)	7.053	238 (0.48)	262 (0.52)	15.75*
Exponential	357 (0.72)	142 (0.28)	3.085	221 (0.44)	279 (0.56)	222.4*

*Note.* 500 random vectors were tested in each cell. Vectors not appearing in the RAMs' conditions stabilized in spurious states. Numbers in parenthesis represent proportions. \* indicate results that significantly differed from their training distribution according to a  $\chi^2(1)$ ,  $\alpha = .001$ . The critical value was 10.83.

Hélie, S., Chartier, S., & Proulx, R. (accepté). Are unsupervised neural networks ignorant? Sizing the effect of environmental distributions on unsupervised learning. *Cognitive Systems Research*.

## 5. Choix du réseau connexionniste

- Le BSB réussi à bien estimer la densité dans toutes les conditions.
- Le réseau compétitif estime toujours une distribution uniforme.
- Les mêmes résultats ont été obtenus avec des modèles sophistiqués (ART1 et NDRAM) dans un environnement complexe (stimuli corrélés, dimensionnalité accrue).

Les mémoires associatives enregistrent l'information nécessaire à l'apprentissage d'une structure Bayésienne.

## 5. Choix du réseau connexionniste

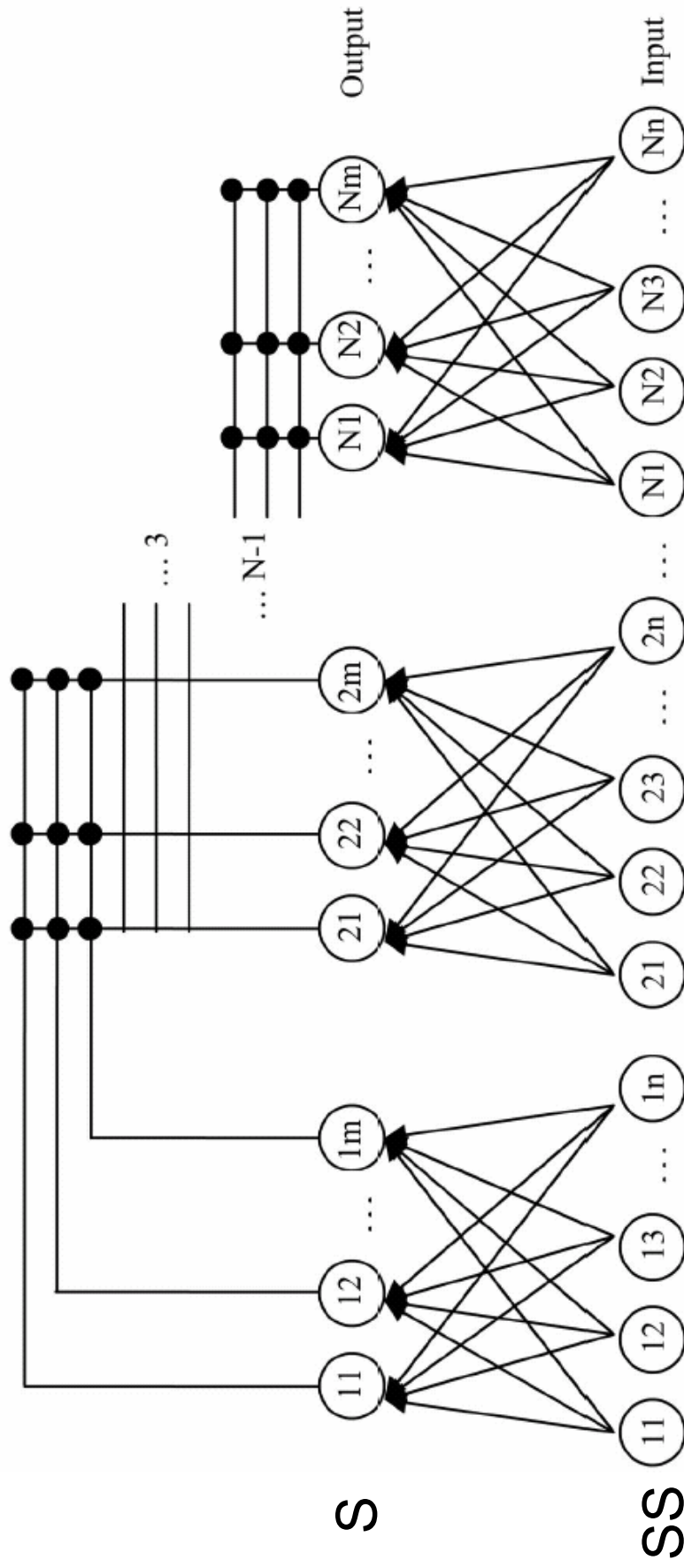
Problème: La matrice de poids de connexions des mémoires associatives est la matrice de covariance entre les unités.

Or, la mémoire implicite doit être modélisée par des représentations sous-symboliques, ce qui proscrit l'utilisation des représentations locales.

Donc, une nouvelle mémoire associative doit être créée.

# 5. Choix du réseau connexionniste

## Joint Probability Extractor (JPEX):

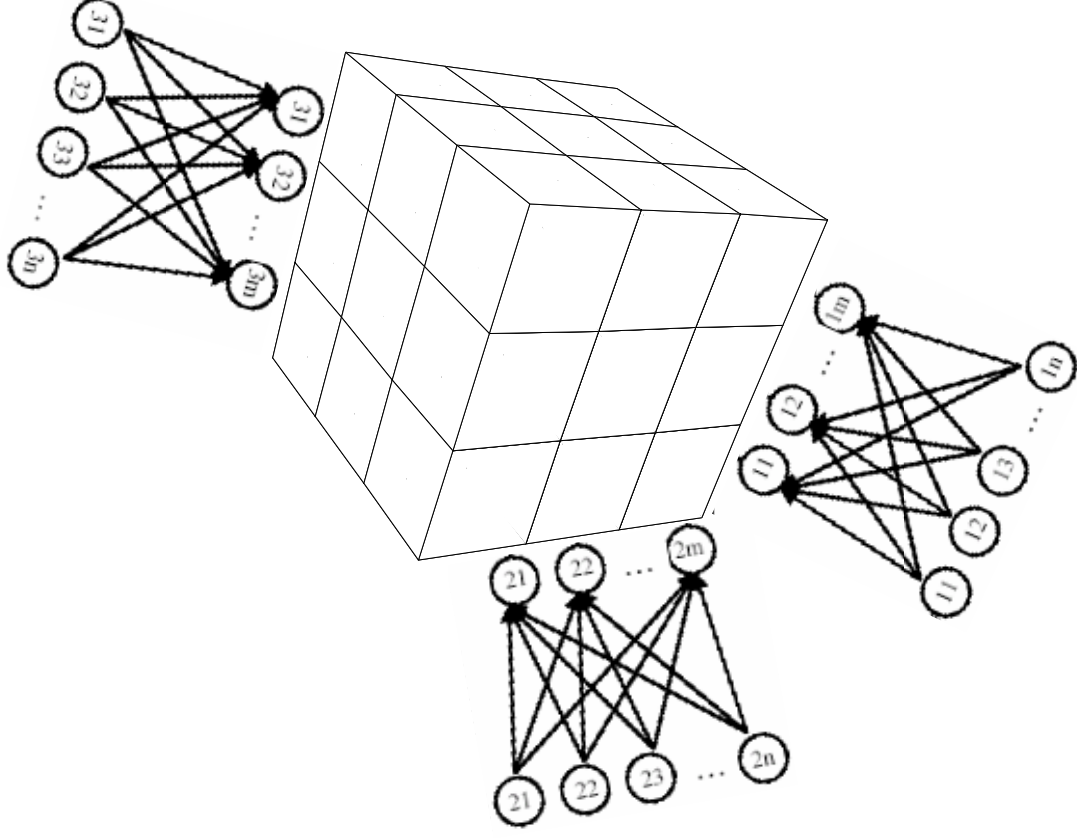


# 5. Choix du réseau connexionniste

Tenseur associatif:

$$\mathbf{V}_{[t+1]} = \mathbf{V}_{[t]} + \bigotimes_{i=1}^N \mathbf{y}_i$$

$$\bigotimes_{i=1}^N \mathbf{y}_i = \begin{cases} \mathbf{y}_1 \mathbf{y}_2^T \mathbf{y}_3 \mathbf{y}_4^T \dots \mathbf{y}_N, & \text{si } N \text{ est impair} \\ \mathbf{y}_1 \mathbf{y}_2^T \mathbf{y}_3 \mathbf{y}_4^T \dots \mathbf{y}_N^T, & \text{sinon} \end{cases}$$









## 5. Choix du réseau connexionniste

Comme un tenseur de rang  $N$  permet d'enregistrer les probabilités conjointes d'ordre  $N - 1$ , les relations non-linéairement séparables peuvent également être apprises...

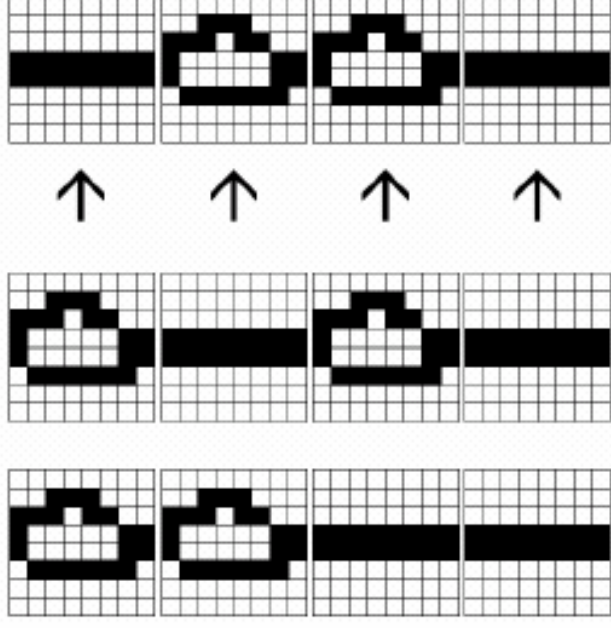
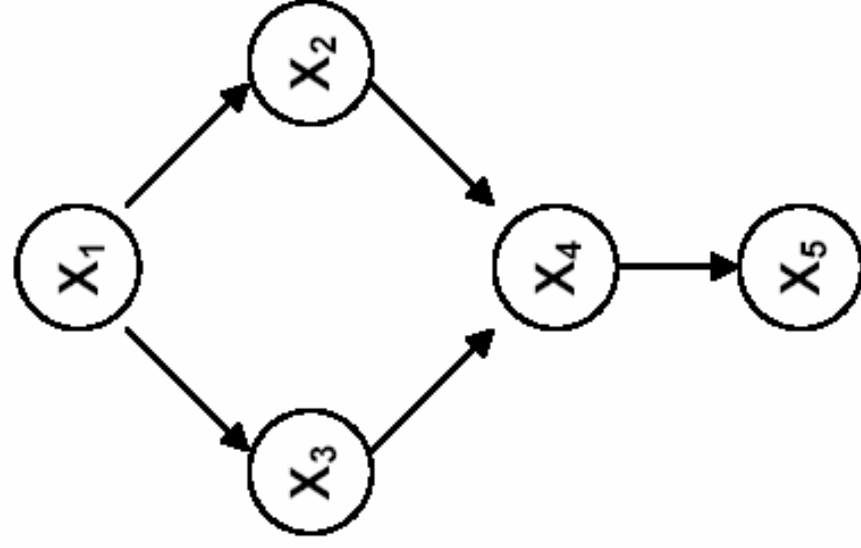


Figure 5: Associations learned in the XOR problem.

## 5. Choix du réseau connexionniste



... ainsi que les facteurs  
causaux indirects (à cause de  
la possibilité de conditionner  
sur  $N - 2$  variables).

## 5. Choix du réseau connexionniste

En bref...

1. Les mémoires associatives sont la seule famille de réseaux connexionnistes non-supervisés pouvant estimer correctement la densité de l'environnement.
2. JPEx est un gBAM utilisant la stabilité des réseaux compétitifs pour faire émerger les symboles.
3. Dans JPEx, un tenseur associatif relie les couches de sorties des  $N$  champs récepteurs, ce qui résulte en un parallélogramme rectangle de contingences contenant les probabilités conjointes d'ordre  $N - 1$ .
4. JPEx est le premier modèle associatif capable d'apprendre les relations non-linéairement séparables.

# Plan

1. Motivation
2. Problématique
3. Hypothèses
4. Plan de travail
5. Choix du réseau connexionniste
6. **Inférence du réseau Bayésien**
7. Modèle complet

## 6. Inférence du réseau Bayésien

Il existe 2 approches pour générer des structures Bayésiennes

1. Approches par contraintes
2. Approches Bayésiennes

## 6. Inférence du réseau Bayésien

Approches par contraintes:

Rappel: Indépendance conditionnelle  $\leftrightarrow$  d-séparation.

→ Si les indépendances conditionnelles présentes dans une distribution sont connues ( $X^2, G^2$ ), il est possible de construire un graphe la représentant.

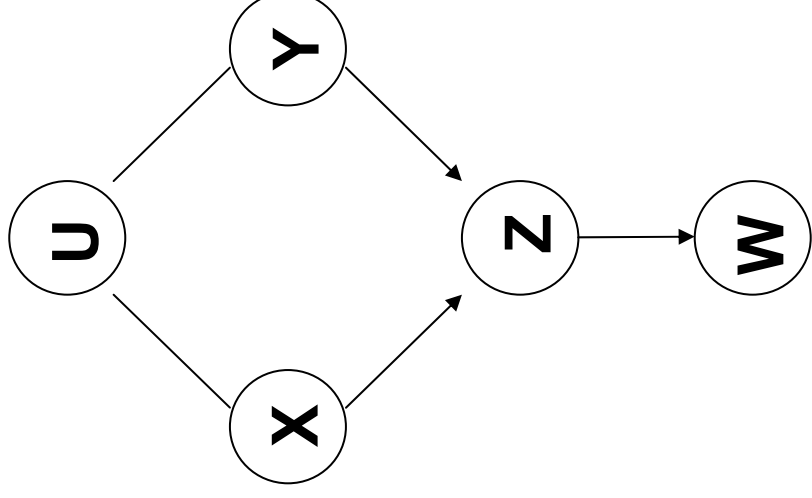
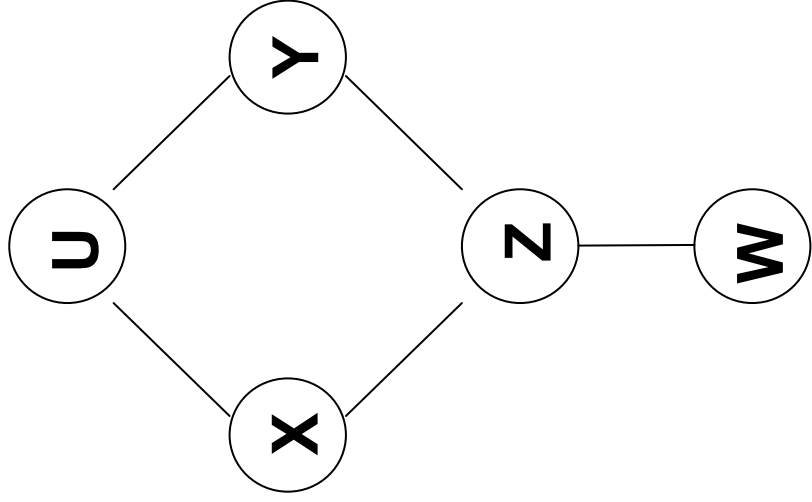
## 6. Inférence du réseau Bayésien

### Ex. Inductive Causation (IC)

1. Construction du squelette:
  1. Lier chaque paire de sommets, ssi les variables qu'ils représentent sont dépendantes étant donné tous les autres ensembles de variables.
2. Direction des arcs:
  1. Mettre un face-à-face à un sommet ssi il est un voisin commun ne rendant pas les variables représentées par les deux autres sommets conditionnellement indépendantes.
  2. Diriger les arcs restant en s'assurant:
    1. De ne pas ajouter de face-à-face;
    2. De ne pas créer de cycle.

## 6. Inférence du réseau Bayésien

$\{I(\{X\}, \{Y\} \mid \{U\}), I(\{U\}, \{Z, W\} \mid \{X, Y\}), I(\{X, Y, U\}, \{W\} \mid \{Z\})\}$





## 6. Inférence du réseau Bayésien

Approches Bayésiennes:

Un réseau Bayésien est le modèle d'une distribution.

→ Il est donc possible d'effectuer une recherche avare (*greedy search*) dans le voisinage du graphe afin de maximiser la vraisemblance de ce dernier.

## 6. Inférence du réseau Bayésien

Ex. dagSearch

Opérations possibles:

- Ajout d'arc;
- Retrait d'un arc;
- Inversion d'un arc.

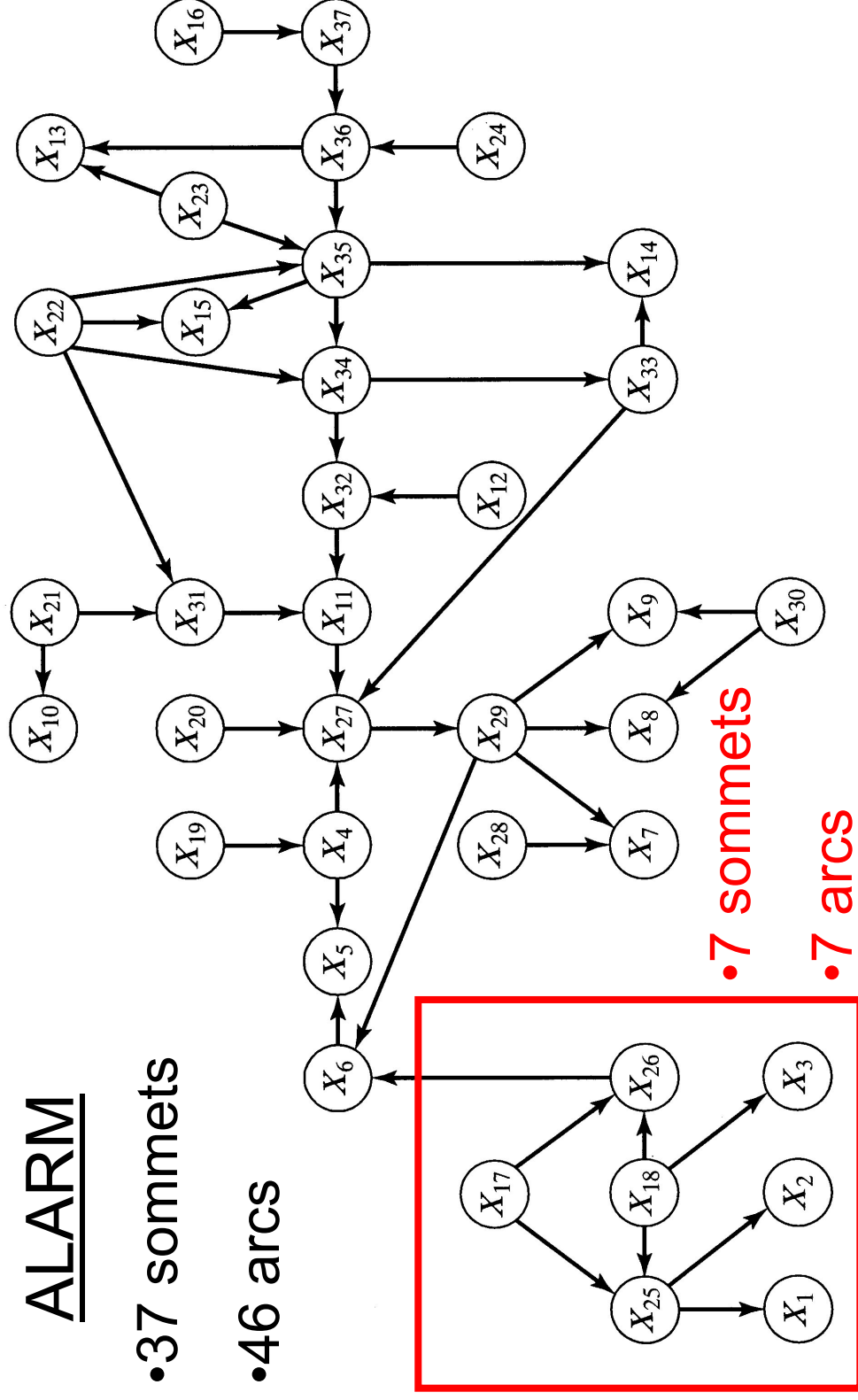
À chaque itération, effectuer l'opération qui augmente le plus la vraisemblance du modèle (sans créer de cycle).

# 6. Inférence du réseau Bayésien

## ALARM

•37 sommets

•46 arcs

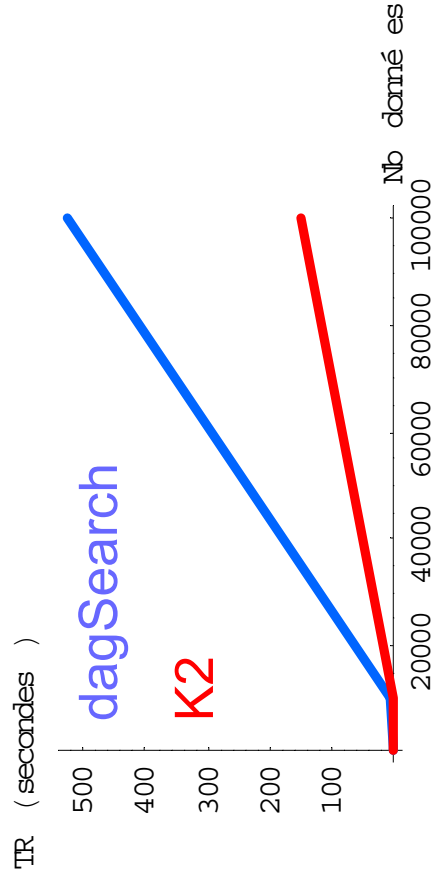


•7 sommets

•7 arcs

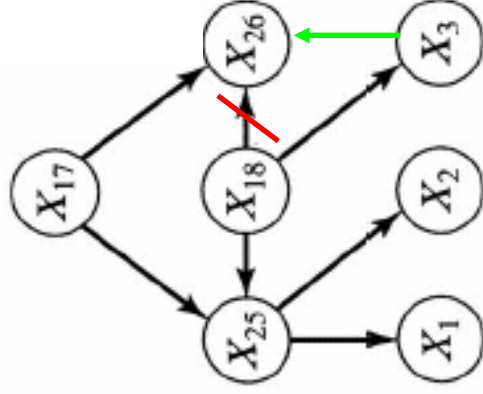
# 6. Inférence du réseau Bayésien

## Résultats partiels



- Avec 500 données, les deux algorithmes sont dans l'erreur.
- Avec 1000 données, K2 trouve le bon réseau alors que dagSearch ne le trouve pas toujours.

- À partir de 10 000 données, les deux algorithmes trouvent le bon réseau.

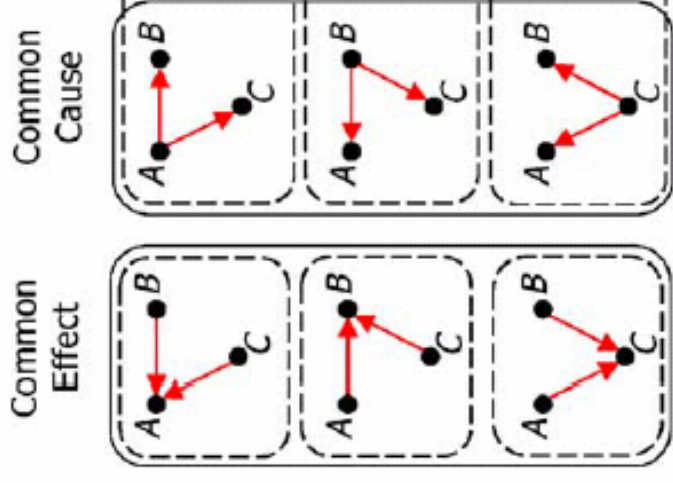


# Plan

1. Motivation
2. Problématique
3. Hypothèses
4. Plan de travail
5. Choix du réseau connexionniste
6. Inférence du réseau Bayésien
7. **Modèle complet**

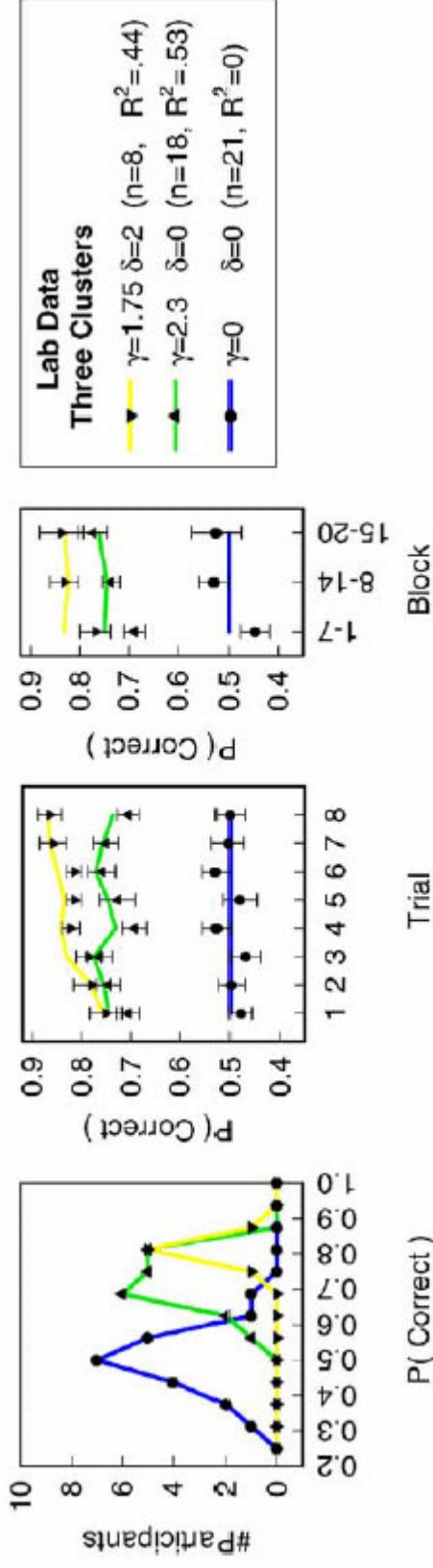
# 7. Modèle complet

Inférence de structures causales (Steyvers et al., 2003: Expérience 1)



Steyvers, M., Tenenbaum, J.B., Wagenmakers, E.-J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, 27, 453-489.

# 7. Modèle complet



Bayésien, optimal

(17%)

Bayésien, 1 essai

(38%)

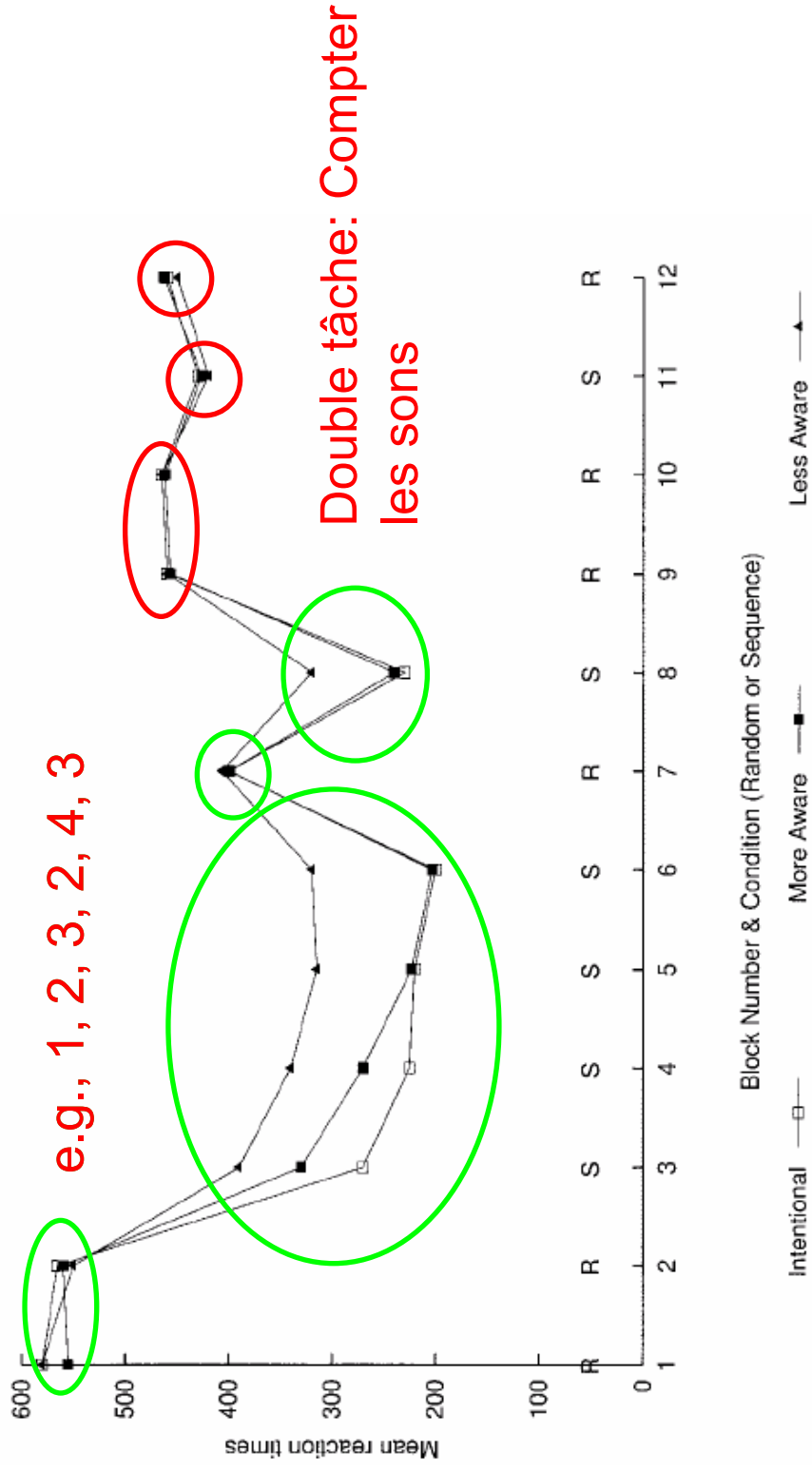
Participants qui n'apprennent pas

(45%)

Steyvers, M., Tenenbaum, J.B., Wagenmakers, E.-J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, 27, 453-489.

# 7. Modèle complet

Apprentissage de séquences (Curran & Keele, 1993: Expérience 1)



Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review*, 112, 159-192.

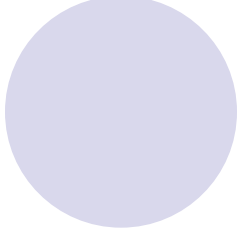
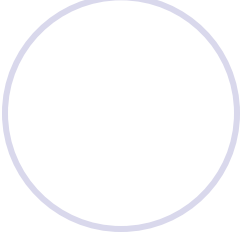
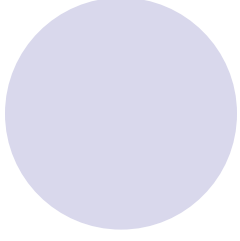
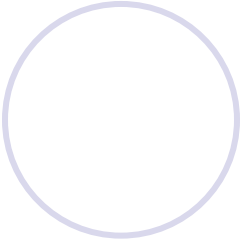
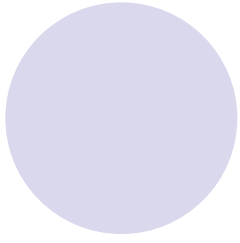




**En bref...**

**Le modèle psychologique final contribuera à expliquer:**

- **La dichotomie qui existe entre les connaissances implicites et explicites.**
- **L'inférence de règles causales à partir de l'information statistique contenue dans l'environnement.**
- **L'interaction entre les connaissances implicites et explicites.**



**Merci!**