



SYSTÈMES DE SYMBOLES CONTRE CONNEXIONISME : LE DÉBAT SUR LA SYSTÉMATIÇITÉ

SÉANCE 10, PHILO DE L'ESPRIT, H2024

JONATHAN SIMON

PLAN

- 1) Introduction: le cœur du débat -- règles contre associations.
- 2) Smolensky: l'alternative connexionniste
- 3) Fodor et Pylyshyn: l'argument de la "systématicité" contre le connexionnisme
- 4) Chalmers, Lake & Baroni: systématicité (limitée) pour les connexionnistes

The background is a dark blue gradient. In the corners, there are white line-art illustrations of circuit boards or neural networks, consisting of lines and small circles.

INTRODUCTION

INTRODUCTION

- La semaine dernière :
- nous avons examiné le débat sur ce que pourrait être l'intelligence artificielle, en termes très généraux. Le critère très général de l'intelligence artificielle de Turing (et ses arguments optimistes), puis les critiques très générales et abstraites de Searle et Dreyfus.

INTRODUCTION

- La vision de Turing de l'intelligence artificielle est "classique" - des systèmes qui manipulent des symboles, en utilisant les relations entre leur signification et leur syntaxe.
- Les critiques de Searle et Dreyfus étaient des critiques de cette approche classique (la possibilité que le traitement syntaxique des symboles puisse suffire à l'intelligence). Mais ni Searle ni Dreyfus ne développe une alternative claire

INTRODUCTION

- Smolensky expose la conception connexionniste - une alternative à la conception classique.
- Fodor et Pylyshyn défendent la conception classique. Ils donnent d'abord des critères cognitifs d'intelligence plus ciblés (par rapport aux critères comportementalistes de Turing). Ensuite, ils affirment que, s'il est facile de voir comment les systèmes classiques atteignent ces critères, le connexionnisme ne peut en principe pas le faire (à moins qu'il ne finisse par être du classicisme déguisé).

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

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 rapides" sont l'exception (plutôt que la règle). En général, nous raisonnons de manière probabiliste et nous comprenons de manière associative.

The background is a dark blue gradient. In the corners, there are white line-art illustrations of circuit boards or neural networks, with lines connecting to small circles.

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".
- 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)
- 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.»

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.

SMOLENSKY

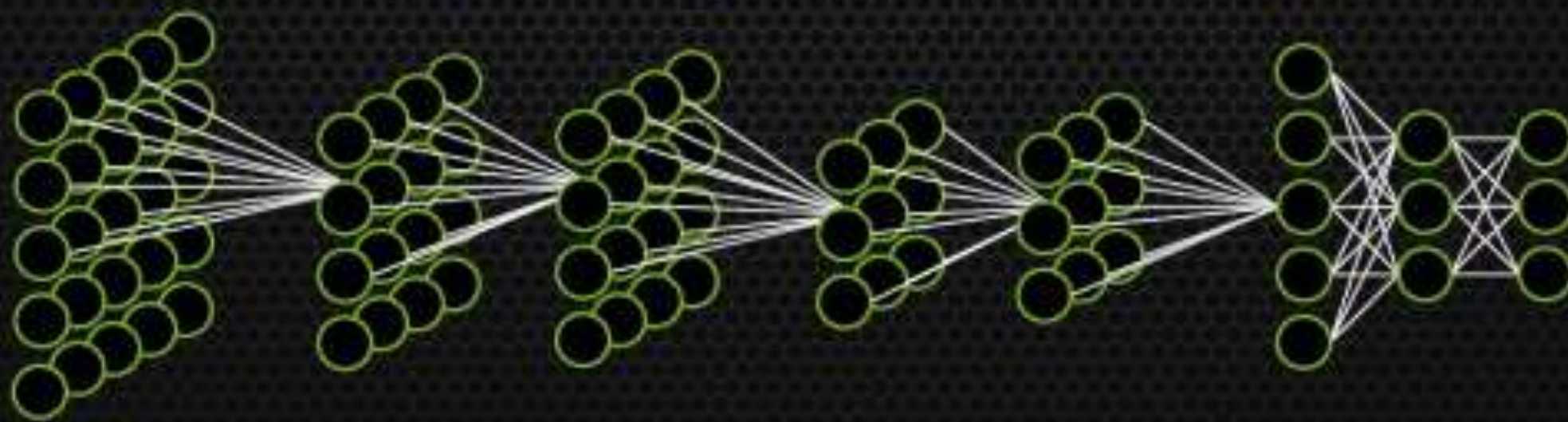
- En sa place, la vision positive:

SMOLENSKY

- La vision positive:



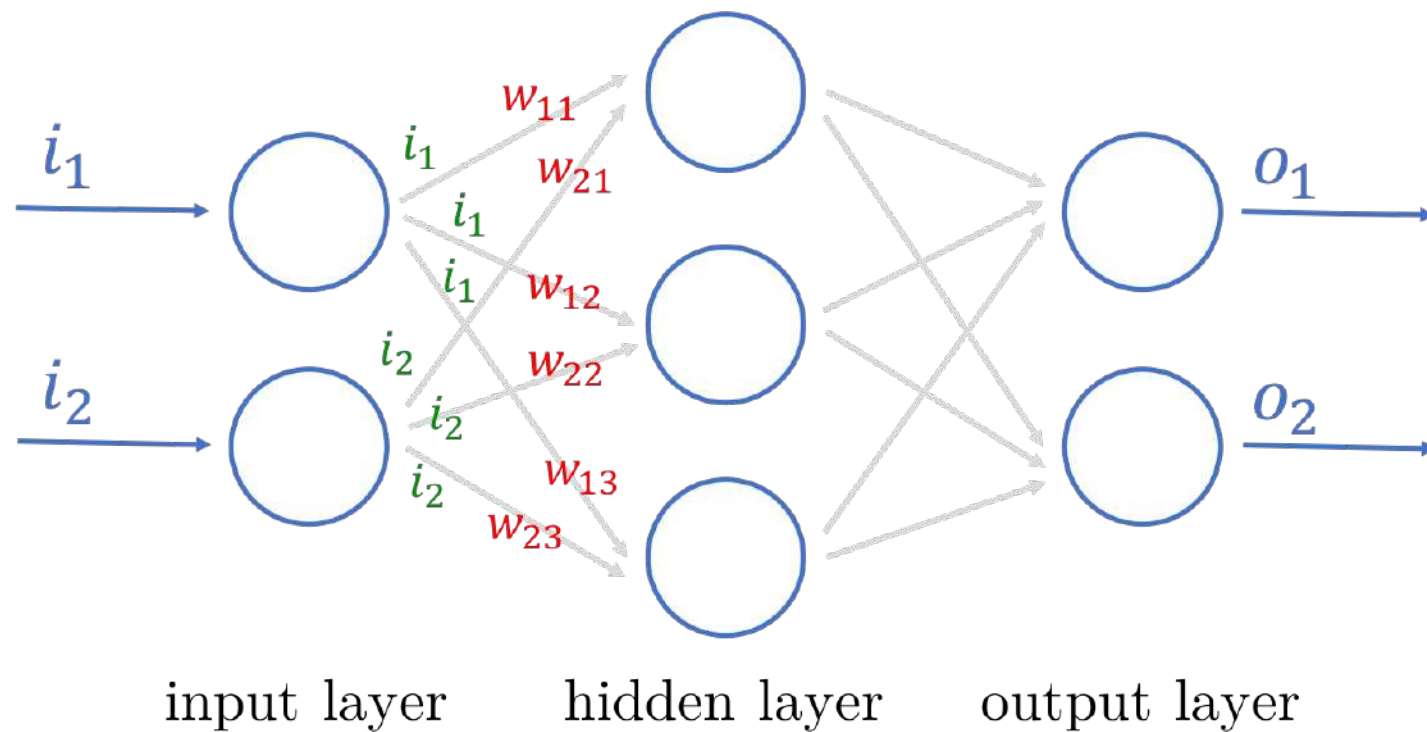
Image



“Sara”

SMOLENSKY

- La vision positive:



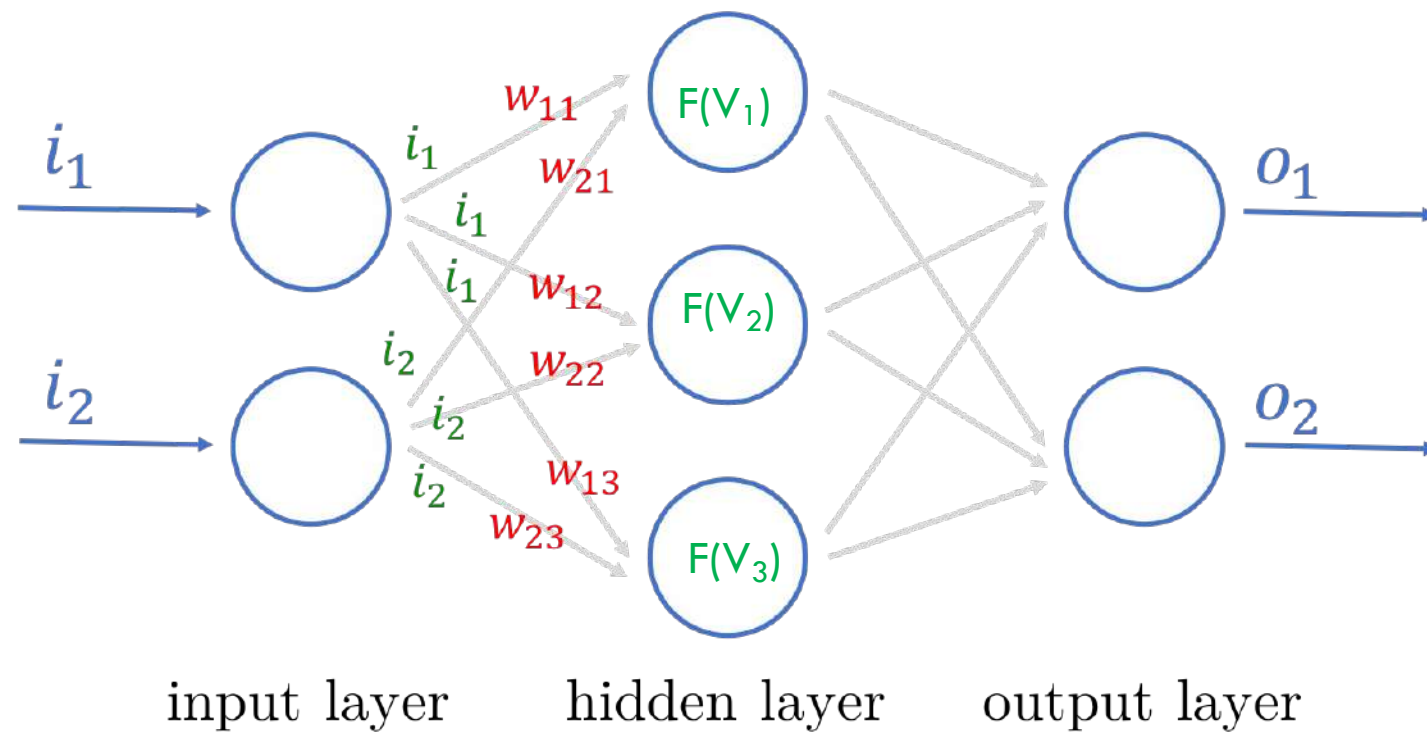
SMOLENSKY

- 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:

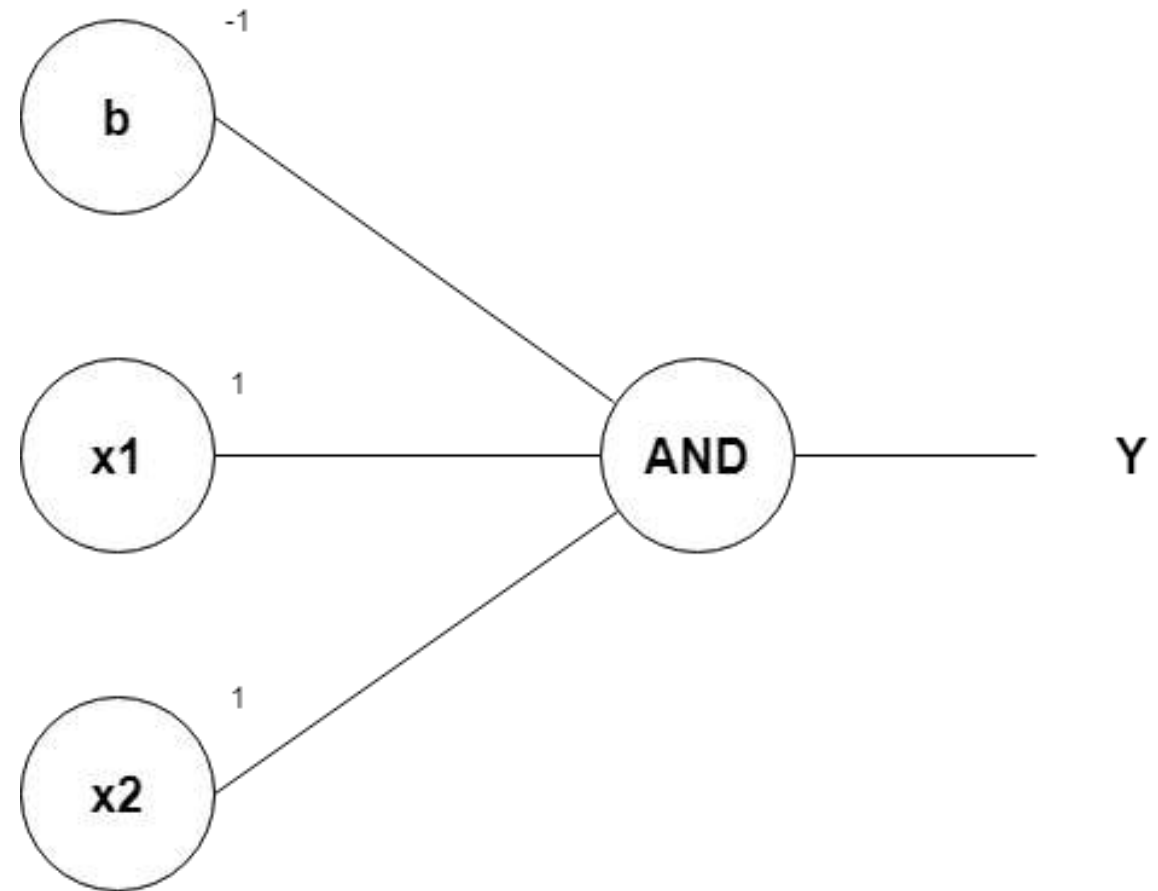
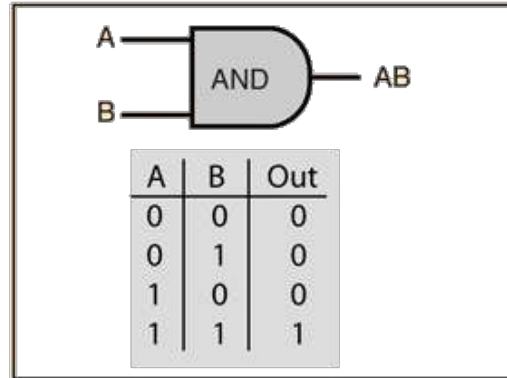
$$\begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \\ w_{13} & w_{23} \end{bmatrix} \cdot \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} (w_{11} \times i_1) + (w_{21} \times i_2) \\ (w_{12} \times i_1) + (w_{22} \times i_2) \\ (w_{13} \times i_1) + (w_{23} \times i_2) \end{bmatrix}$$

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.

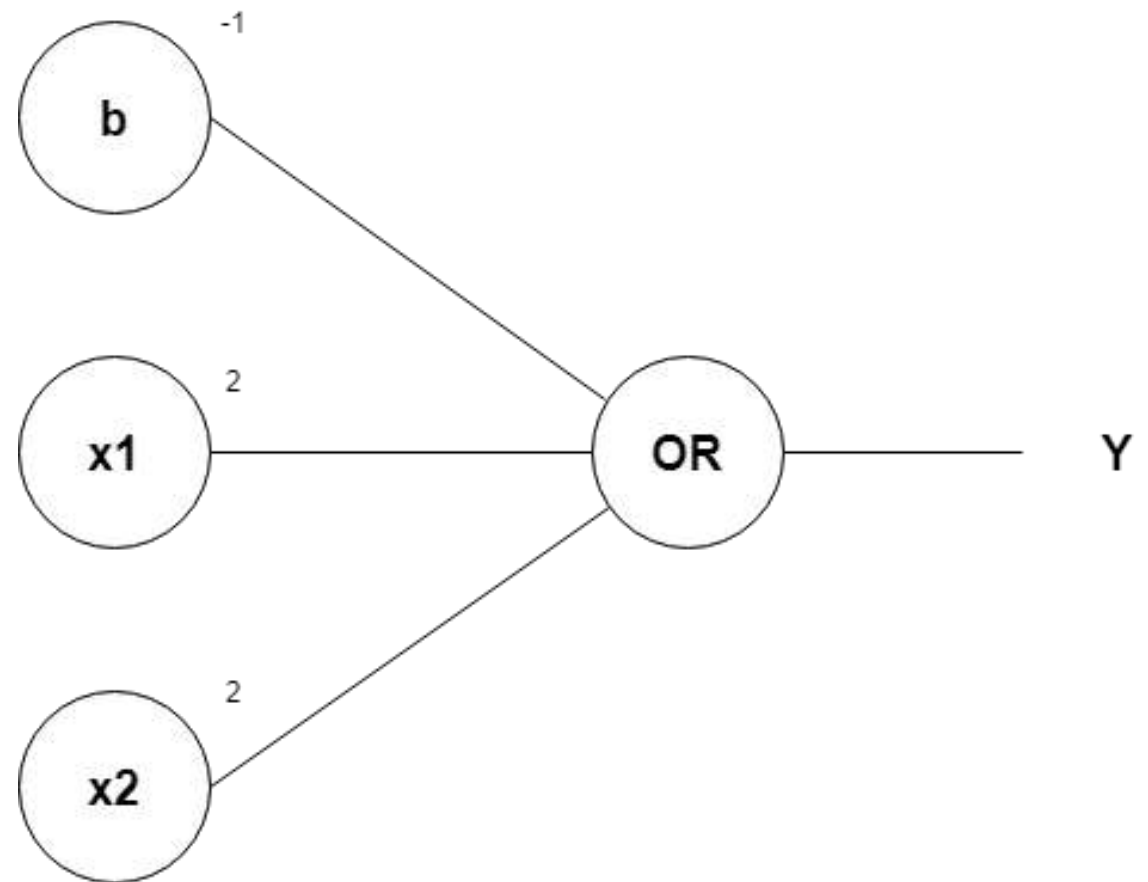
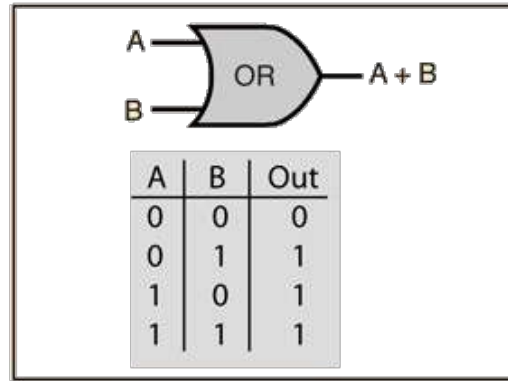
SMOLENSKY

- “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 dark blue gradient. In the corners, there are white line-art graphics resembling circuit boards or neural networks, with lines connecting to small circles.

FODOR ET PYLYSHYN: LA SYSTÉMATIÉTÉ DE LA PENSÉE

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 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



(a) some plants surrounding a lightbulb



(b) a lightbulb surrounding some plants

Figure 1. An example from Winoground. The two sentences contain the same words but in a different order. The task of understanding which image and caption match is trivial for humans but much harder for vision and language models. Every model that we tested (UNITER, ViLLA, VinVL, VisualBERT, ViLT, LXMERT, ViLBERT, UniT, FLAVA, CLIP, VSE++, and VSRN) fails to correctly pair the images and captions, except the large checkpoint of ViLLA by a very thin margin (0.00013 confidence).

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 connexionnistes 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) ?

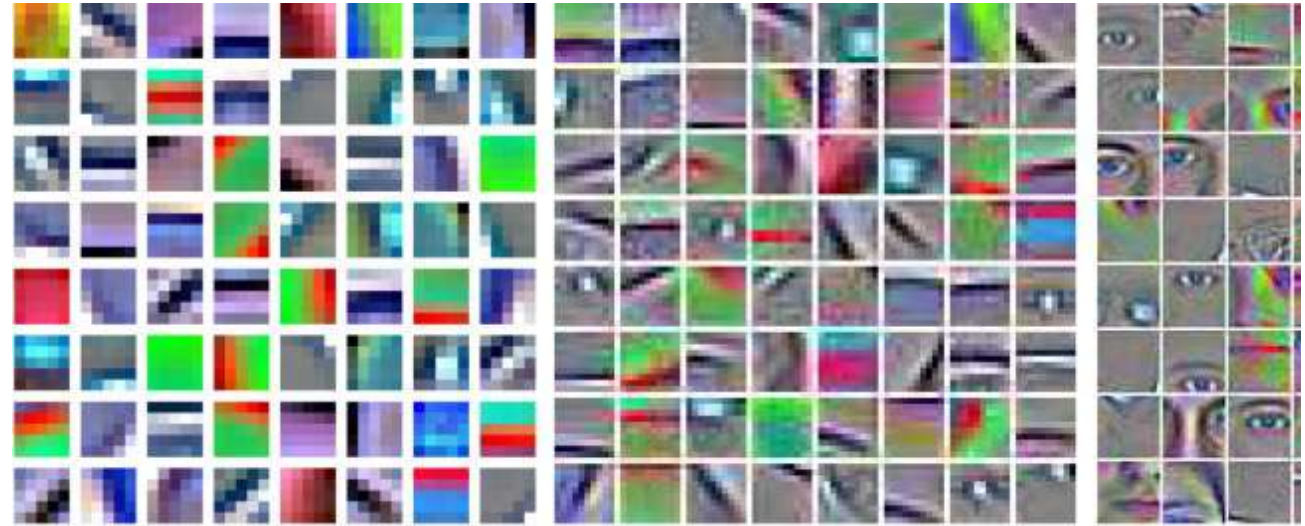
FODOR ET PYLYSHYN

- 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 "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)

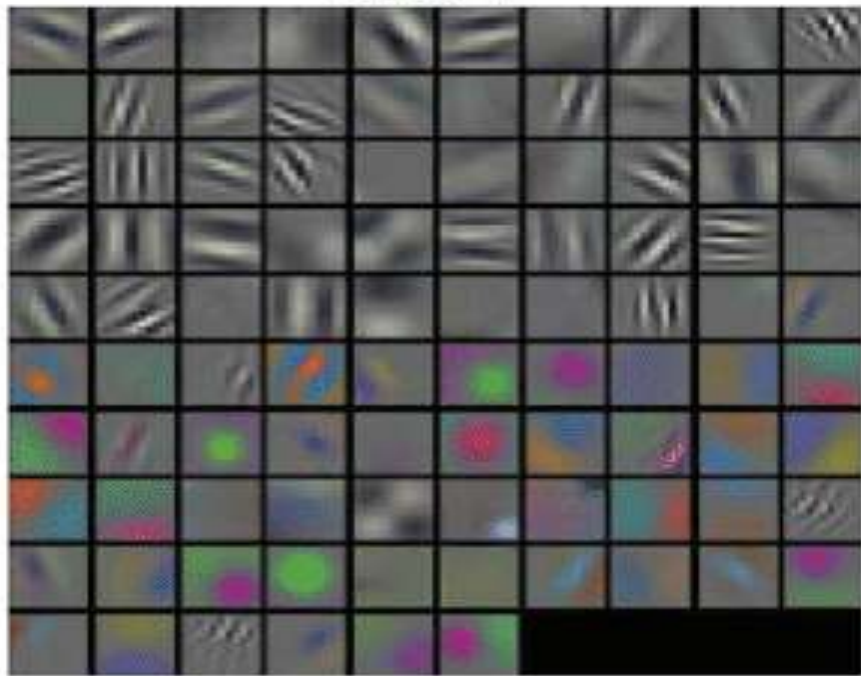


(d)

(c)

ImageNet-CNN

Conv 1



Pool 2



Pool 5

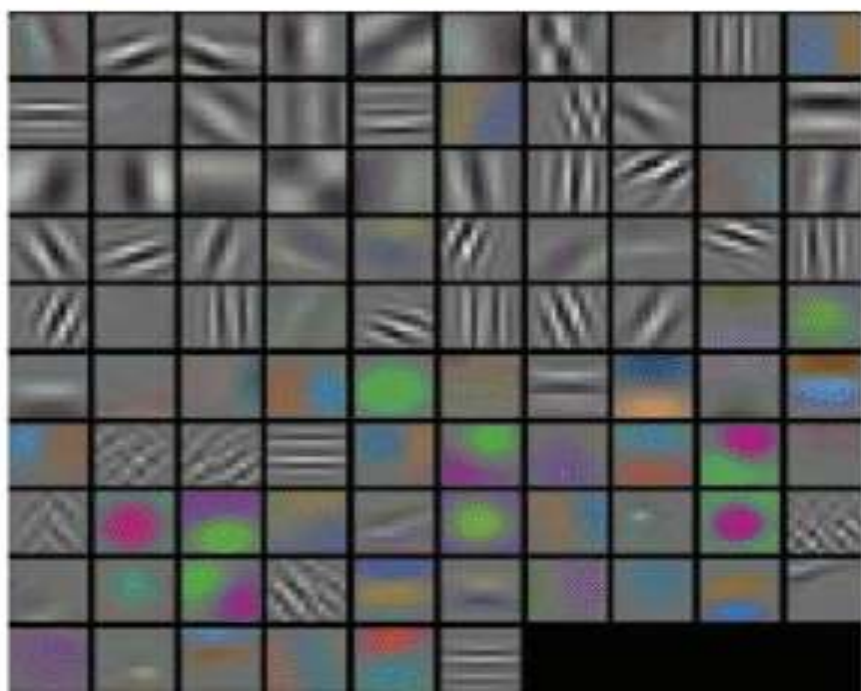


FC 7



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 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 ?)

The background is a dark blue gradient. In the corners, there are white line-art illustrations of circuit boards or neural networks, with lines connecting to small circles.

CHALMERS, LAKE & BARONI: SYSTÉMATIÉTÉ
(LIMITÉE) POUR LES CONNEXIONNISTES

CHALMERS, LAKE ET BARONI

- Chalmers: 1) F&P prouvent trop : si leur argument fonctionne, il montre que les systèmes connectionnistes ne peuvent même pas mettre en œuvre des ordinateurs classiques (ce que nous savons qu'ils peuvent faire).
- 2) Rapport sur un modèle connexionniste qui présente une sensibilité de structure limitée (systématicité) sans structure partielle-ensemble

CHALMERS, LAKE ET BARONI

- Lake et Baroni:) Rapport sur un modèle connexionniste qui présente une sensibilité de structure limitée (systématicité) sans structure partielle-ensemble