



# SYSTÈMES DE SYMBOLES CONTRE CONNEXIONISME : LE DÉBAT SUR LA SYSTÉMATIÇITÉ

SÉANCE 5, ESPRITS NUMÉRIQUES

PHI6385 A24

JONATHAN SIMON

# PLAN

- 1) Introduction: le cœur du débat -- règles contre associations.
- 2) Dreyfus: Savoir faire
- 3) Smolensky: l'alternative connexionniste
- 4) Fodor et Pylyshyn: l'argument de la « systématité » contre le connexionnisme

The background is a dark blue gradient. In the corners, there are white line-art illustrations of circuit traces and nodes. A large, faint, light-blue circular graphic is centered in the upper half of the page.

# INTRODUCTION

# INTRODUCTION

- Règles contre associations (Rationalisme / Nativisme vs Empirisme)
- IA symbolique (classique) : symboles et règles. Rationalisme, déduction, nativisme
- IA connectioniste : nœuds et associations, approximation statistique (correspondance des modèles, ajustement des courbes), empirisme, induction

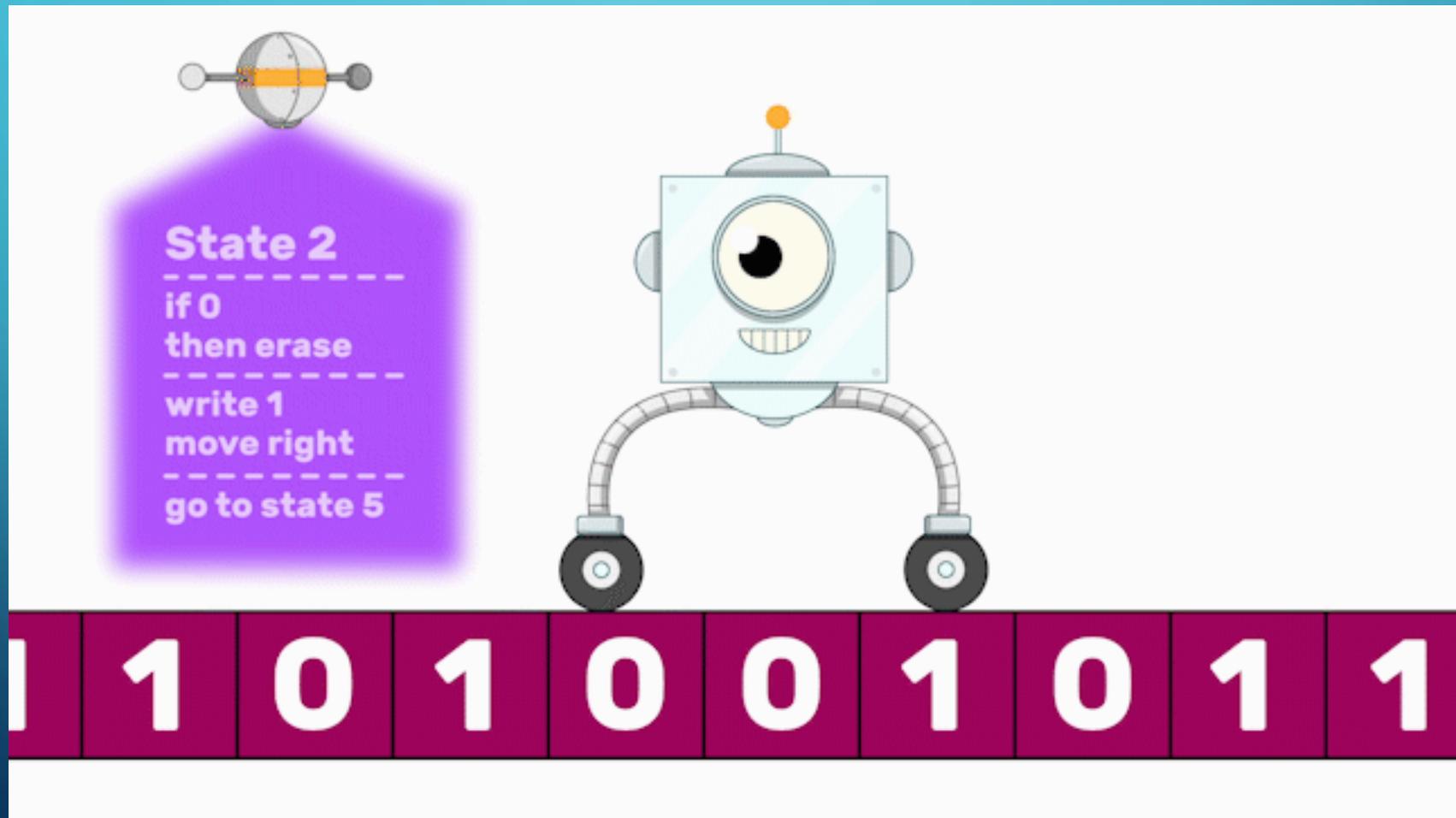
# INTRODUCTION

- Règles contre associations (Rationalisme vs Empirisme)
- IA symbolique (classique) : règles. Rationalisme:
- nous sommes câblés pour faire du raisonnement (logique) basé sur des règles, qui utilise souvent la structure syntaxique des représentations : nous déduisons P de « P et Q » non pas parce que nous avons remarqué (dans notre expérience) que le premier est généralement vrai lorsque le second l'est, mais parce que nous comprenons que cela fait partie de ce que « et » signifie

# L'IA CLASSIQUE

- Systemes de traitement de symboles (GOFAI)
- Les états d'une machine de Turing (ou similaire) représentent directement et explicitement les pensées, les plans, les instructions

# MACHINES DE TURING (RAPPEL)



# MACHINES DE TURING (RAPPEL)

TABLE 1

	$S_1$	$S_2$
Input : pièce de 5 cents	N'émettre aucun output Passer en $S_2$	Emettre un coca-cola Passer en $S_1$
Input : pièce de 10 cents	Emettre un coca-cola Rester en $S_1$	Emettre un coca-cola et une pièce de 5 cents Passer en $S_1$

# L'IA CLASSIQUE

- Systèmes de traitement de symboles (GOFAI)
- Les états d'une machine de Turing représentent directement les pensées, les plans, les instructions
- L'IA précise beaucoup de faits et de règles formelles : une théorie formelle de toute la réalité du sens commun

# L'IA CLASSIQUE

- Systemes de traitement de symboles (GOFAI)
- (Remarque : compris comme un langage machine, ce qui est explicite dans un état n'est toujours qu'une série de 1 et de 0... les règles explicites ou symboliques peuvent être comprises soit comme intégrées dans les règles du tableau machine, soit comme exprimées dans une interprétation de plus haut niveau des 1 et des 0... les symboles composés de 1 et de 0).

# L'IA CLASSIQUE

- Reste en usage aujourd'hui :
- la programmation informatique standard et explicite (par exemple, écrire un programme en Python : les termes que tu utilises sont des symboles, les commandes que tu donnes, et la logique sous-jacente du système, sont les règles...).

```
<!--      <button id ="save1Button" onclick="save1()">SAVE SLOT
1</button>
      <button id ="load1Button" onclick="load1()">LOAD SLOT
1</button><br />
      <button id ="save2Button" onclick="save2()">SAVE SLOT
2</button>
      <button id ="load2Button" onclick="load2()">LOAD SLOT
2</button><br />
      <button id ="resetButton" onclick="reset()">RESET ALL
PROGRESS</button><br />

      <button id ="freeClipsButton" onclick="cheatClips()">Free
Clips</button><br />
      <button id ="freeMoneyButton" onclick="cheatMoney()">Free
Money</button><br />
      <button id ="freeTrustButton" onclick="cheatTrust()">Free
Trust</button><br />
      <button id ="freeOpsButton" onclick="cheatOps()">Free
Ops</button><br />
      <button id ="freeCreatButton" onclick="cheatCreat()">Free
Creativity</button><br />
      <button id ="freeYomiButton" onclick="cheatYomi()">Free
Yomi</button><br />
      <button id ="resetPrestige" onclick="resetPrestige()">Reset
Prestige</button><br />
```

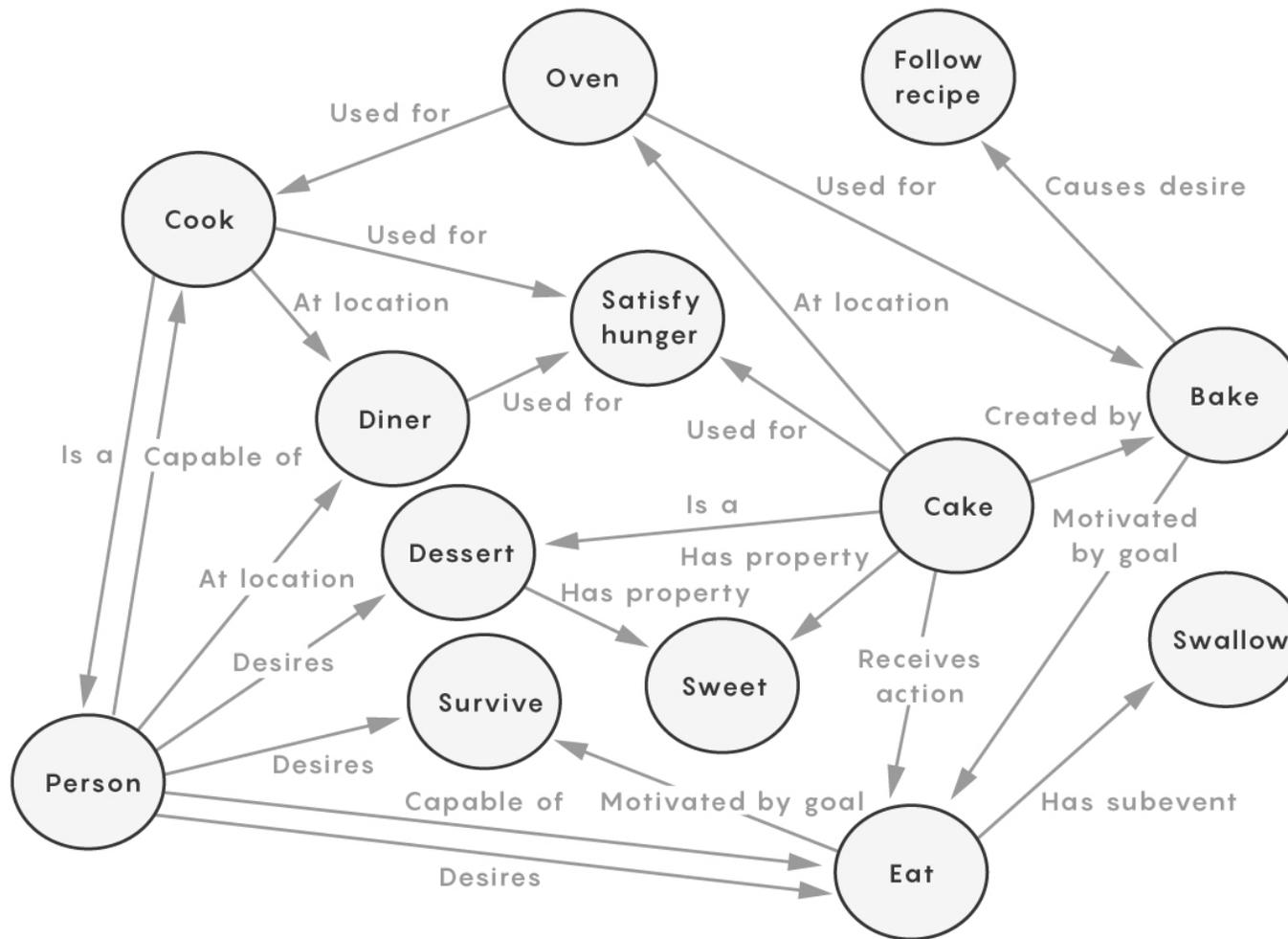
---

```
<button id ="resetProgressAllItemsButton" />
```

# L'IA CLASSIQUE

- Reste en usage aujourd'hui même en IA : les graphes de connaissances (knowledge graphs),
- il existe même une approche à l'IA appelée «ontologie appliqué».

# IA CLASSIQUE / GOFAI





# The Next Generation of Enterprise AI

The Cyc Advantage  
LOGIC-BASED MACHINE REASONING



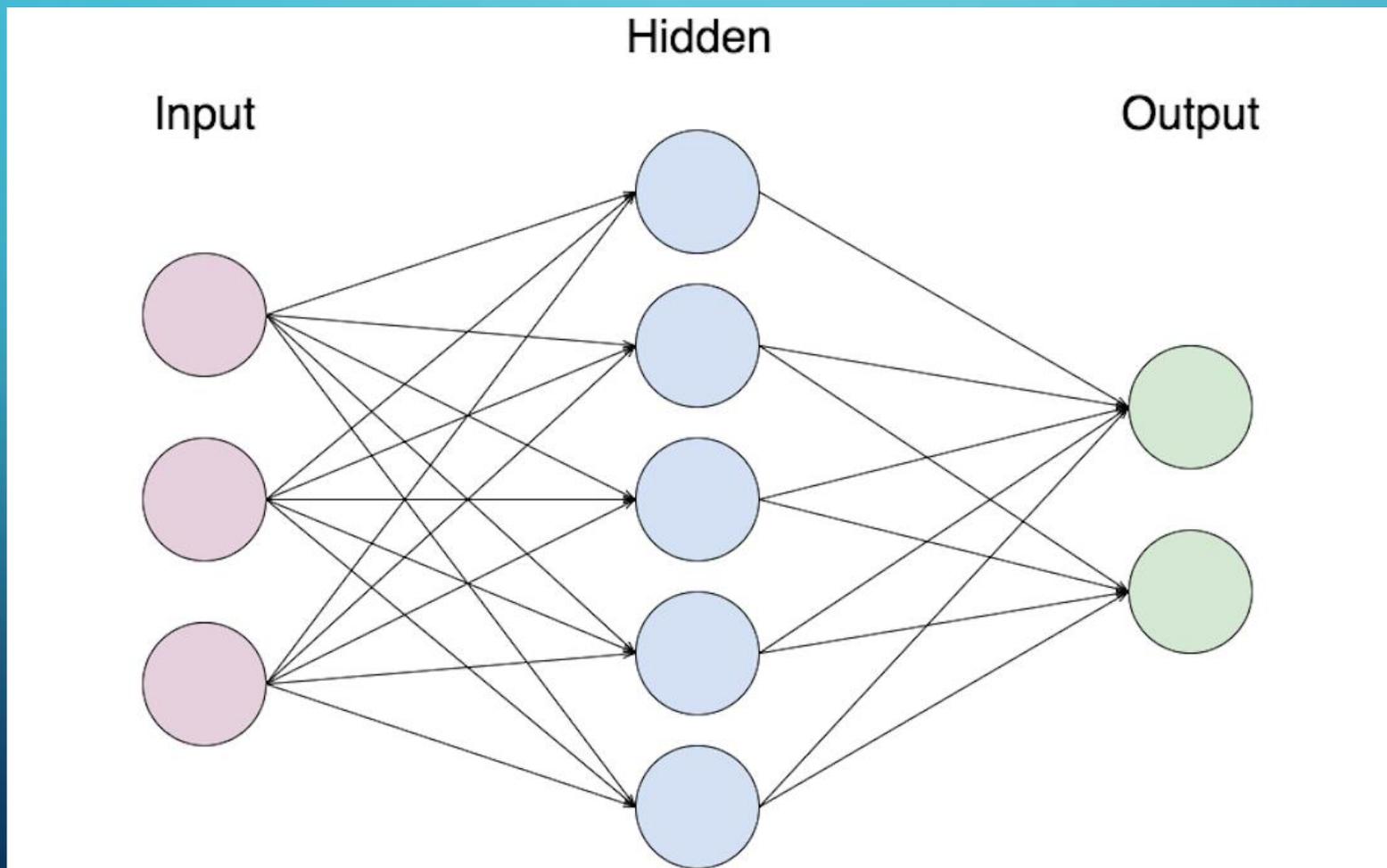
# INTRODUCTION

- Règles contre associations (Rationalisme vs Empirisme)
- IA connexioniste : associations. Empirisme:
- Notre compréhension est basée sur un apprentissage associatif (empirique) sensible au contexte. Les règles logiques « dures et fermes » sont l'exception (plutôt que la règle). En général, nous raisonnons de manière probabiliste et nous comprenons de manière associative.

# CONNECTIONNISME

- L'approche rivale, née dans les années 50 , 60 et 70.
- Réseaux de neurones : structure informatique modelée sur les réseaux de neurones biologiques
- L'unité de calcul de base est modelée sur le tir des neurones (beaucoup d'opérations très simples et subsymboliques - pas de représentation explicite des faits ou des règles au niveau de la pensée humaine, au lieu de cela, ceux-ci sont codés implicitement dans des poids neuronaux)

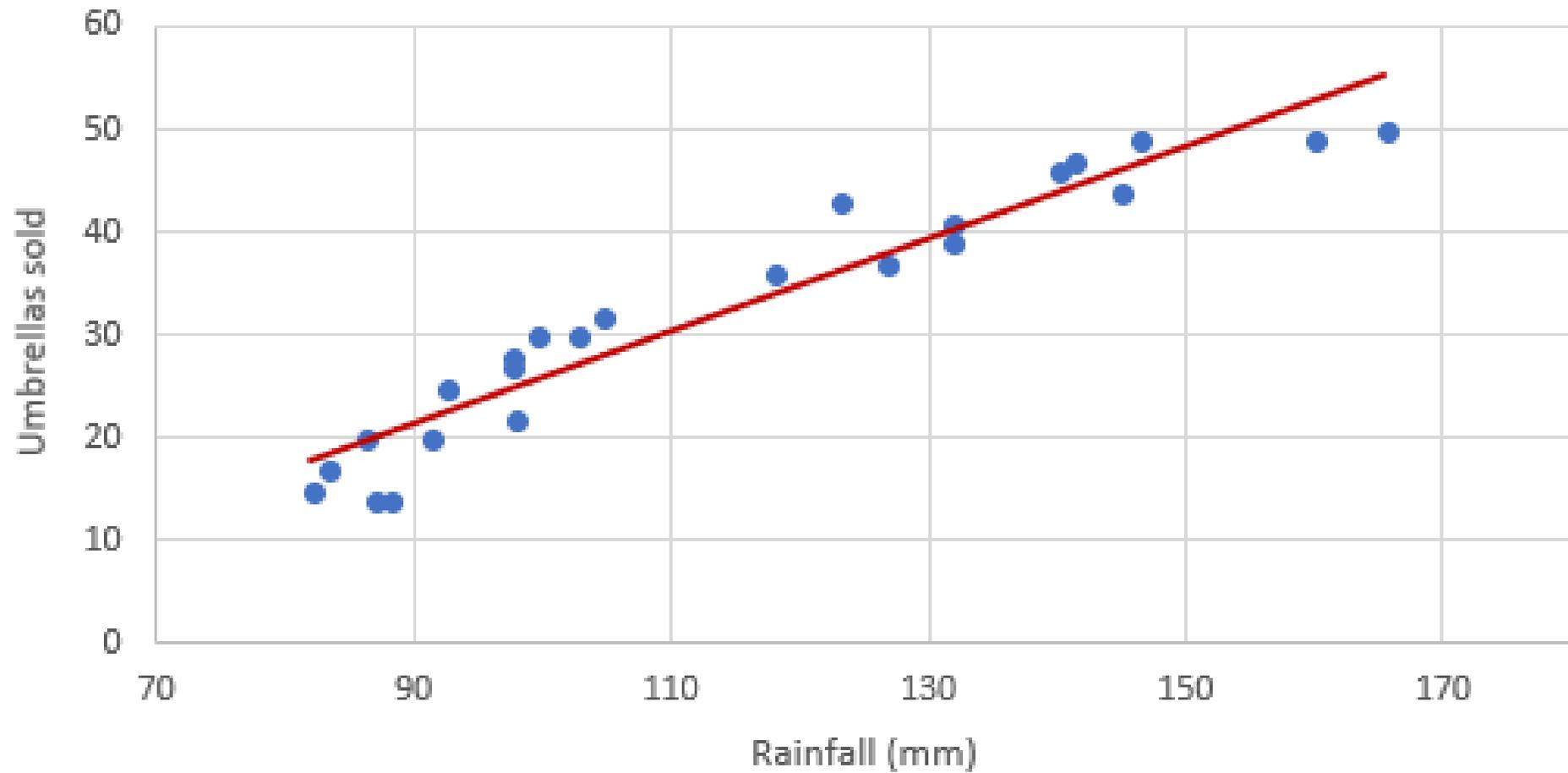
# CONNECTIONNISME



# CONNECTIONNISME

- Ces méthodes sont toutes, en fait, des moyens d'approximer des fonctions (distributions de données). La régression linéaire est la plus simple :

# Linear regression

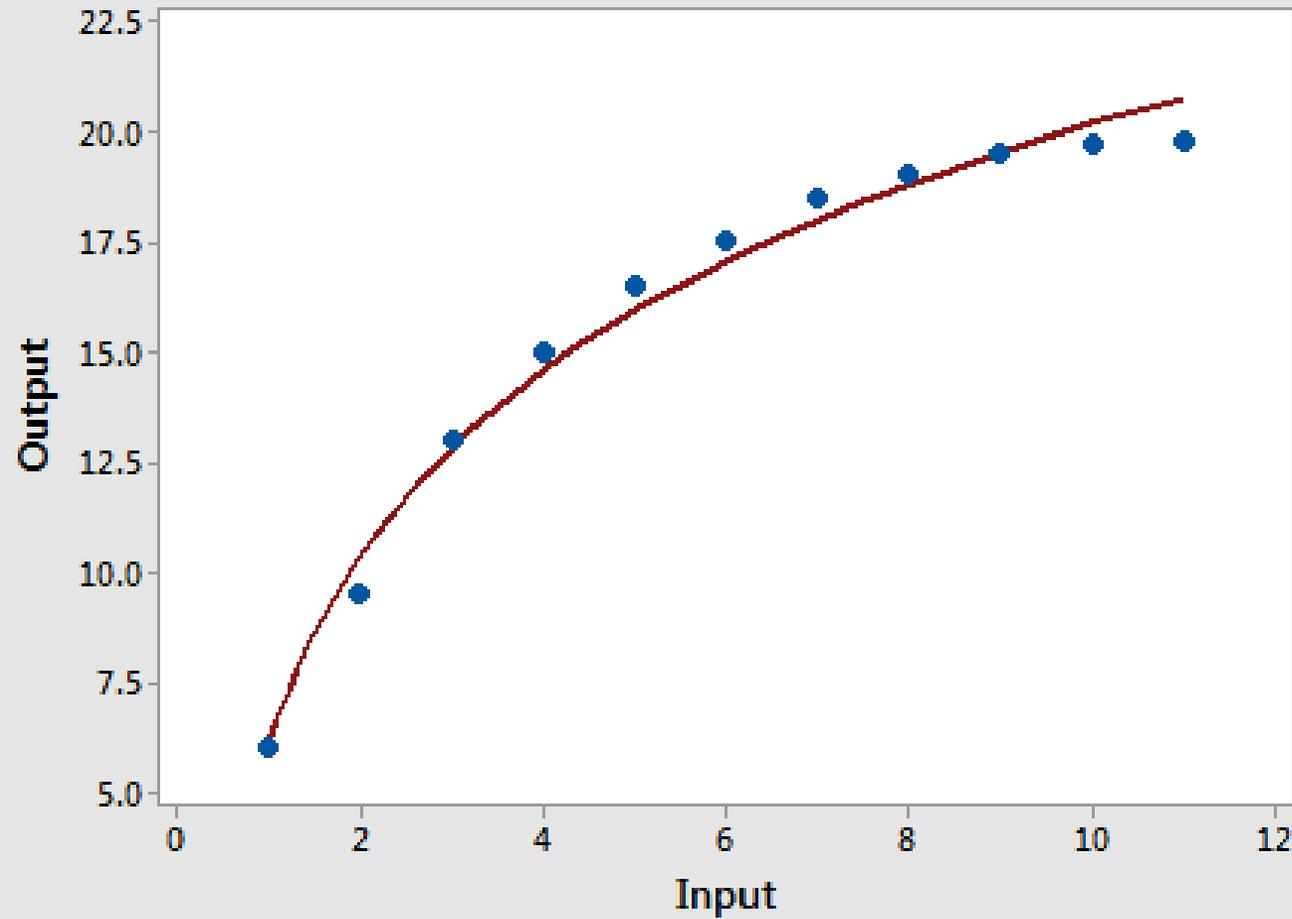


# CONNECTIONNISME

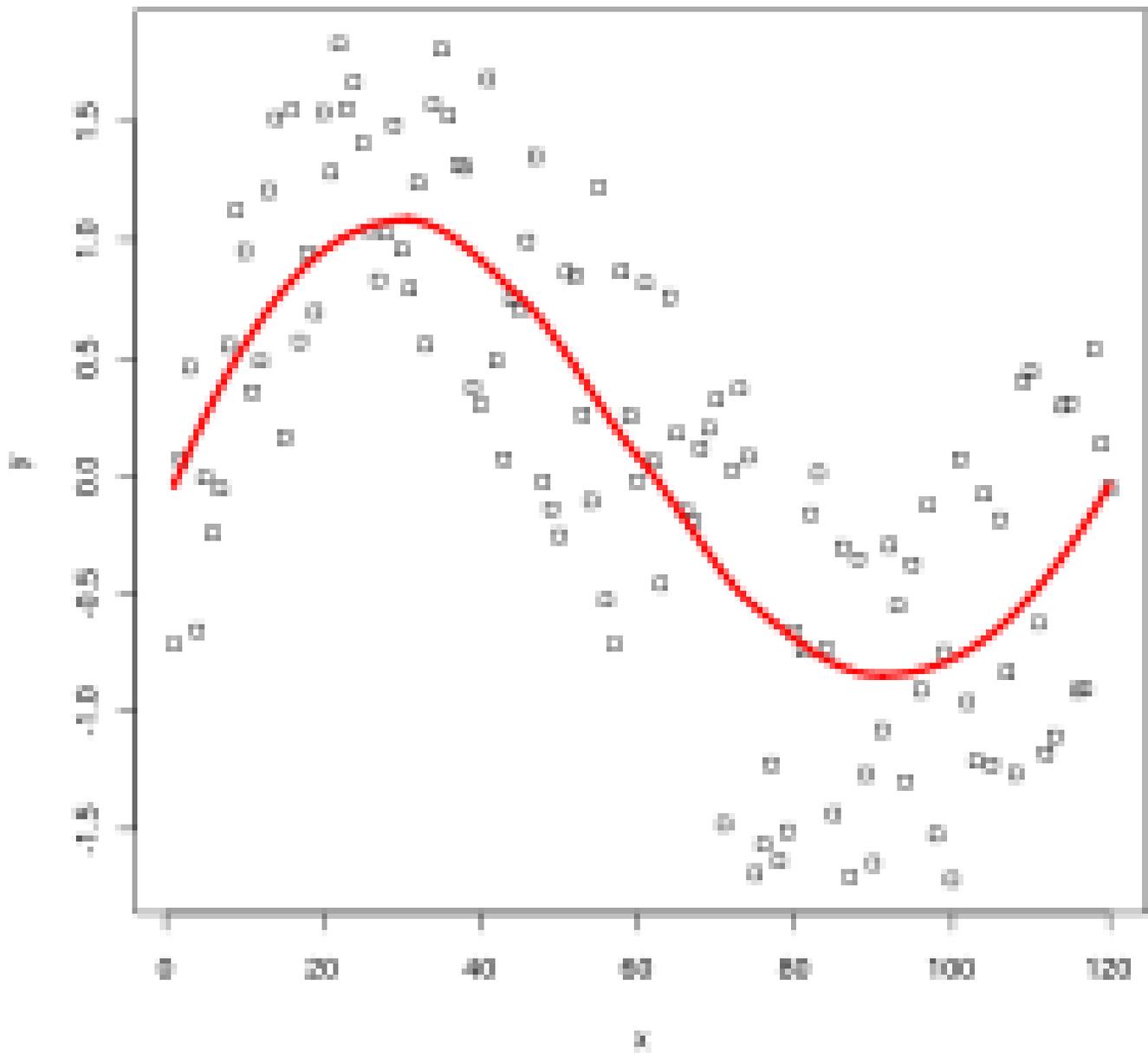
- Plus on ajoute de termes, plus l'approximation peut devenir sophistiquée :

## Fitted Line Plot

$$\text{Output} = 6.099 + 14.06 \log_{10}(\text{Input})$$



S	0.565293
R-Sq	98.6%
R-Sq(adj)	98.5%



# Linear Regression Curve

Daily Chart - E-mini S&P 500 Future (ES)



# CONNECTIONNISME

- Il existe un théorème, le théorème d'approximation universelle, qui dit qu'un réseau neuronal représenté comme ci-dessus peut approximer n'importe quelle fonction : plus on veut que l'approximation soit compliquée, plus on ajoute de neurones...

# CONNECTIONNISME

- Note que tu peux utiliser une telle approximation pour faire des prédictions, c'est-à-dire pour remplir les cas pour lesquels tu n'as pas encore de données

# CONNECTIONNISME

- Notez: il s'agit d'un débat sur le logiciel (*software*), pas sur le matériel (*hardware*):
- le même ordinateur peut mettre en œuvre les deux (et en fait, ton ordinateur le fait !)... cela peut toujours être une machine de Turing : la question est de savoir comment comprendre ses états
- - les «neurones» des réseaux neuronaux ne sont que des représentations numériques, *modélées* sur des neurones.

# CONNECTIONNISME

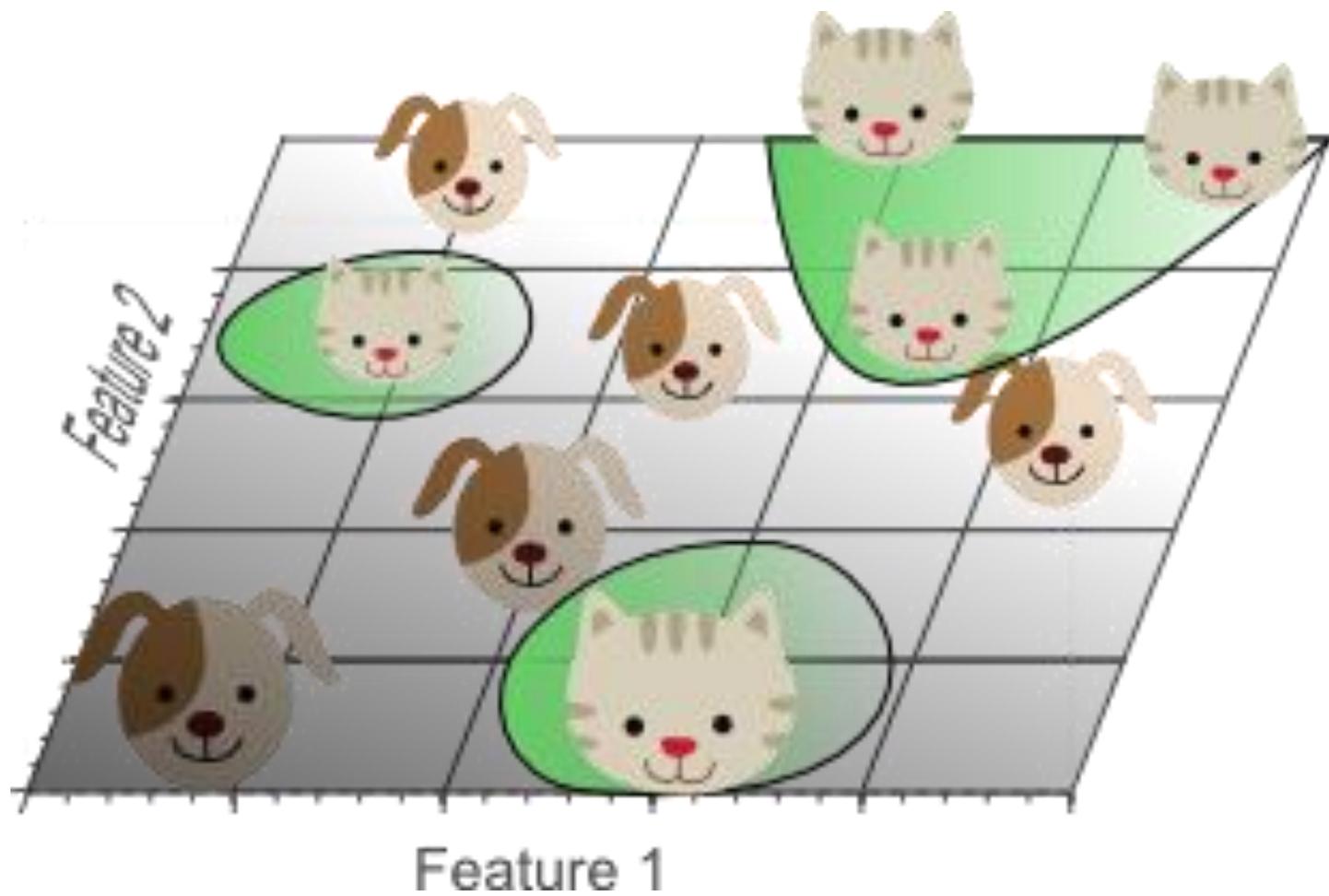
- IA Classique: Les règles du tableau et les états de la machine de bas niveau (au niveau des 1 et des 0), lorsqu'ils sont lus à un niveau supérieur, correspondent à des mots ou à des déclarations et à des règles pour raisonner avec eux.
- Connectionisme: les états internes sont des collections de chiffres (matrices, tenseurs) qui représentent des groupes de neurones (une matrice représente une couche, une ligne de la matrice représente un neurone), et les opérations (généralement la multiplication des matrices) représentent des événements de tir neuronal : des événements où chaque neurone transmet un signal unique aux neurones auxquels il est directement connecté dans la couche suivante

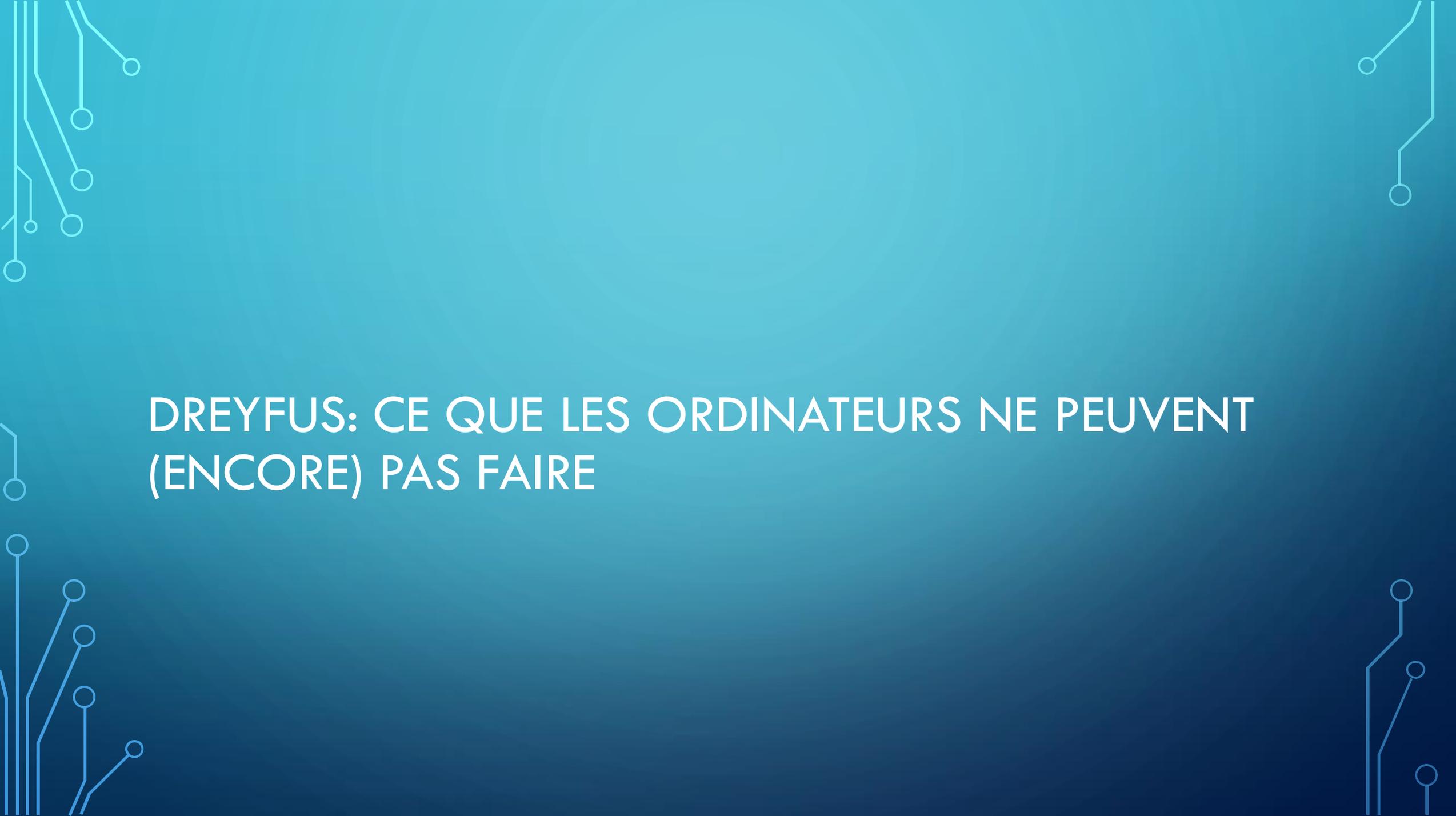
# COMPARER LES DEUX:

- Trier les chats des chiens :
- Supposons que nous ayons des images en pixels que nous voulons alimenter à un système : soit une image de chat, soit une image de chien
- GOFAI: Il faut d'abord des règles pour convertir les motifs de pixels en bords, puis les motifs de bords en formes, puis les motifs de formes en, par exemple, oreille de chien, nez de chat, etc. A la dernière étape, vous avez terminé. Mais essayez de penser à une règle que vous pourriez écrire explicitement, pour déduire quand une forme est en forme d'oreille de chien plutôt que de chat...

# COMPARER LES DEUX:

- Trier les chats des chiens :
- Supposons que nous ayons des images en pixels que nous voulons alimenter à un système : soit une image de chat, soit une image de chien
- Réseaux neuronaux : vous l'entraînez sur beaucoup, beaucoup d'images (pour lesquelles vous lui dites si c'est un chien ou un chat). Ensuite, il ajuste ses poids (les valeurs exactes dans les matrices), en ajustant une frontière très complexe (côté chien, côté chat) dans un espace de représentation de valeurs de pixels. Ensuite, lorsque vous lui montrez une nouvelle image, il repère simplement de quel côté de la limite il se trouve.



The background is a solid teal color with a subtle gradient. In the four corners, there are decorative white line-art elements resembling circuit traces or neural network connections, with small circles at the end of the lines.

# DREYFUS: CE QUE LES ORDINATEURS NE PEUVENT (ENCORE) PAS FAIRE

# HUBERT DREYFUS

- Dreyfus: dans la tradition de Heidegger et Merleau-Ponty
- A commencé à critiquer l'IA classique pendant les années 60 et 70, bien avant que ce soit populaire (de le critiquer)

# HUBERT DREYFUS

- Cette deuxième édition (avec sa nouvelle introduction) a été écrite en 1992, à la fin de la première vague d'IA classique, début de la première montée du connectionisme.
- Beaucoup de ses critiques préfigurent (ou même motivent) le connectionnisme, mais il critique aussi le connectionnisme (et ses critiques des deux écoles restent valables aujourd'hui !)

# HUBERT DREYFUS

- Critiques primaires :
- l'intelligence humaine est une question de savoir-faire pratique, et ce savoir-faire pratique n'est pas équivalent à une quelconque quantité de *connaissances-que* (connaissance des faits)

# HUBERT DREYFUS

- Si un expert doit expliquer au système ce qui est pertinent / approprié, le système ne sera jamais pleinement compétent / adaptable
- Lorsque nous apprenons quelque chose de nouveau, comme les échecs, nous appliquons des règles qu'un expert nous a expliquées. Mais à mesure que nous acquérons de l'expérience, nous apprenons à improviser, à adapter et à contourner les règles. Pour les vrais experts, la bonne chose à faire est simplement évidente, ce n'est pas une question de raisonnement...

The background is a solid teal color with a subtle gradient. In the corners, there are decorative white line-art elements resembling circuit traces or neural network connections, with small circles at the end of the lines.

# SMOLENSKY: LA CONCEPTION CONNEXIONNISTE

# SMOLENSKY

- Le processeur (des règles) consciente(s) et le processeur intuitif:
  - Smolensky distingue deux capacités cognitives ("machines virtuelles") : le "processeur conscient" (l'interprète conscient des règles) et le "processeur intuitif".

# SMOLENSKY

- Processeur conscient: . « Les règles exprimées dans le langage naturel peuvent fournir une formalisation satisfaisante du savoir culturel. L'application consciente d'une règle peut être modélisée comme une interprétation séquentielle de telles règles par une machine virtuelle appelée «l'agent interprétant consciemment une règle». Ces règles sont formulées en termes de concepts consciemment utilisés afin de décrire le domaine de la tâche - elles sont formulées à un niveau conceptuel .»

# SMOLENSKY

- Au début, Smolensky ne conteste pas cette image (que la cognition linguistique, qui suit explicitement des règles, est classique : la manipulation syntaxique de symboles, avec des axiomes et des règles spécifiques à un domaine)

# SMOLENSKY

- L'alternative de Smolensky est (d'abord) seulement une alternative à la conception classique *du processeur intuitif* : «Ce dernier est vraisemblablement responsable de l'ensemble des comportements animaux et d'une très grande partie des comportements humains : la perception, la motricité automatique, la pratique langagière courante, l'intuition dans la résolution de problèmes et le jeu -bref, pratiquement toutes les performance demandant des habiletés.»

# SMOLENSKY

- Il continue: «Le transfert de responsabilités d'un agent interprétant consciemment une règle au processeur intuitif durant l'acquisition d'habiletés s'avère l'un des phénomènes les plus frappants et les plus étudiés par les sciences cognitive». (pense à Dreyfus)
- La cible primaire de Smolensky est l'affirmation selon laquelle le processeur intuitif fonctionne de la même manière que l'interprète conscient des règles

# SMOLENSKY

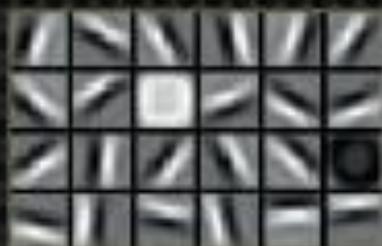
- Autrement dit, Smolensky rejette: «L'hypothèse de l'interprétation inconsciente d'une règle: Les programmes fonctionnant au sein d'un processeur intuitif ont une syntaxe et une sémantique comparables à celles des programmes gérés par un agent interprétant consciemment les règles.»

# SMOLENSKY

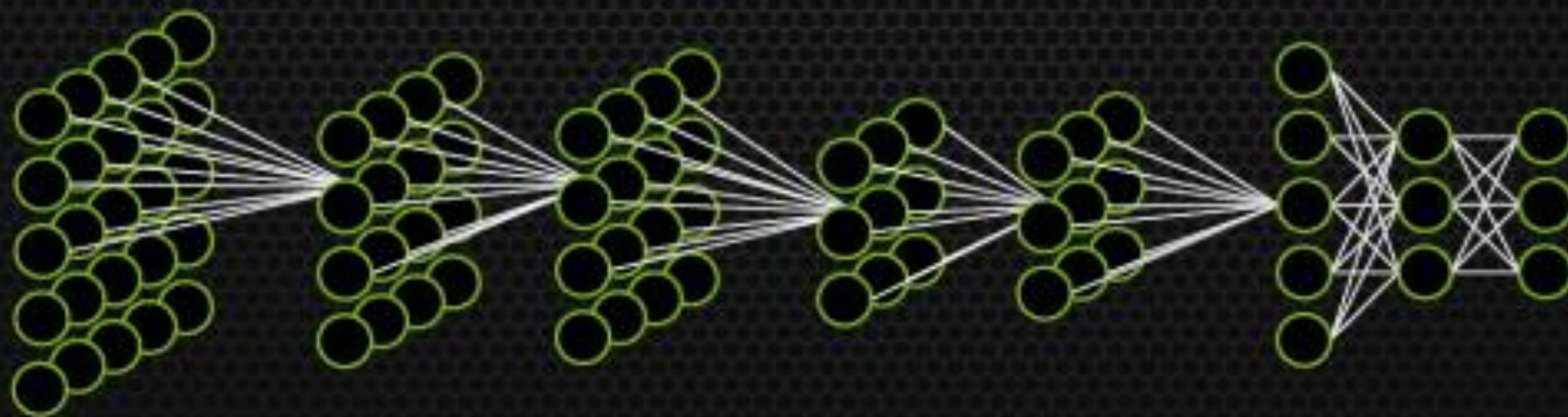
- En sa place, la vision positive:

# SMOLENSKY

- La vision positive:



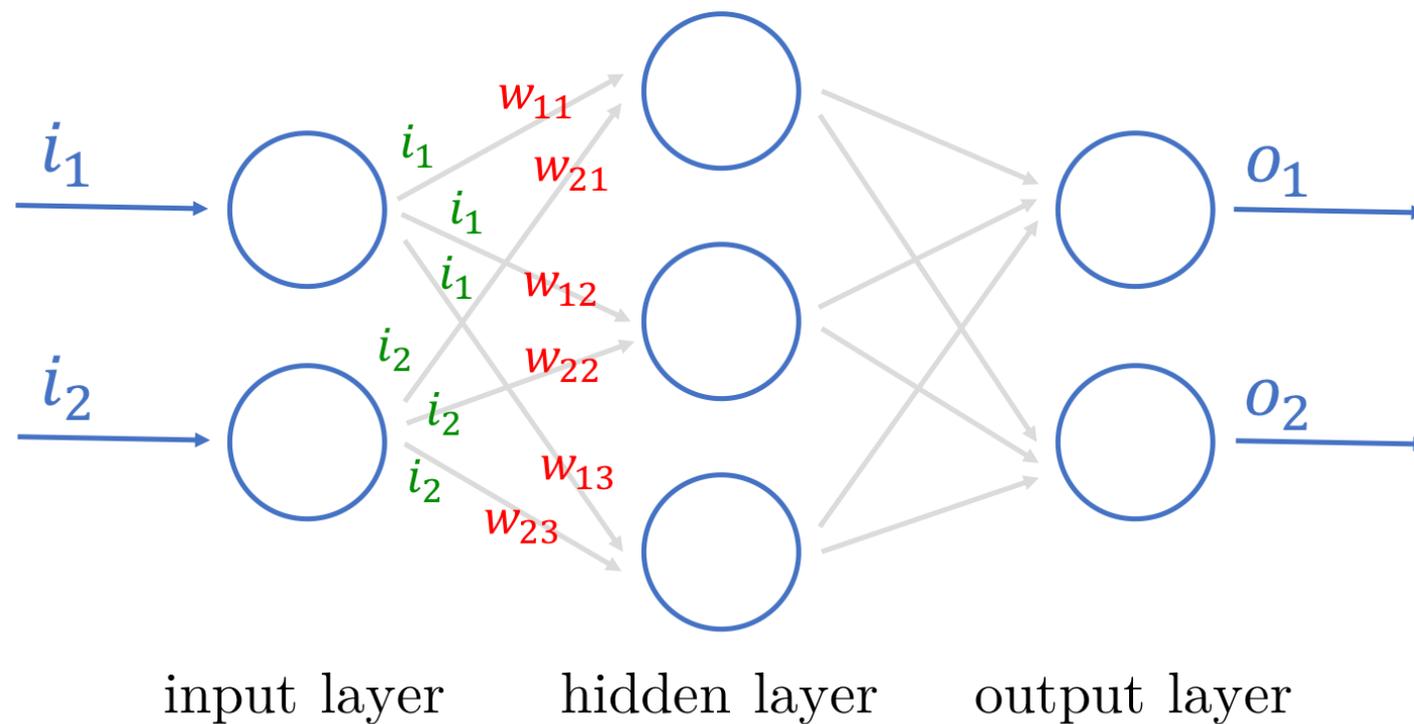
Image



“Sara”

# SMOLENSKY

- La vision positive:



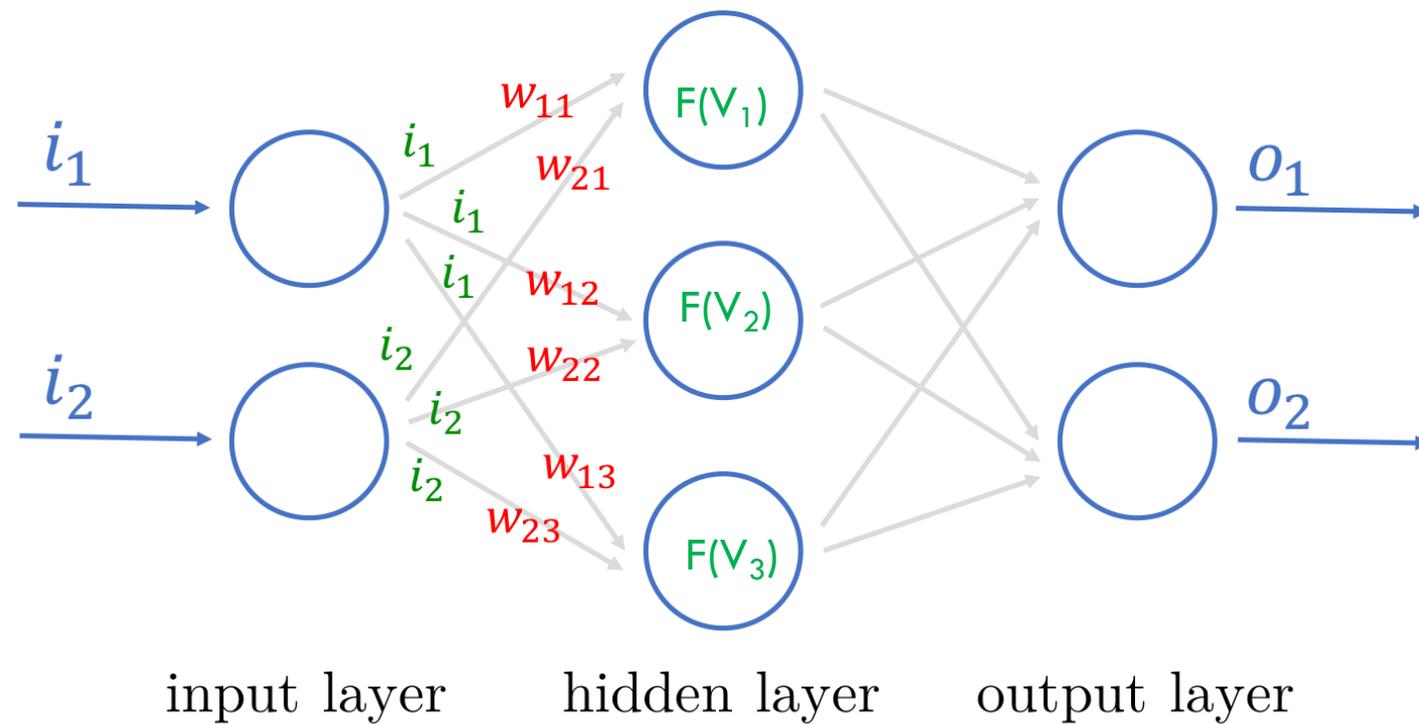
# SMOLENSKY

- La vision positive:

$$\begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \\ w_{13} & w_{23} \end{bmatrix} \cdot \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} (w_{11} \times i_1) + (w_{21} \times i_2) \\ (w_{12} \times i_1) + (w_{22} \times i_2) \\ (w_{13} \times i_1) + (w_{23} \times i_2) \end{bmatrix} \begin{matrix} =V_1 \\ =V_2 \\ =V_3 \end{matrix}$$

# SMOLENSKY

- La vision positive:



# SMOLENSKY

- «L'hypothèse du système dynamique connexionniste: À n'importe quel moment, l'état du processeur intuitif est précisément défini par un vecteur de valeurs numériques ( une pour chacune des unités). La dynamique du processeur intuitif est régie par une équation différentielle. Les paramètres numériques de cette équation constituent le programme ou le savoir du processeur. Dans les systèmes d'auto-apprentissage ces paramètres changent en fonction d'une autre équation différentielle.»

# SMOLENSKY

- C'est juste une façon de caractériser l'algèbre matricielle que nous utilisons finalement pour décrire les réseaux de neurones

# SMOLENSKY

- Niveaux d'analyse (des systèmes cognitifs) : conceptuel, sous-conceptuel
- Les paradigmes de l'analyse (des systèmes cognitifs) : symbolique, sous-symbolique

# SMOLENSKY

- Dans ces débats, les « niveaux d'explication » sont remis en question. Il existe différents modèles de calcul : celui d'une machine de Turing /Von Neumann, contre celui d'un réseau neuronal parallèle et distribué.
- Mais vous pouvez utiliser l'un ou l'autre pour mettre en œuvre l'autre «virtuellement». Vous pouvez faire des mathématiques matricielles sur une machine de Turing, donc vous pouvez y implémenter n'importe quel réseau de neurones.

# SMOLENSKY

- La vision positive:

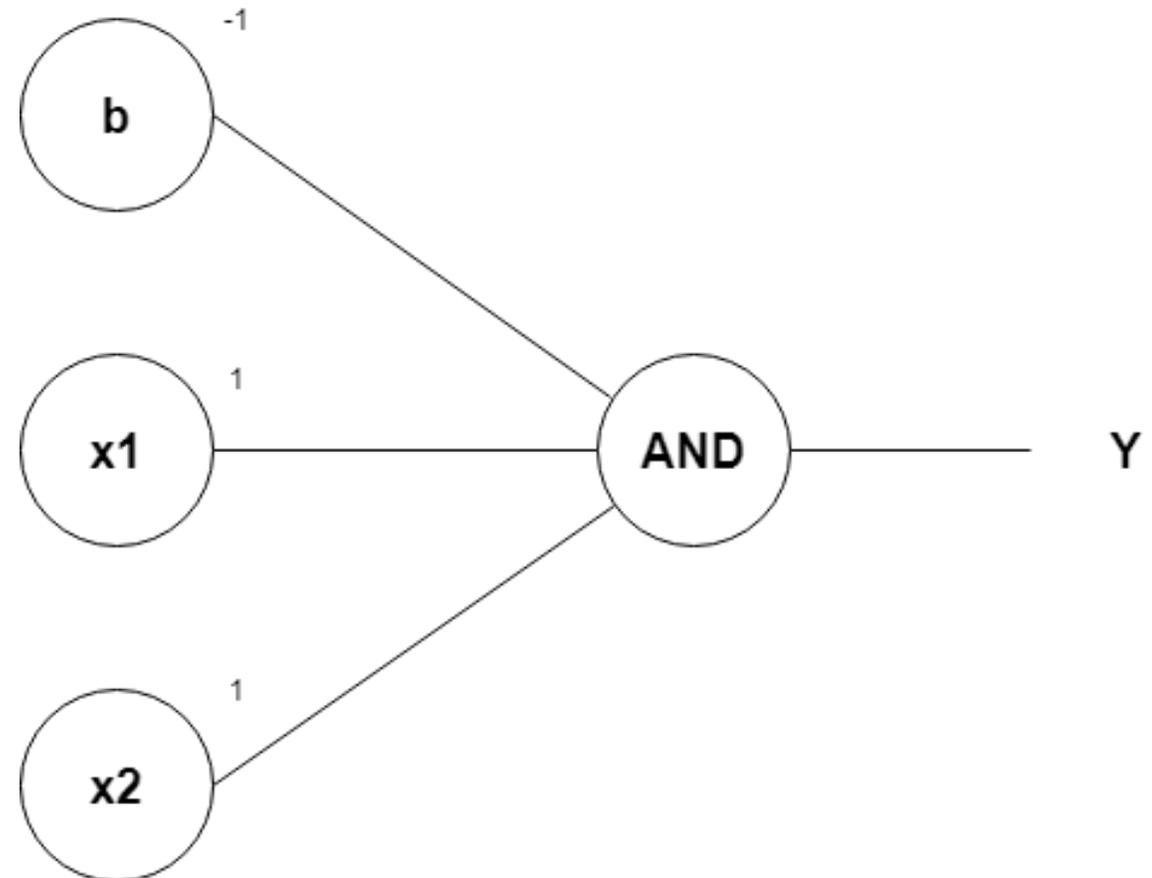
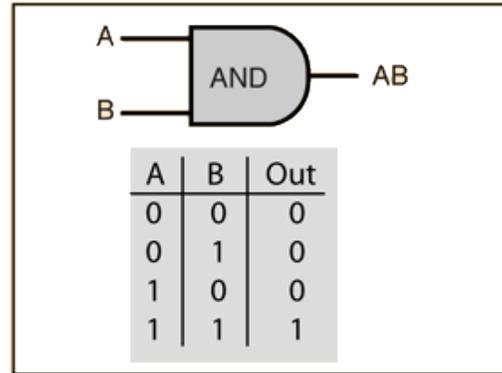
$$\begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \\ w_{13} & w_{23} \end{bmatrix} \cdot \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} (w_{11} \times i_1) + (w_{21} \times i_2) \\ (w_{12} \times i_1) + (w_{22} \times i_2) \\ (w_{13} \times i_1) + (w_{23} \times i_2) \end{bmatrix}$$

## SMOLENSKY

- Vous pouvez faire des mathématiques matricielles sur une machine de Turing, donc vous pouvez y implémenter n'importe quel réseau de neurones.
- Et vous pouvez implémenter une machine de Turing, des portes logiques, etc., en utilisant un véritable réseau neuronal distribué en parallèle.

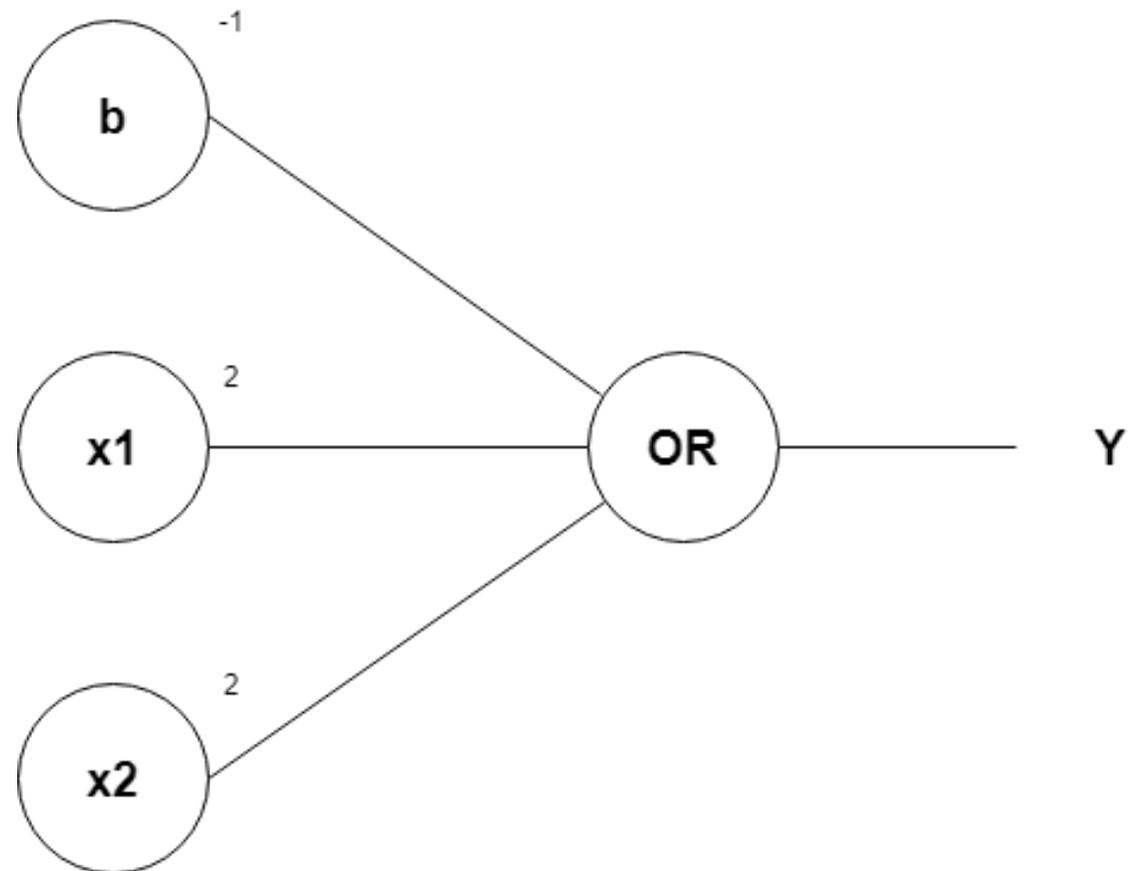
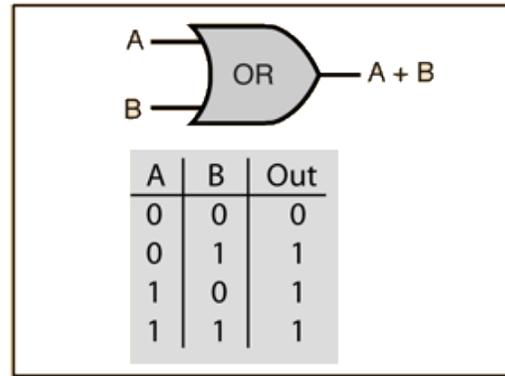
# SMOLENSKY

- “et” (AND)



# SMOLENSKY

- “ou” (OR)



# SMOLENSKY

- Mais, comme le note Smolensky, il n'y aurait pas vraiment d'alternative connexionniste, si les réseaux connexionnistes se contentent de mettre en œuvre des systèmes véritablement classiques.
- L'affirmation principale de Smolensky est qu'il n'existe pas de compte rendu complet et précis de la cognition au niveau conceptuel. Pour comprendre le processeur intuitif, nous devons nous tourner vers le traitement distribué au niveau subconceptuel. L'histoire qui en résulte au niveau conceptuel ne sera qu'une approximation (comme la mécanique newtonienne classique n'est qu'une approximation de la théorie relativiste ou quantique)

# SMOLENSKY

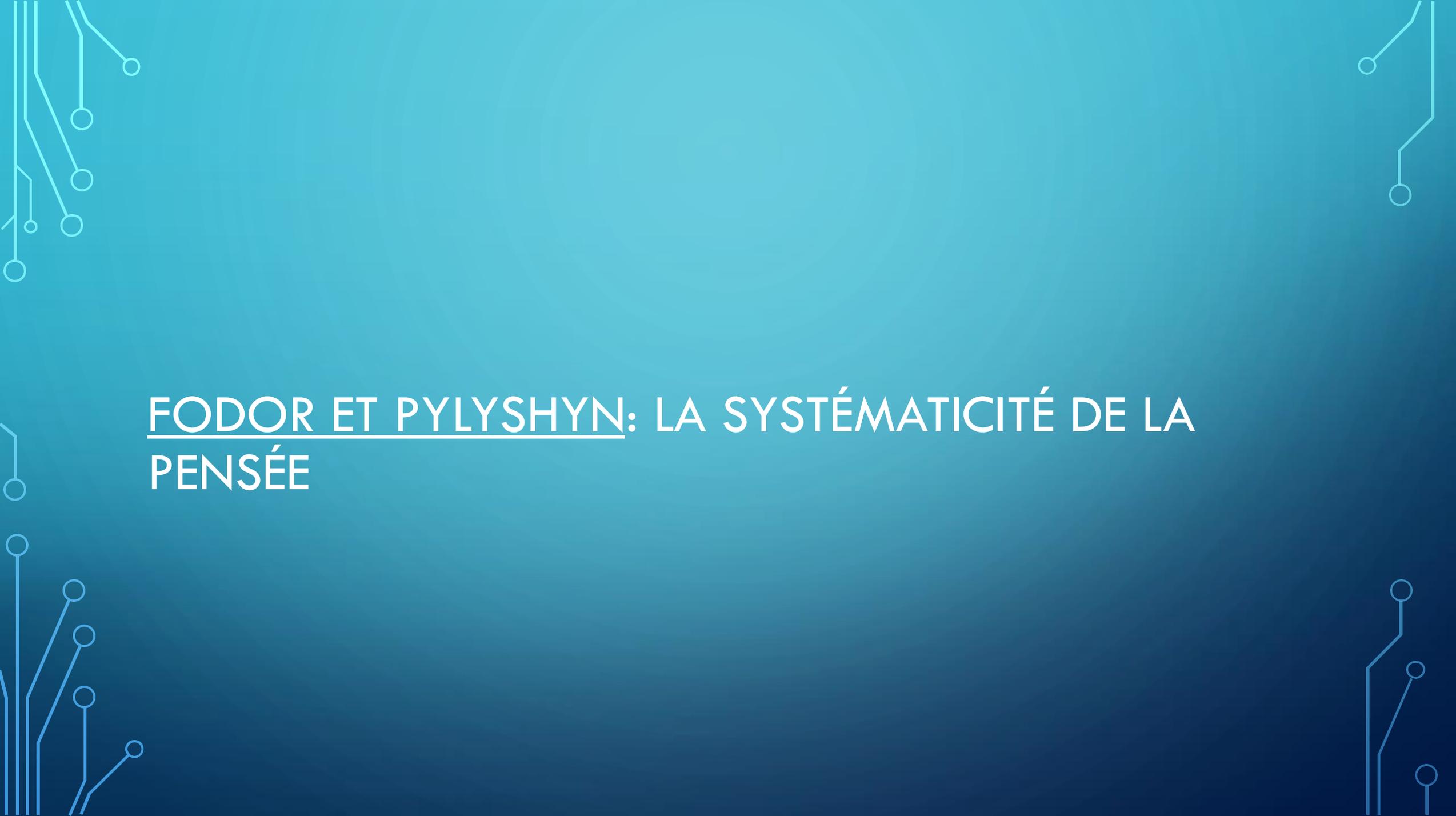
- Notez qu'à ce stade, Smolensky ne parle que de l'activité cognitive du processeur intuitif. L'interprète de règles conscient pourrait encore être parfaitement classique. Ou pas...

# SMOLENSKY

- (En fait, cela fait encore l'objet de débats aujourd'hui. Certains, comme Gary Marcus, disent que le «processeur conscient» doit être symbolique et basé sur des règles, et au mieux seulement neuronal dans sa mise en œuvre (qui doit être strictement et non seulement approximativement basée sur des règles), tandis que d'autres (comme Bengio et LeCun) disent que le processeur conscient sera plus comme le processeur intuitif, seulement une approximation d'un système basé sur des règles...)

# SMOLENSKY

- Si nous laissons les choses ici, Fodor et Pylyshyn n'auraient rien à objecter. Mais comme le note Smolensky, s'il y avait vraiment deux systèmes distincts, il n'y aurait aucun compte rendu sur la façon de relier les deux, sur la façon dont les connaissances passent de l'interprétation des règles à l'intuition, sur la façon dont les concepts que nous utilisons lors de l'interprétation des règles sont fondés sur l'intuition, etc. Il suggère donc qu'un compte rendu subsymbolique du processeur intuitif pourrait conduire à un compte rendu subsymbolique du processeur conscient. C'est là qu'interviennent Fodor et Pylyshyn.

The background is a solid teal color. In the four corners, there are decorative white line-art patterns resembling circuit boards or neural networks, with lines connecting to small circles.

# FODOR ET PYLYSHYN: LA SYSTÉMATIÉ DE LA PENSÉE

# FODOR ET PYLYSHYN

- L'idée centrale : la pensée (et le langage) sont productifs, systématiques et compositionnels.
- Pour que ces caractéristiques soient réellement mises en œuvre, il doit y avoir une structure de partie et d'ensemble parmi les éléments syntaxiques qui représentent les pensées, et cette structure doit être causalement pertinente et aider à expliquer les processus de pensée
- Mais les systèmes connectionnistes n'ont pas ce genre de structure partie-et-ensemble (au niveau cognitif)

# FODOR ET PYLYSHYN

- Prémisse 1) La pensée (et le langage) sont productifs, systématiques et compositionnels.
  - Le débat ici : la pensée a-t-elle vraiment ces caractéristiques, ou ne s'agit-il que d'approximations utiles ?

# FODOR ET PYLYSHYN

- Prémisse 1) La pensée (et le langage) sont productifs, systématiques et compositionnels.
- Prémisse 2) Si la pensée est productif, systématique et compositionnel, il a une structure (causalement pertinent) de partie et d'ensemble
- Prémisse 3) Mais les systèmes connexionnistes n'ont pas une structure (causalement pertinent) de partie et d'ensemble
- Conclusion) Donc, la pensée n'est pas connexionniste

# FODOR ET PYLYSHYN

- Prémisse 2) Si la pensée est productif, systématique et compositionnel, il a une structure (causalement pertinent) de partie et d'ensemble
  - Le débat ici : y a-t-il des moyens (connectionnistes) de réaliser ces caractéristiques, sans une structure (causalement pertinent) de partie et d'ensemble?

# FODOR ET PYLYSHYN

- Prémisse 3) Mais les systèmes connexionnistes n'ont pas une structure (causalement pertinent) de partie et d'ensemble
  - Le débat ici : est-il vrai qu'aucun système connexionniste n'a la structure adéquate (sauf s'il s'agit en fait d'une simple mise en œuvre d'une architecture classique) ?

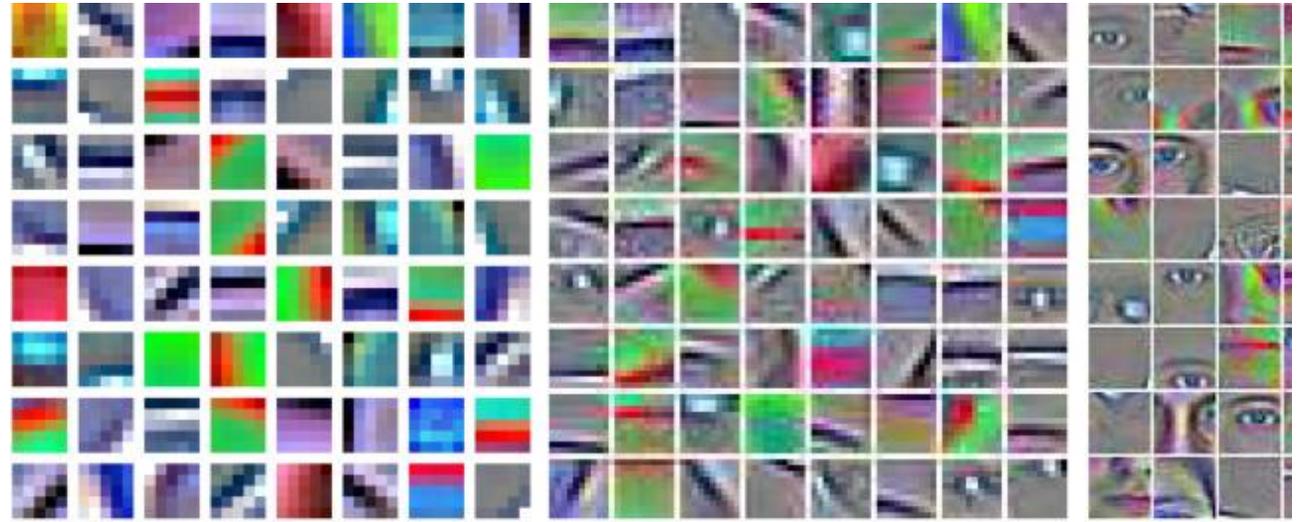
# FODOR ET PYLYSHYN

- Qu'est-ce qu'on veut dire par « une structure (causalement pertinent) de partie et d'ensemble » ?
- Dans le traitement classique, les contenus comme "vache brune" ou "vache et brune" sont représentés par des symboles composés. Ils ont des parties. Le symbole de la "vache" est une partie du symbole de la "vache brune". Les règles de traitement classique peuvent fonctionner sur cette structure, c'est-à-dire que si vous pensez que "P et Q", alors concluez que "P" peut être mis en œuvre de manière syntaxique : si vous acceptez un contenu qui est un composé de deux contenus entourant le symbole "et", alors acceptez le contenu atomique à gauche.

# FODOR ET PYLYSHYN

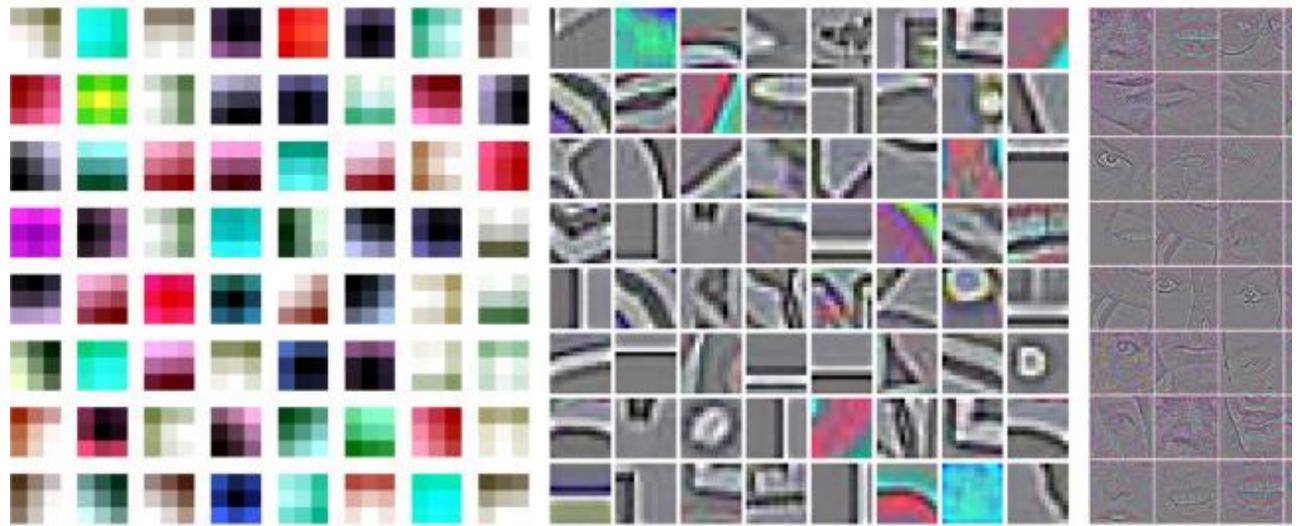
- Qu'est-ce qu'on veut dire par « une structure (causalement pertinent) de partie et d'ensemble » ?
- Dans le traitement connexionniste, Les contenus sont représentés par des modèles de tir de neurones ou de groupes de neurones. L'idée étant que le neurone ou l'amas de neurones tire beaucoup pour signaler la présence de ce qu'il représente.
- En général, ce n'est pas parce que le neurone N1 signale la vache et le N2 le marron que l'on peut raisonner efficacement avec la "vache marron", mais parce qu'il faut un autre neurone N3 pour le signaler. (donc pas de structure de composition automatique). C'est précisément ce que font les Deep Nets : ils ajoutent des couches supplémentaires de neurones pour représenter plus de contenu.

# FODOR ET PYLYSHYN



(a)

(b)

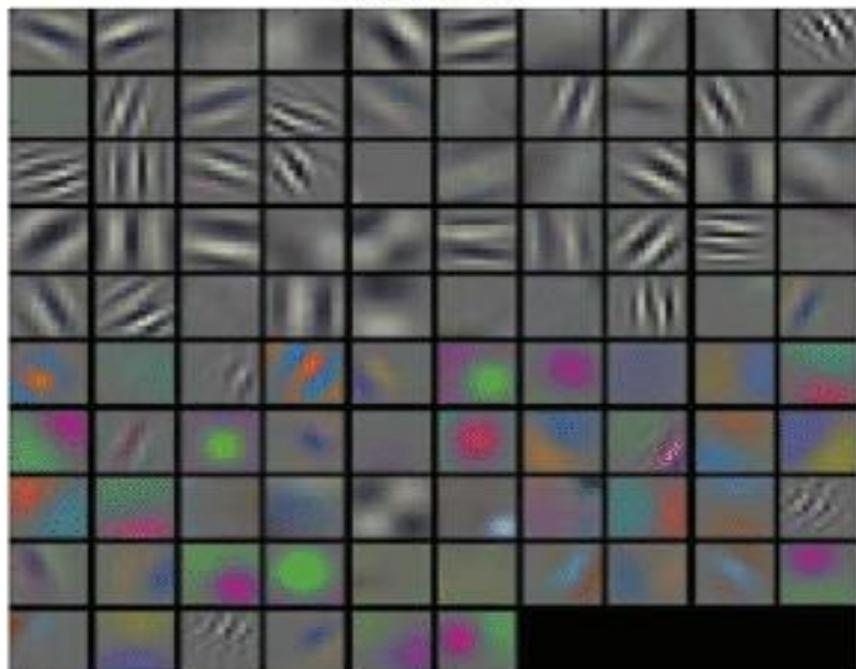


(d)

(e)

ImageNet-CNN

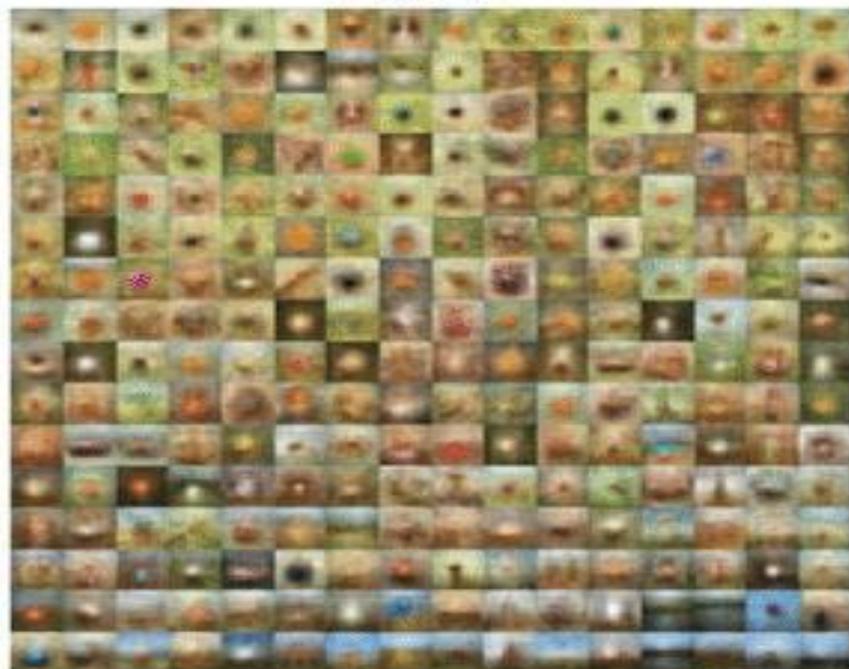
Conv 1



Pool 2



Pool 5

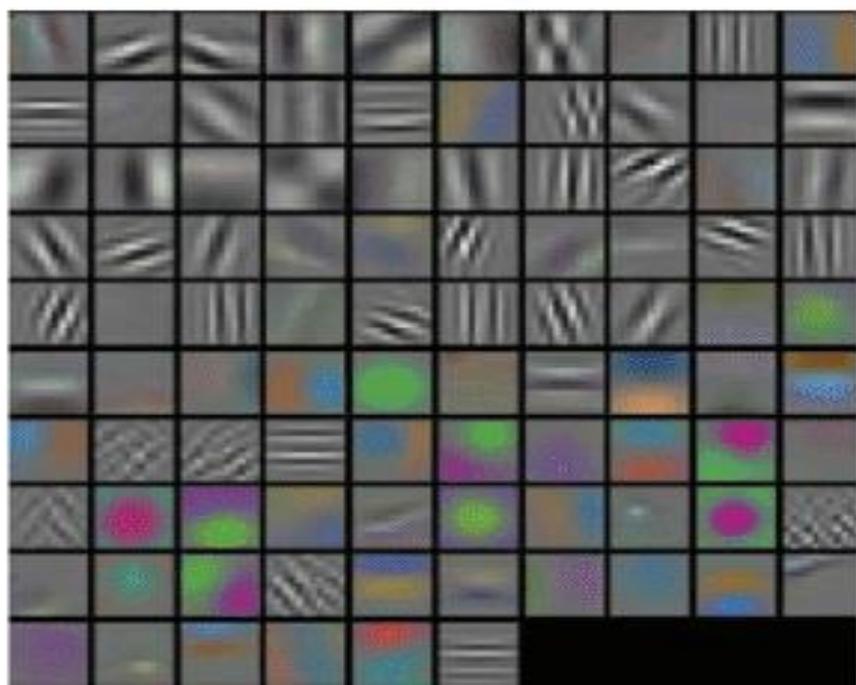


FC 7



Places-CNN

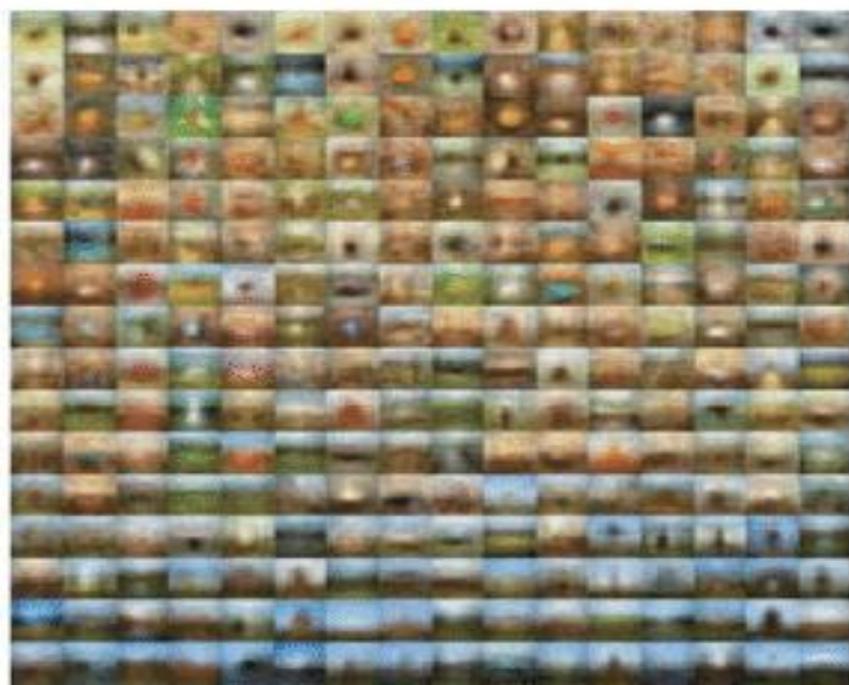
Conv 1



Pool 2



Pool 5



FC 7



# FODOR ET PYLYSHYN

- La pensée (et le langage) sont productifs, systématiques et compositionnels:
- Productivité : nous pouvons générer et comprendre des phrases arbitrairement longues
- Systématicité : si nous comprenons une phrase en utilisant les mots A, B et C, nous pouvons comprendre toutes les phrases utilisant seulement ces mots. (Si vous comprenez «Jean aime Marie», vous pouvez comprendre «Marie aime Jean»)
- Compositionnalité : les mots apportent la même contribution au sens indépendamment de leur contexte

# FODOR ET PYLYSHYN

- La productivité et la compositionnalité sont controversées. Peut-on vraiment comprendre des phrases de 400 mots ?
- Les mots ont-ils vraiment la même signification dans tous les contextes ?  
(grain de café, tasse de café, plante de café...)

# FODOR ET PYLYSHYN

- Il est beaucoup plus difficile de nier le systematicité
- si je comprends "Jean aime Marie", je peux comprendre "Marie aime Jean".
- Mais attention : même si nous admettons que le tir du neurone pour le brun et le tir pour la vache constituent une représentation composite de la vache brune, nous avons ici un problème : considerez: «Jean aime Marie»: comment distinguer entre cela et «Marie aime Jean»)

# FODOR ET PYLYSHYN

- Nous pouvons ajouter des "marqueurs de sujet" : John-sujet aime Mary-objet. Mais maintenant, nous avons besoin de différents neurones (ou clusters) pour représenter le sujet-John et l'objet-John. Et que dire de "Jean aime Marie et Susan aime Jean et Marie aime Susan" (chaque personne est à la fois sujet et objet mais cela ne suffit pas..)

# FODOR ET PYLYSHYN

- Réponse possible : appel à la "synchronisation neurale". Permettre que chaque neurone ait quelques créneaux temporels différents. Laissez le neurone John décider de tirer pendant le cycle un ou le cycle deux ou le cycle trois. Laissez différents neurones tirer sur le même cycle pour coordonner dans le sens...
- Question : cela ne ferait-il que mettre en place une structure classique ?

# FODOR ET PYLYSHYN

- Question : y a-t-il un moyen pour un système connexionniste d'avoir une structure partielle et entière, sans se contenter de mettre en place une structure classique (que dirait Smolensky ?)