Intelligence Artificielle

Antoine Cornuéjols

AgroParisTech – INRA MIA Paris-Saclay

EKINOCS research group

Outline

1. La perspective de l'Intelligence Artificielle
2. Notions de base en science des données
3. L'induction d'arbres de décision
4. L'induction d'arbres de régression
5. Les forêts aléatoires

L'IA/AA dans tous les médias aujourd'hui

Que savez-vous de l’IA ?

- Aide à la décision
  - Robotique
    - Pour la médecine
- Chatbot : assistant intelligent, conversation
  - Analyse de texte (Traitement automatique du langage)
  - La décision
- Apprentissage par l’expérience : à partir de données, interactions avec le monde
  - Réseaux de neurones
  - Prédiction
    - Diagnostics en médecine
    - Repliement de la protéine

Que savez-vous de l'histoire de l'IA ?

- Date de naissance du terme IA ?
  - 1956 : workshop à Dartmouth college
    - Démonstration automatique de théorème
    - Souris cybernétique
    - GPS : General Problem Solver
- (1956 – 1968) I = méthodes générales de raisonnement (résolution de problème)
  - Jeux

D'où vient l'I.A. ?

Les pionniers : réseaux de neurones et raisonnement

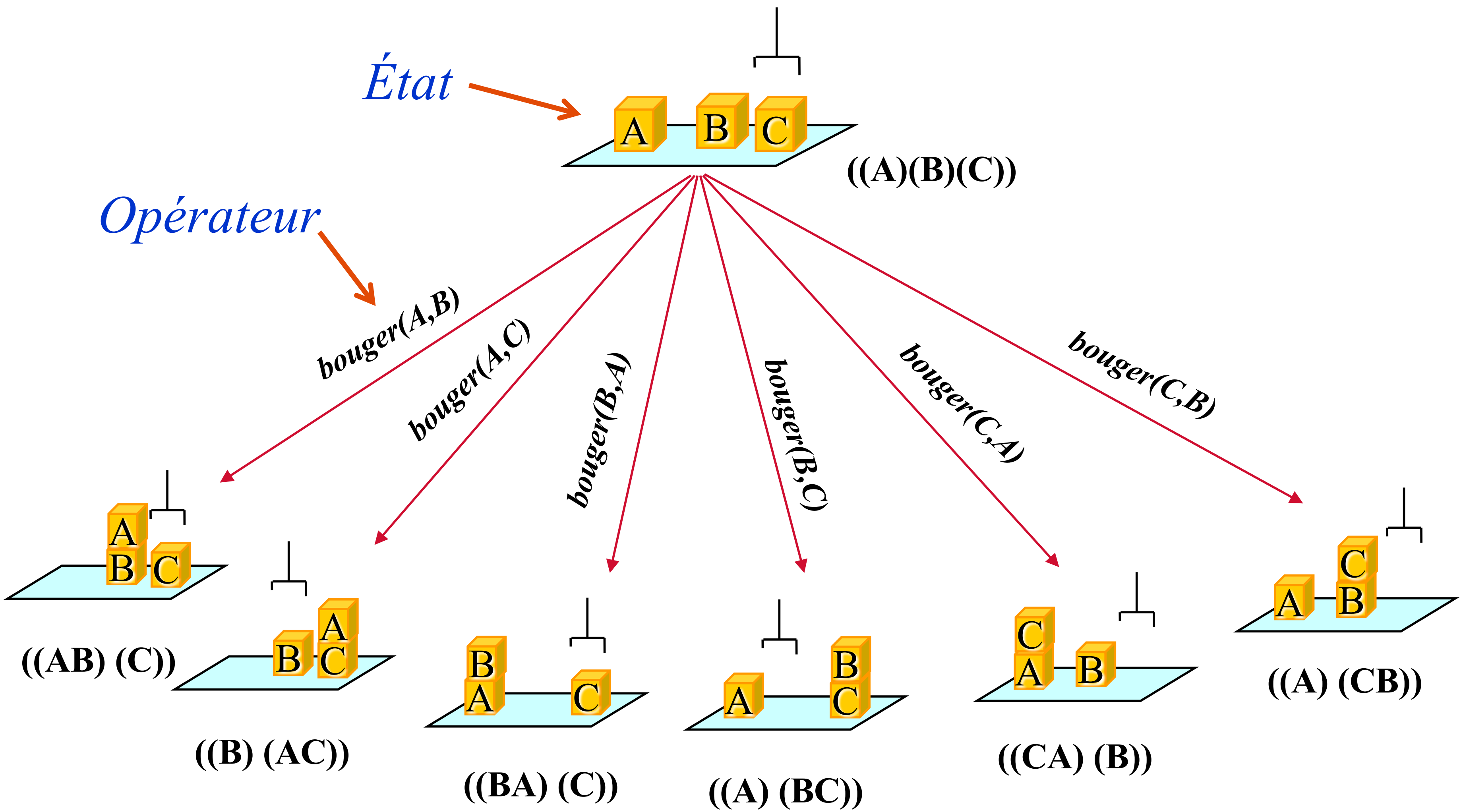
(1956 – 1969)

L’espoir

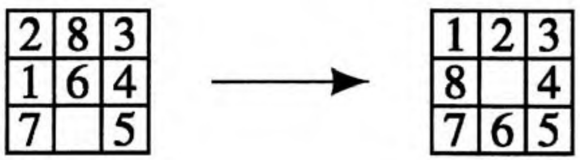
L’intelligence met en jeu

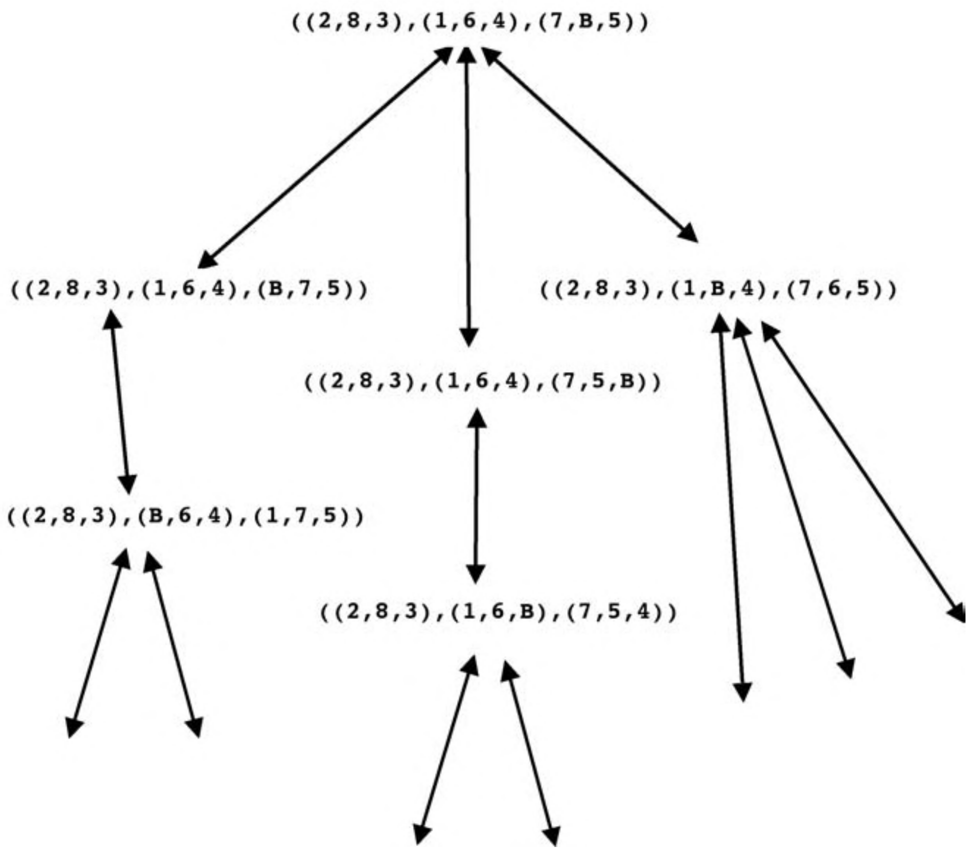
des **processus généraux** de raisonnement

Raisonnement / résolution de problèmes



Raisonnement / résolution de problèmes

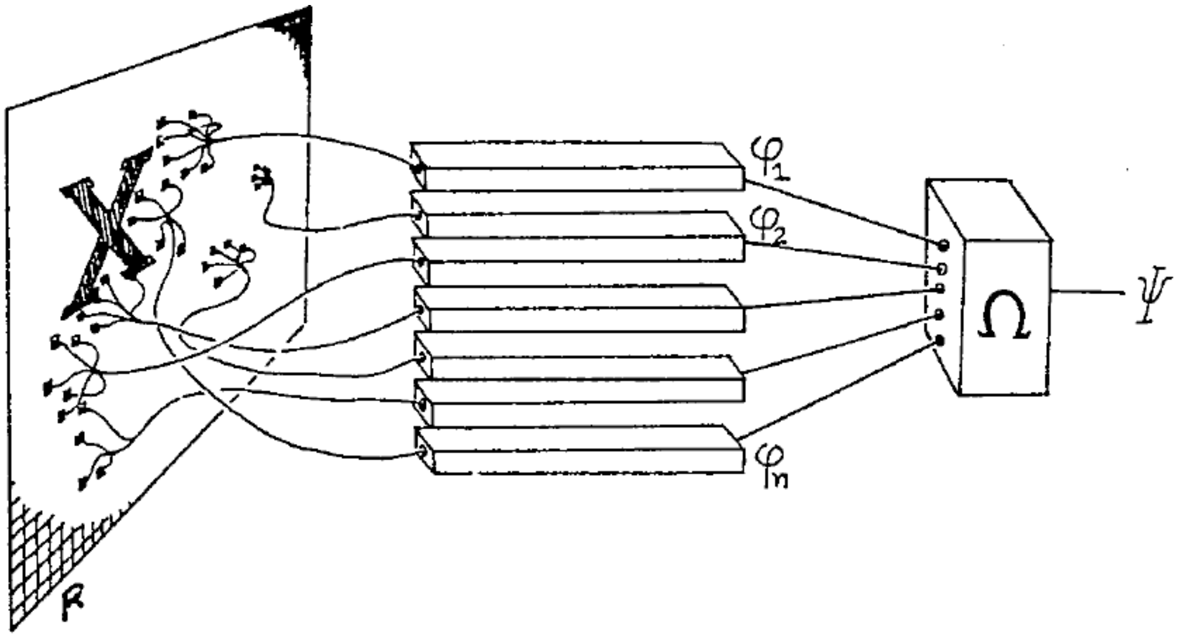




- Recherche dans un graphe

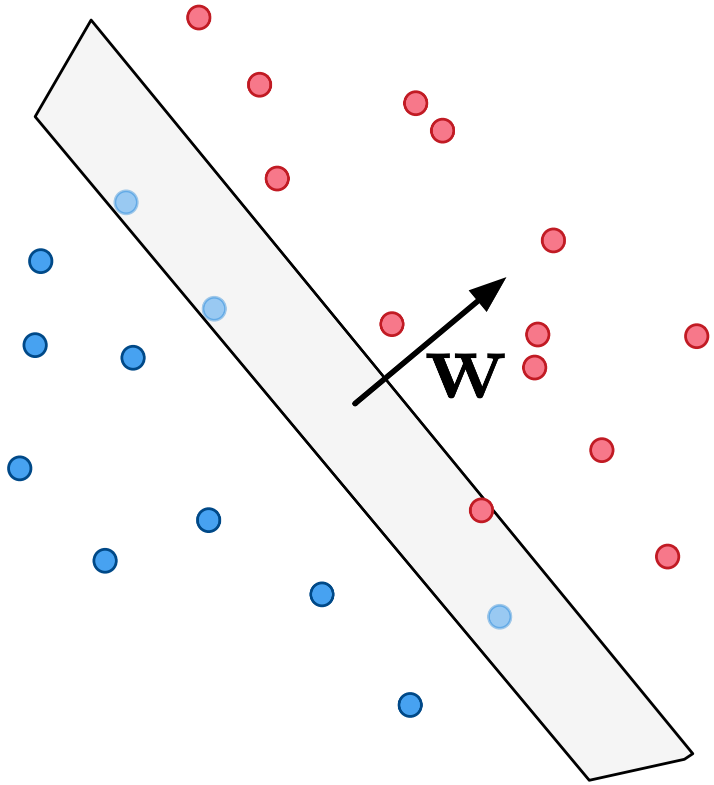
Premier connexionnisme : le perceptron

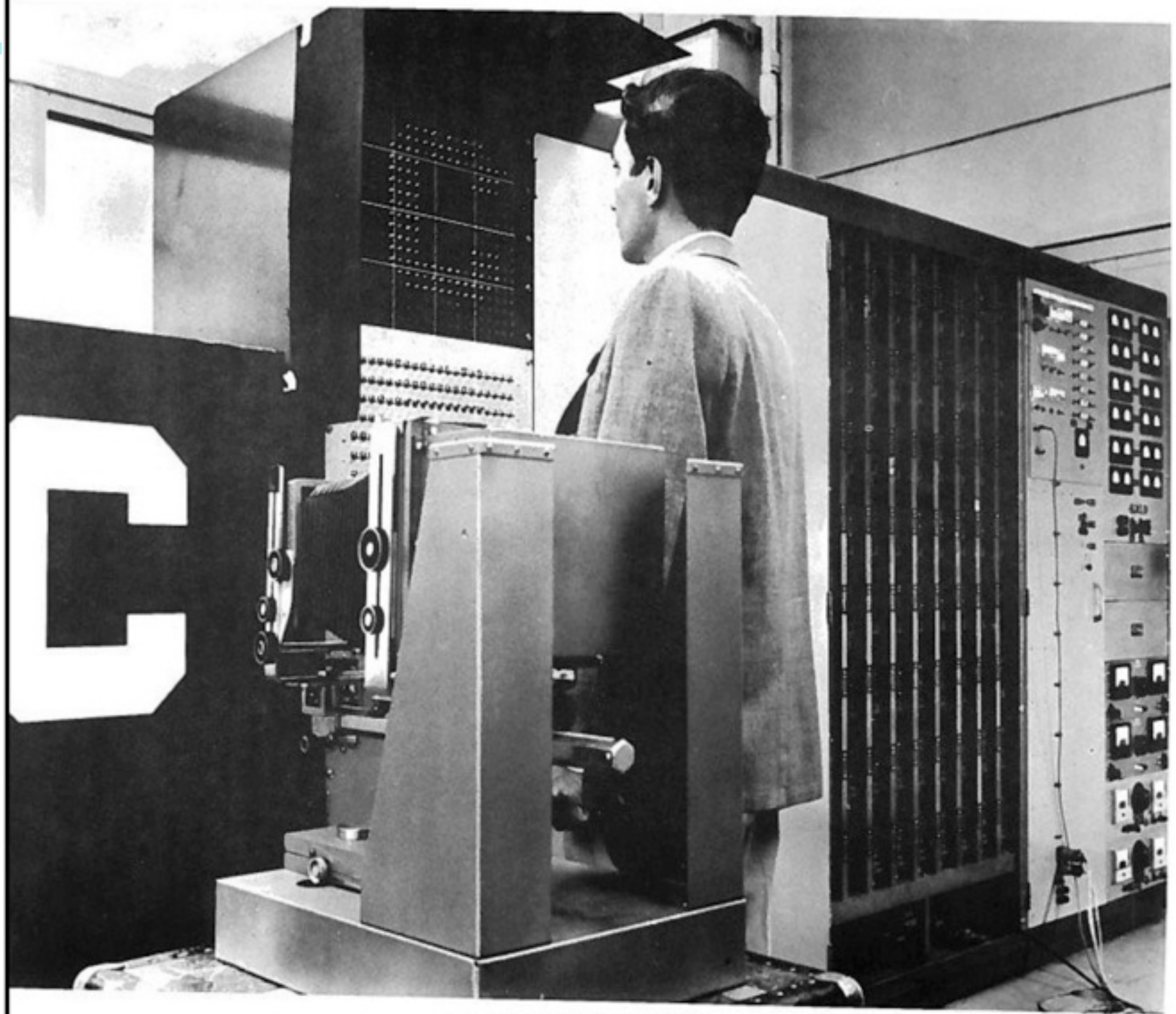
- Frank Rosenblatt (1958 – 1962)



$$\Psi(\mathbf{x}) = \sum_{i=1}^{n} w_i \ \phi_i(\mathbf{x})$$

The perceptron: a linear discriminant





THE MARK I PERCEPTRON

NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI)—The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau's $2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

...

FIGURE 10.19 : *Article du New York Times daté du 8 juillet 1958 parlant du perceptron et de son financement. Les ordinateurs les plus puissants de l'époque (l'IBM 704 coûtant à l'époque 2 000 000 de dollars(20 000 000 aujourd'hui)) ne pouvaient pas réaliser plus de 12 000 multiplications/seconde (un smartphone d'aujourd'hui, en 2020 peut en réaliser environ 35 milliards !). Le perceptron était donc réalisé sur une machine de type analogique, dont le coût à l'époque était de 100 000 dollars, soit environ 1 000 000 aujourd'hui. On notera le ton de l'article bien proche des articles actuels sur le « deep learning » :*

La marine a révélé aujourd'hui l'existence d'un ordinateur électronique, embryon d'une machine qui, selon elle, pourra marcher, parler, voir, écrire, se reproduire et être conscient de son existence. L'embryon, l'ordinateur « 704 » du bureau météorologique à 2 000 000 dollars, a appris à différencier la droite de la gauche après une cinquantaine de tentatives dans la démonstration de la marine à destination des journalistes. Le service a déclaré qu'il utiliserait ce principe pour construire la première de ses machines à penser Perceptron qui sera capable de lire et d'écrire. Elle devrait être terminée dans un an environ, pour un coût de 100 000 dollars. Le Dr Frank Rosenblatt, concepteur du Perceptron, a effectué la démonstration. Il a affirmé que la machine serait le premier appareil à penser comme le cerveau humain. Comme les êtres humains, le Perceptron fera des erreurs au début mais deviendra plus performant au fur et à mesure qu'il gagnera en expérience, a-t-il dit. Le Dr Rosenblatt, un psychologue au Cornell Aeronautical Laboratory, à Buffalo, a déclaré que les Perceptrons pourraient être lancés sur les planètes en tant qu'explorateurs mécaniques de l'espace.

...

Démonstration de théorèmes

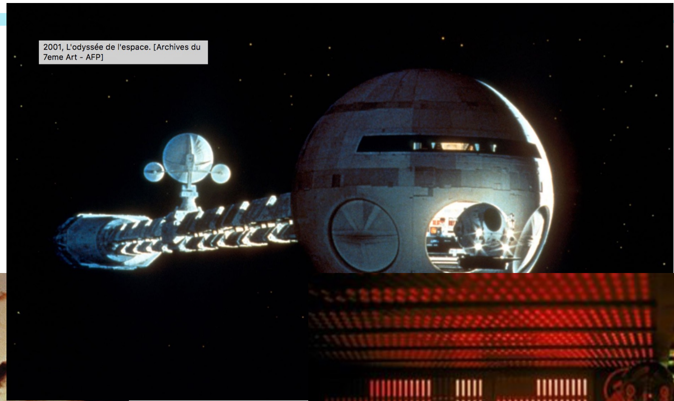
- Démonstration de théorème
- Planification
- Raisonnement

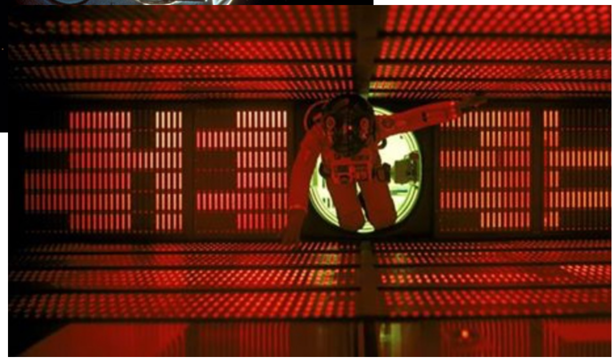
15

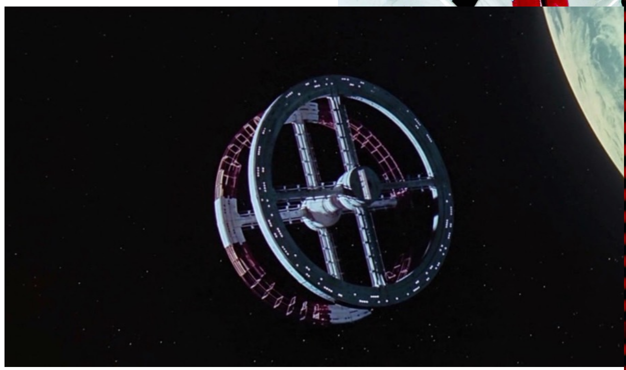
1 9 6 8

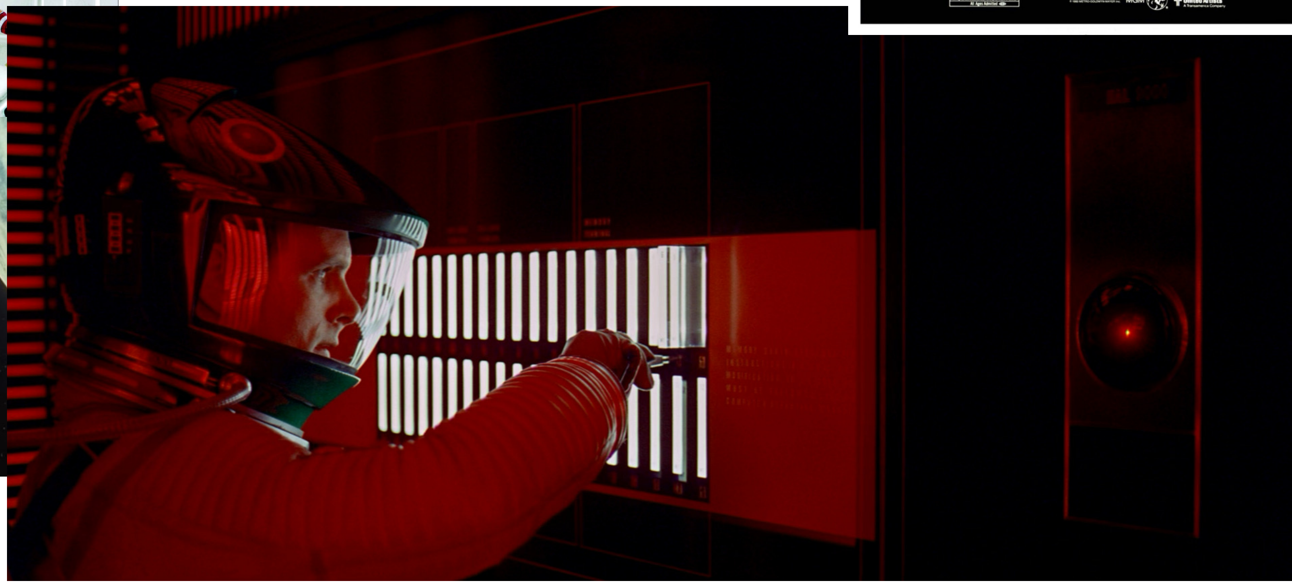
Une année extraordinaire … en cache une autre

- 1968 …







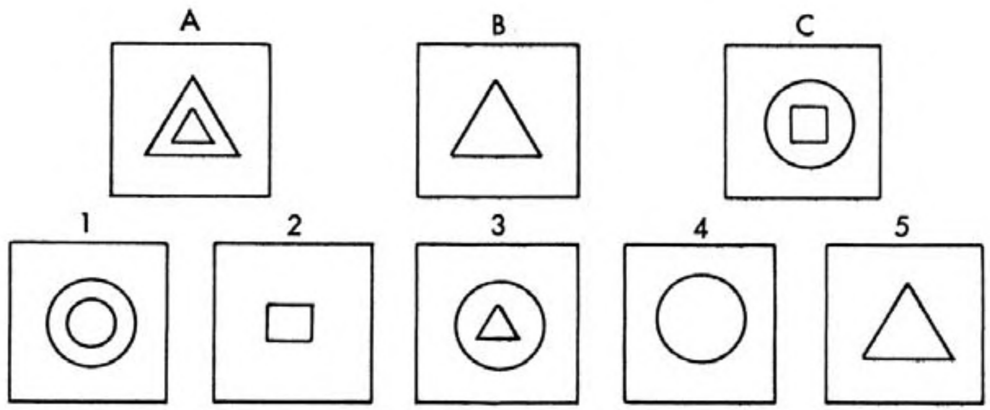


2001 : Odyssée de l’espace

- Vision
- Communication
  - Lecture sur les lèvres
  - Conversation
- Planification
- Raisonnement
  - Joue (et gagne) aux échecs
- Auto-reprogrammation
  - Tue des astronautes
- Émotion
  - Peur

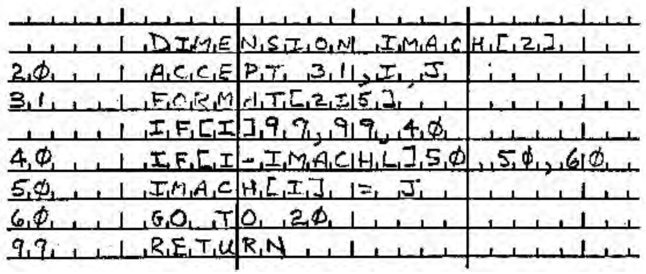
1 9 6 8 : que sait-on faire ?

- Raisonnement
  - Résolution de problème
  - Mémoire associative
- Jeu de dames
- **Analogie**
- Conversation (?)
  - Eliza
- Apprentissage
  - Reconnaissance de caractères
  - Jeu de dames



1 9 6 8 : que sait-on faire ?

- Raisonnement
  - Résolution de problème
  - Mémoire associative
- Jeu de dames
- Analogie
- Conversation (?)
  - Eliza
- **Apprentissage**
  - **Reconnaissance de caractères**
  - Jeu de dames

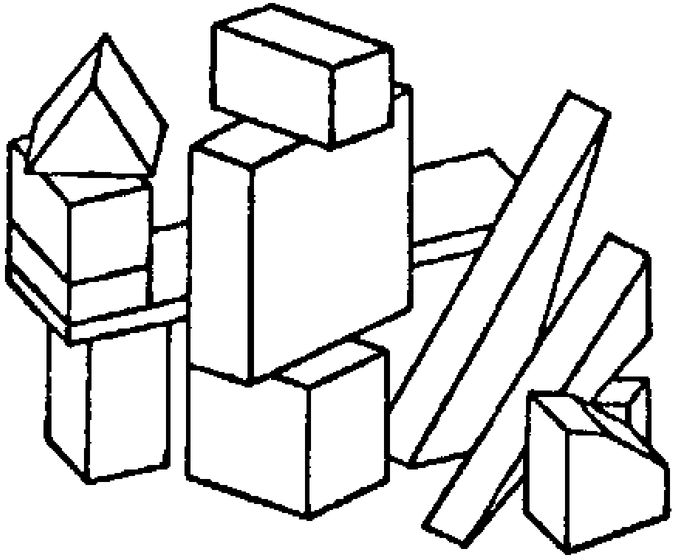


```
      DIMENSION IMACM[2]
20    ACCEPT 31,I,J
31    FORMAT[215]
      IF[I]79,99,40
40    IF[I-IMACHL]50,50,60
50    IMACH[I]=J
60    GO TO 20
99    RETURN
```

Taux de reconnaissance = 98% !!!

- Stanford AI Lab

Figure 8.1: Site of the Stanford AI Lab from 1966 until 1980. (Photograph courtesy of Lester Earnest.)



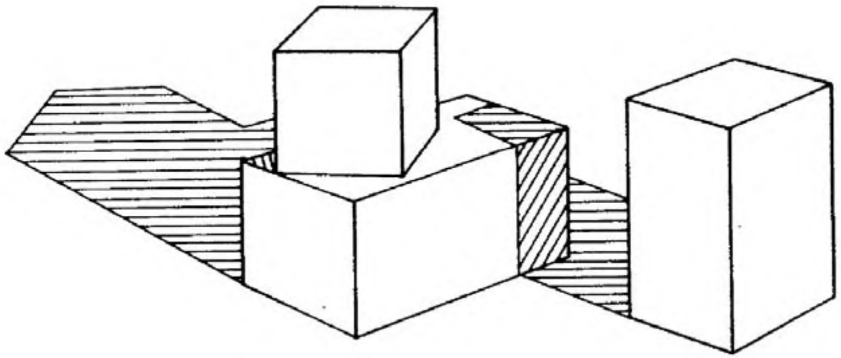


Figure 9.15: A scene with shadows analyzed by Waltz's program. (Illustration used with permission of David Waltz.)

Vision + planification +

interface par langage

pseudo naturel

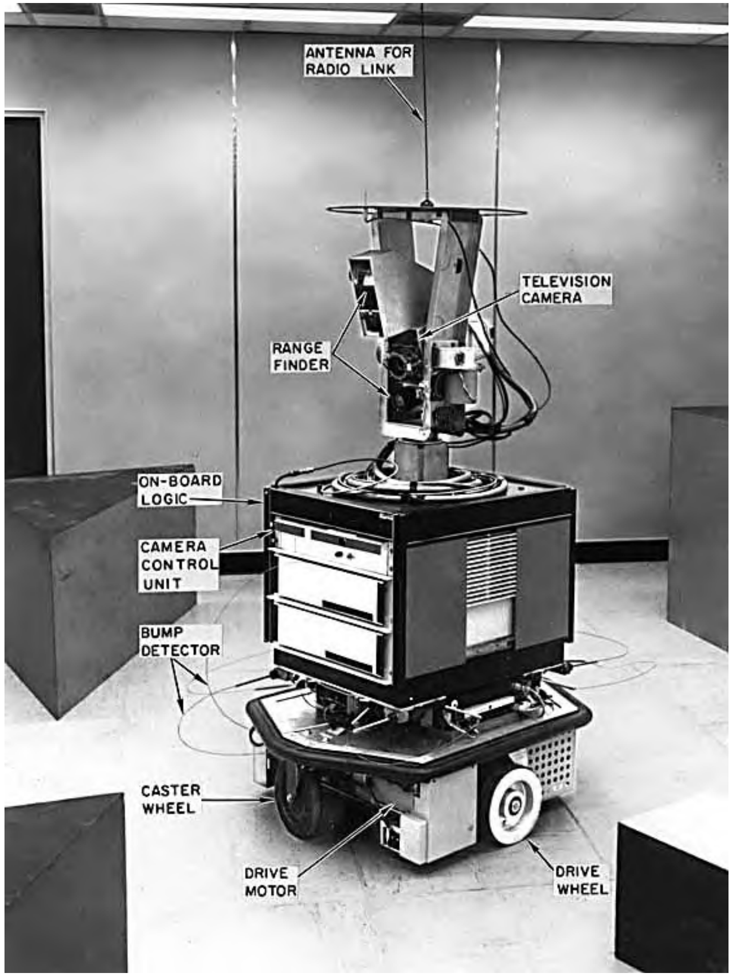


Figure 12.3: Shakey as it existed in November 1968 (with some of its components labeled). (Photograph courtesy of SRI International.)

SHRLDU Interface en langage pseudo naturel

1. pick up a big red block.

In order to pick up the red block, it had to clear it off by finding a space for the green one and moving the green one away.

OK.

The system answers "OK" when it carries out a command.

2. grasp the pyramid.

There are three pyramids in the scene, so the program does nothing.

I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.

3. find a block which is taller than the one you are holding and put it into the box.

The system has to figure out what is meant by "one" and "it," as well as using deduction to decide which block it is holding.

BY "IT," I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.

Printed out to keep us informed.

The hand has to set down the red block to pick up the taller blue one.

OK.

4. what does the box contain?

THE BLUE PYRAMID AND THE BLUE BLOCK.

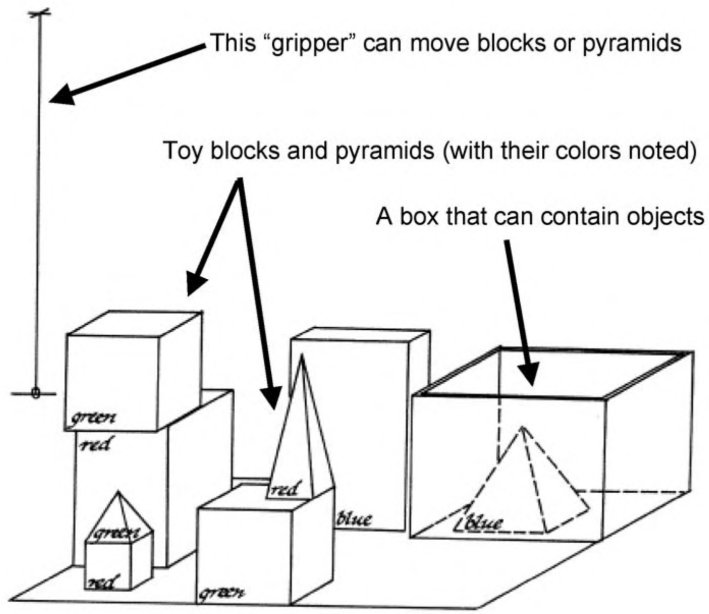
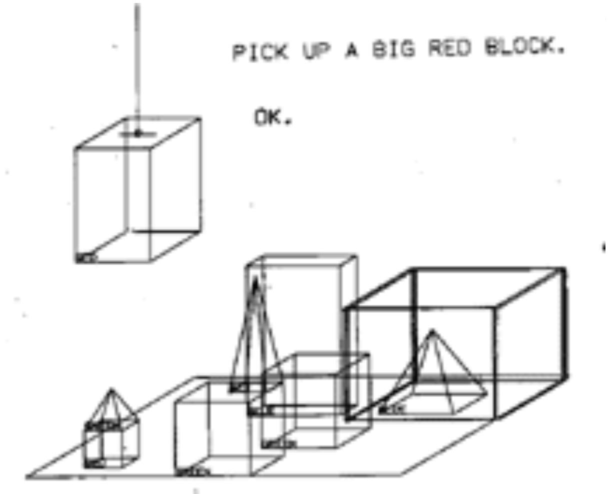
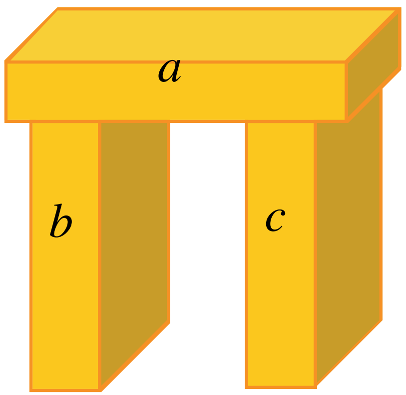


Figure 13.2: SHRDLU's world. (Illustration used with permission of Terry Winograd.)

ARCH [Winston, 1970]

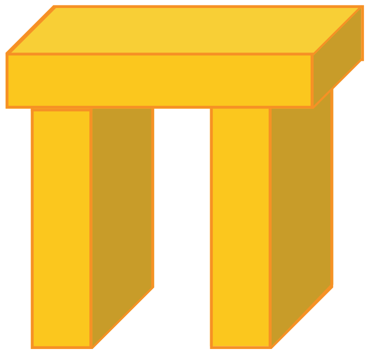
- **Apprentissage** de **concept** (e.g. arche) dans un monde de blocs





ARCH [Winston, 1970]

- Les exemples ne sont **pas choisis au hasard**



[Ross Quillian, 1968 : Semantic memory]

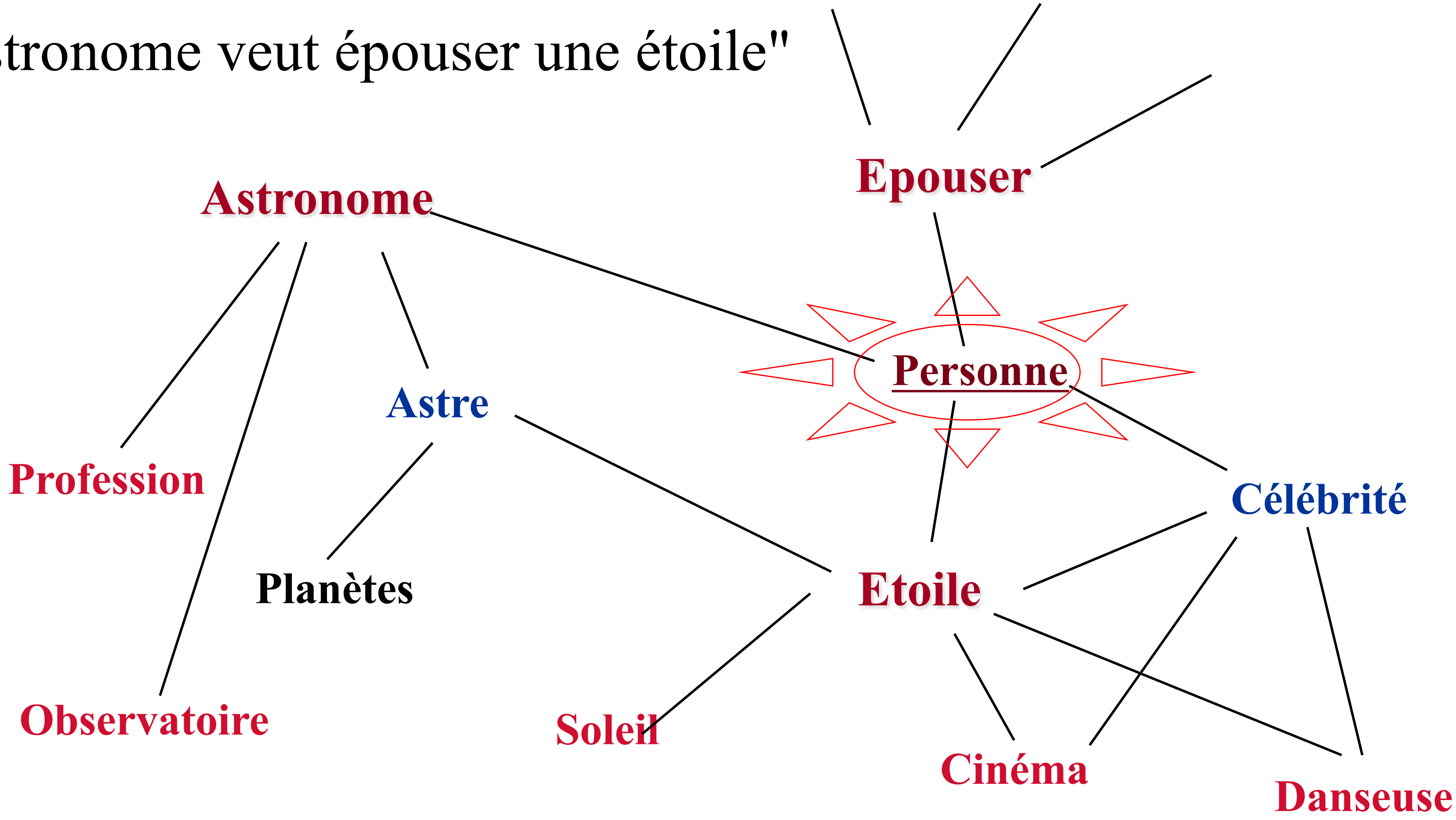
- Résolution d'ambiguïtés
  (par intersection de propagation de marqueurs)

"L'astronome voulait épouser une étoile"

- Un concept acquiert un sens à travers le réseau sémantique dans lequel il s'insère et les relations qu'il a avec d'autres concepts

Les réseaux sémantiques

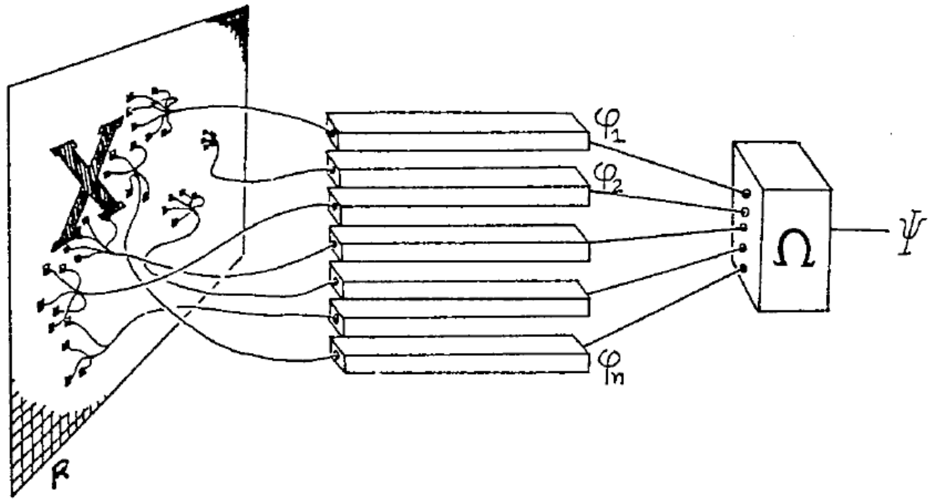
"L'astronome veut épouser une étoile"



Inférence par propagation d'*activité* ou de *marqueurs* : focalise l'attention

1 9 6 8 : qu'est-ce qu'on ne sait pas faire ?

- **Raisonnement**
  - Lourd. Pas au niveau des experts
- **Apprentissage**
  - Limité [Minsky et Papert]
    - 1969 « Perceptrons »
- **Traduction** automatique



Perspective historique : apprentissage automatique

1950s

1960s

1970s

- Expériences de pensée sous contrainte de réalisabilité computationnelle
  - **Opérateurs** sur des représentation
  - Recherche dans un **espace d'états**. Buts / sous-buts
  - **Apprentissage** par mutations aléatoires (mais guidée par ressemblance) ~ AG
  - Ivresse : comprendre la pensée
  - *« Our problem, our joint problem, is to discover what transformations must be made on the available data in order to preserve intact the significant features and to discard the irrelevant details ».*
- Principes, théorèmes et démonstrations (Checker. Problèmes « jouets »)
  - Reconnaissance des formes. Plutôt numérique (bayésien, perceptron)

D’où vient l’I.A. ?

Knowledge is power

(1970 – 1985)

Knowledge is **power**

Perspective historique : apprentissage automatique

1950s

- Expériences de pensée sous contrainte de réalisabilité computationnelle
  - **Opérateurs** sur des représentation
  - Recherche dans un **espace d'états**. Buts / sous-buts
  - **Apprentissage** par mutations aléatoires (mais guidée par ressemblance) ~ AG
  - Ivresse : comprendre la pensée
  - « *Our problem, our joint problem, is to discover what transformations must be made on the available data in order to preserve intact the significant features and to discard the irrelevant details* ».

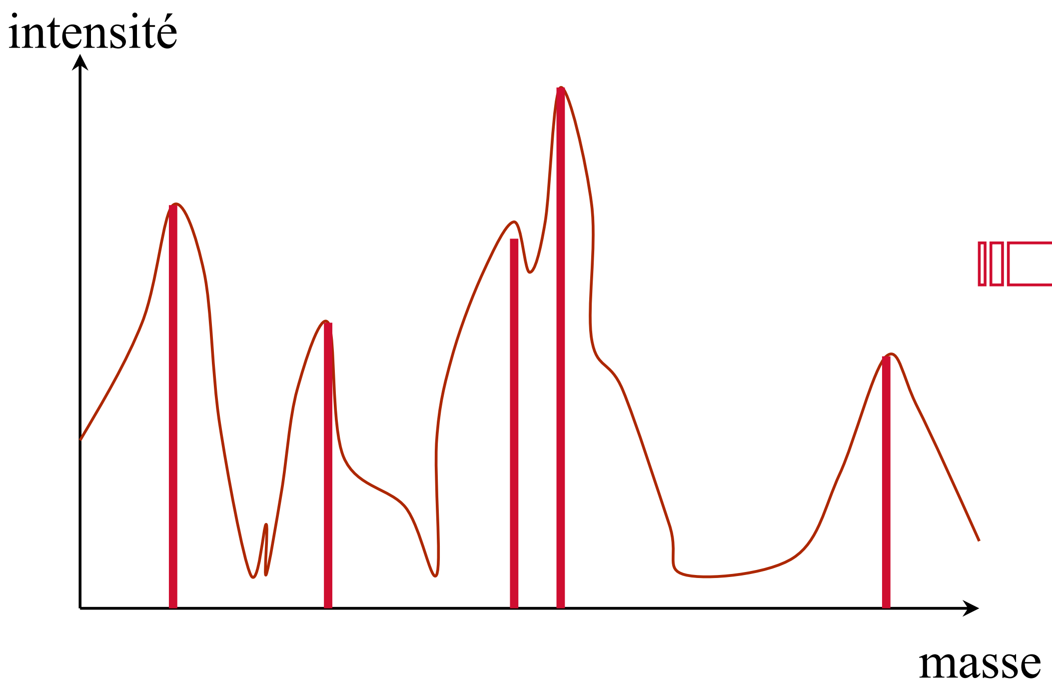
1960s

- Principes, théorèmes et démonstrations (Checker. Problèmes « jouets »)
  - Reconnaissance des formes. Plutôt numérique (bayésien, perceptron)

1970s

- Expertise. Sciences cognitives : plausibilité. Intégration dans le raisonnement
  - **Modèles de mémoire**. Réseaux sémantiques. Représentation des connaissances
  - **Règles** de production. Moteur d'inférence.
  - **Mécanismes** d'apprentissage et de **généralisation**
  - **Apprentissage et raisonnement** : heuristiques, macro-opérateurs, chunking

- Le système DENDRAL
  - Pour la NASA : 1965 - …
  - Y a-t-il de la vie sur Mars ?
  - Spectrographie de masse



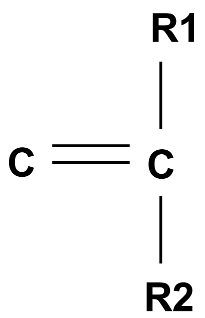
Formule développée du composé chimique ?

- Exemples de connaissances
  - Règle :

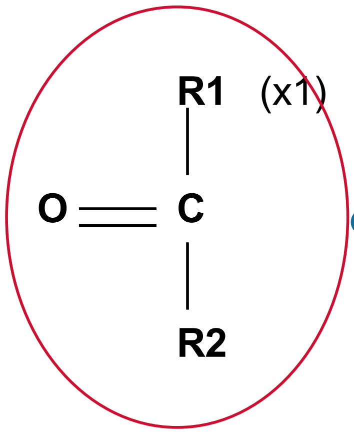
Si le spectre de la molécule présente deux pics x1 et x2 tels que :

1. x1 - x2 = M + 28
2. x1 - 28 est un pic élevé
3. x2 - 28 est un pic élevé
4. au moins l'un des pics x1 et x2 est élevé

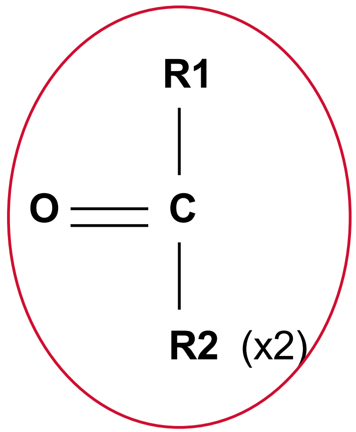
Alors la molécule contient un groupe cétone



Se décompose en :



ou en :



Les systèmes experts : MYCIN

MYCIN : Mes recommandations thérapeutiques seront basées sur les infections probables et les organismes infectieux potentiels suivant :

L'organisme peut être : 1° pseudomonia aeruginosa, 2° klebsiellia pneumoniae, 3° escherichia coli, 4° bacteroides fragilis, 5° enterobactoriaceae, 6° proteus non mirabilis

Voulez-vous me donner la gravité de la maladie de M. Fred Smith sur une échelle de 0 à 4 ?

Médecin : 3

MYCIN : M. Fred Smith présente-t-il des réactions allergiques à un quelconque agent microbien ?

Médecin : Non

MYCIN : Quel est son poids en kilos ?

Médecin : 70

Les systèmes experts : MYCIN

MYCIN : Ma recommandation thérapeutique sera en premier choix : pour couvrir les points <1 2 3 5 6>, prescrire de la Gentamicine, dose 119 mg 98 h IV, pendant 10 jours;

Pour couvrir le point 4, prescrire de la Clindamycine, dose 595 mg 96 h IV, pendant 14 jours.

Désirez-vous éventuellement éliminer certains médicaments ?

Médecin : Oui, la Clindamycine.

MYCIN : Pour couvrir les points <2 3 4 5 6> donner du Chloramphénicol, dose 1 g 96 h IV, pendant 14 jours;

** Surveiller le taux de globules blancs du patient.

Pour couvrir le point 1, donner de la gentamicine, dose 119 mg 98 h IV, pendant 10 jours.

** Modifier la dose en cas de faiblesse rhénale.

Les explications des SEs

MYCIN : L'infection du malade a t-elle été contractée à l'hôpital ?

Utilisateur : pourquoi ?

TEIRESIAS : [pourquoi est-ce important de savoir cela ?]

Cela aidera à déterminer la catégorie à laquelle appartient l'organisme responsable de l'infection :

- il a déjà été établi que 1° l'organisme est à Gram négatif, 2° c'est un bâtonnet, 3° il se développe à l'air ou sans air,

- donc si l'infection par cet organisme n'a pas été contactée lors du séjour à l'hôpital

alors il est presque certain (0,8) que l'organisme est de la catégorie des enterobactoriaceae; de plus, il se peut (0,1 ; règle 37) que ce soit la pseudomonia aeruginosa.

Utilisateur : pourquoi n'avez-vous pas envisagé la possibilité que l'organisme soit un streptocoque ?

TEIRESIAS : la règle 33 aurait du être utilisée pour identifier l'organisme comme un streptocoque, mais la seconde clause "la morphologie est de type coque" était déjà connue comme fausse et la règle 33 n'a jamais été essayée.

BORIS [Dyer,82]

"Richard hadn't heard from his college roomate Paul for years. Richard had borrowed money from Paul which was never paid back. But now he had no idea where to find his old friend. When a letter finally arrived from San Francisco, Richard was anxious to find out how Paul was."

BORIS [Dyer,82]

"Richard hadn't heard from his college roomate Paul for years. Richard had borrowed money from Paul which was never paid back. But now he had no idea where to find his old friend. When a letter finally arrived from San Francisco, Richard was anxious to find out how Paul was."

Comment traiter un non événement ?

"John walked in the room and Mary was not there" (John désirait voir Mary et s'attendait à la trouver dans la pièce -> pourquoi n'y est-elle pas ? …)

Ici, on est prêt à comprendre que Richard se sent une obligation vis-à-vis de Paul. Il faut donc des connaissances sur les relations sociales et sur la psychologie.

Comment traiter "had" (possession ?) et "old" (âge de son ami ?. Non. Mais pourquoi et comment?)

Comment infère-t-on que cette lettre vient de Paul alors que ce n'est pas dit explicitement ?

Pourtant …

- Ingénierie des connaissances
- Processus très lourd
- Peu systématisé
- Maintenance difficile

Dépasse **le quantum d'action** = la thèse

D’où vient l’I.A. ?

Pourquoi ne pas tout apprendre ?

(1985 – … ?)

L’espoir (suite de la suite)

L’intelligence met en jeu beaucoup de **connaissances**

que l’on obtiendra par des **processus généraux** d’**apprentissage**

Plan

1. D'où vient l'IA ?
    - Préhistoire : la cybernétique
    - I.A.= raisonnement
    - I.A. = connaissances
    - I.A. et apprentissage
2. Apprentissage artificiel
    - Apprendre est difficile
    - L'apprentissage « profond »
3. L'I.A. partout : sommes-nous prêts ?

1- Apprendre des connaissances pour les systèmes experts

Illustration: LEX (Tom Mitchell)

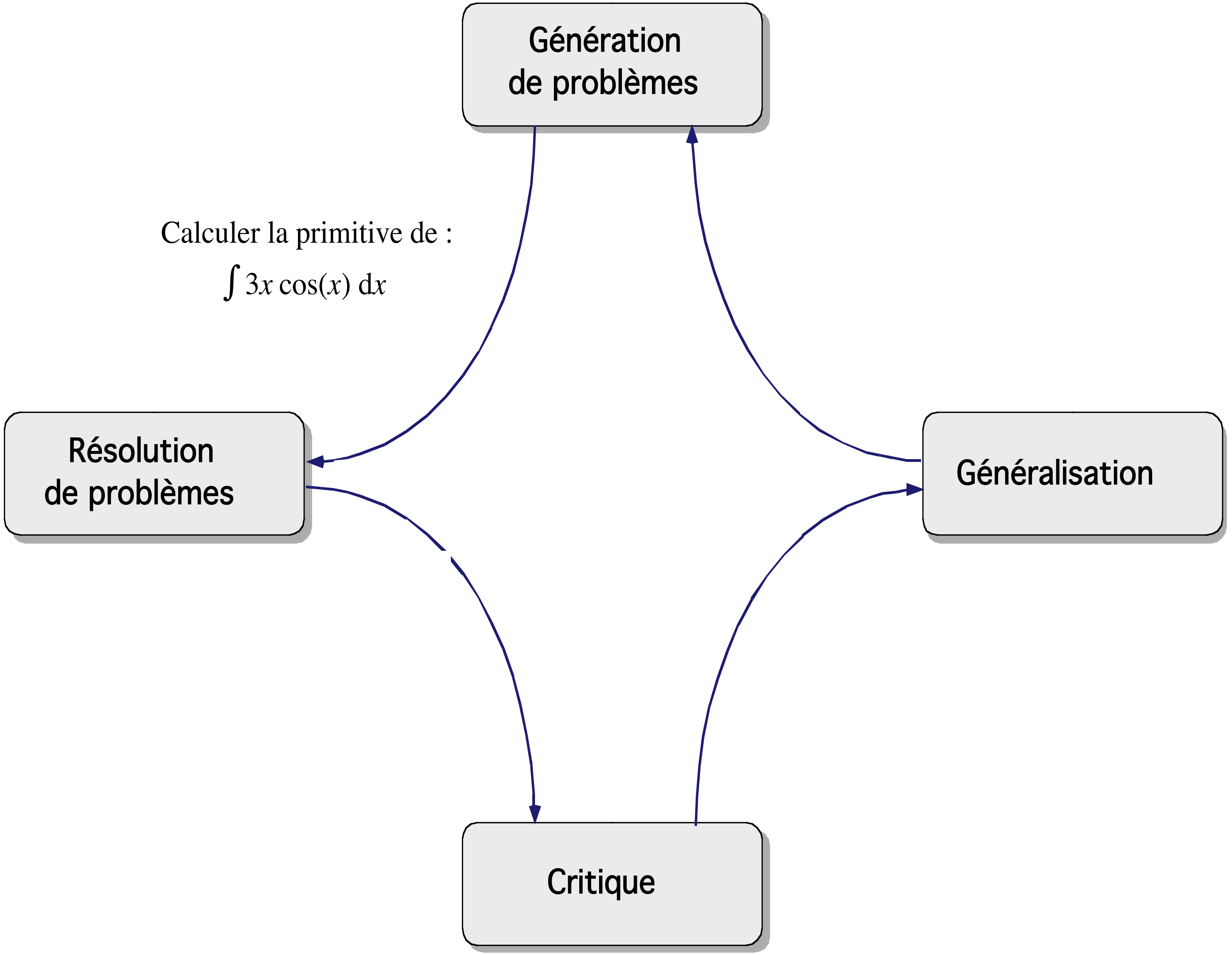


Illustration: LEX (Tom Mitchell)

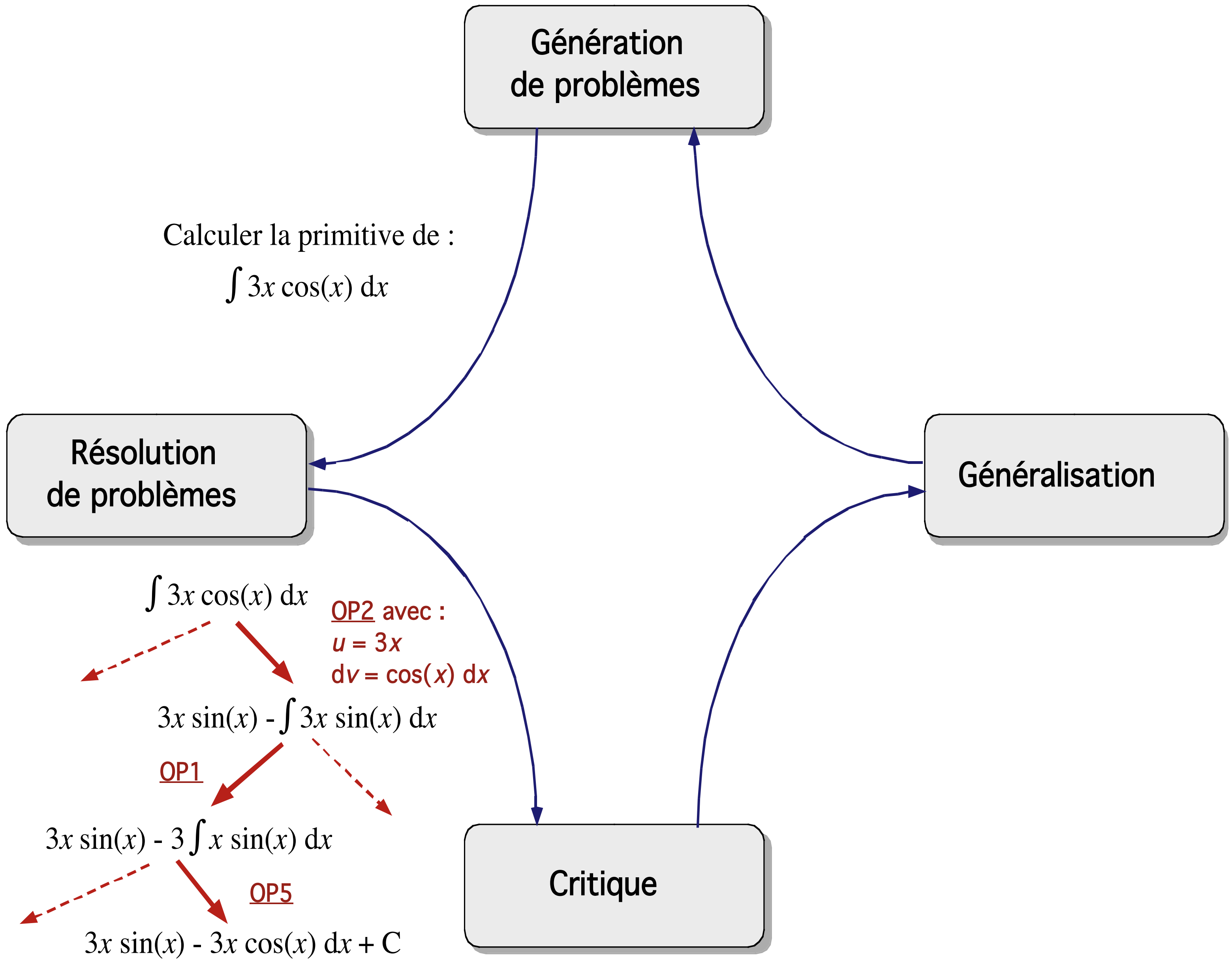


Illustration: LEX (Tom Mitchell)

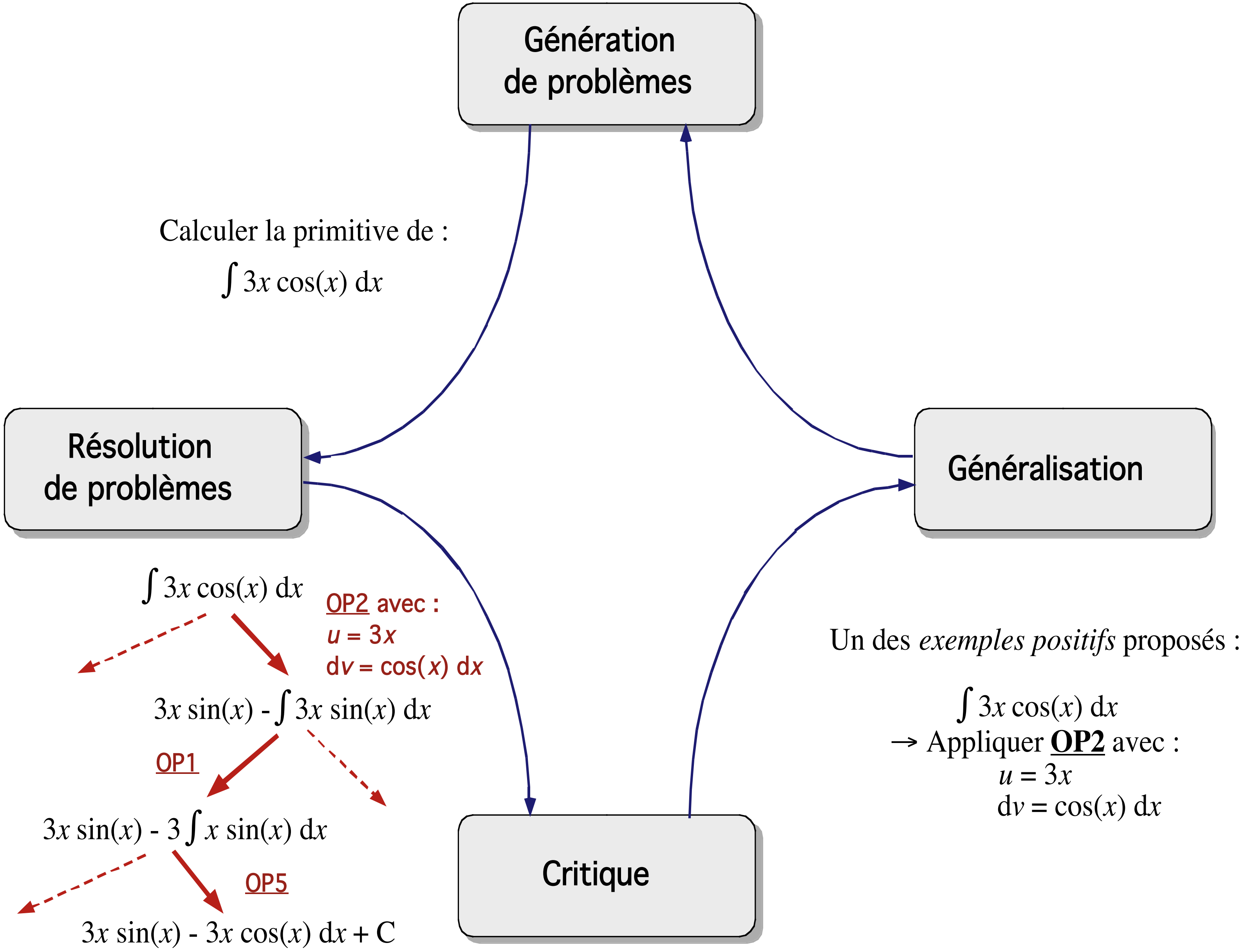
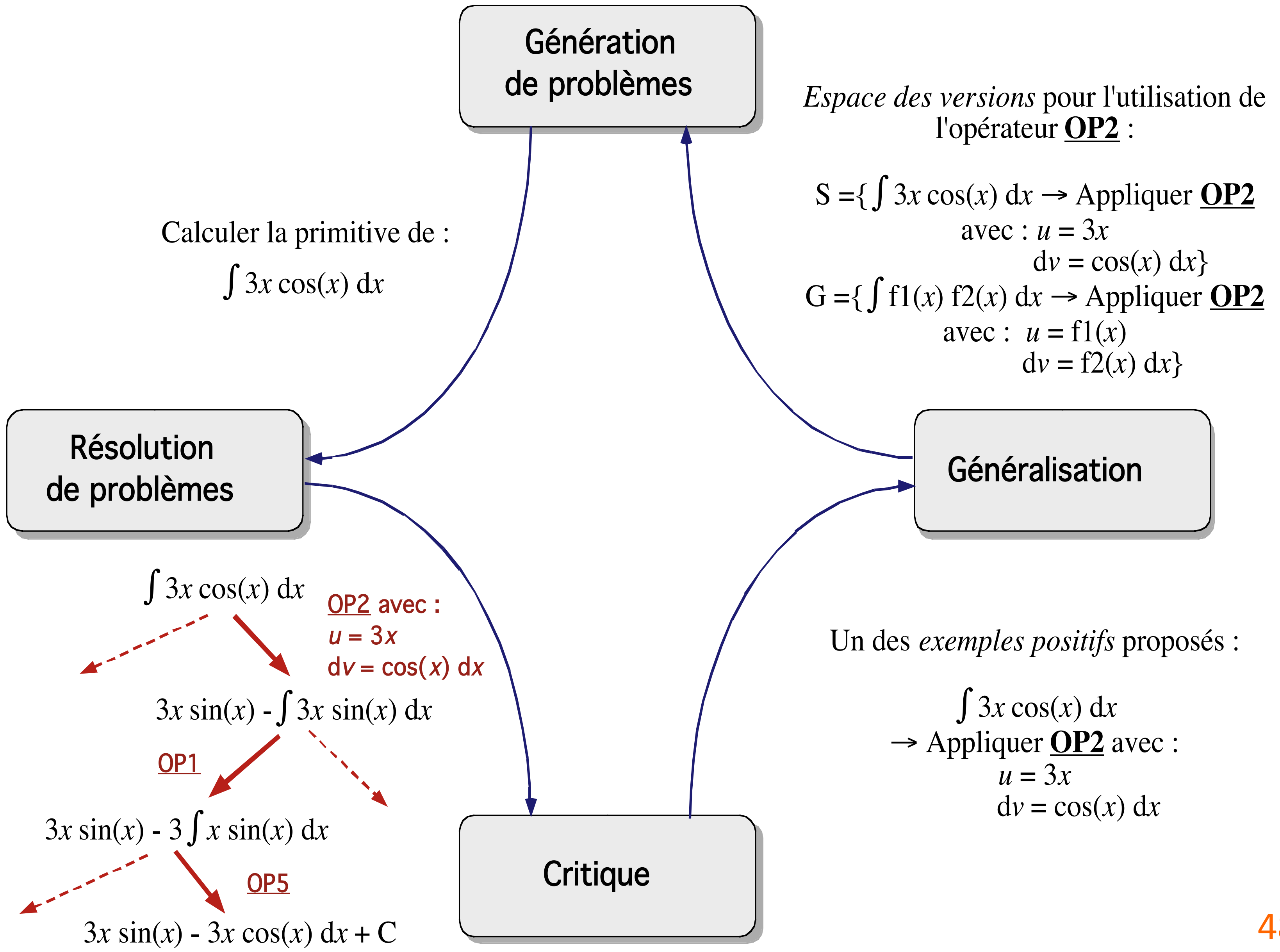


Illustration: LEX (Tom Mitchell)



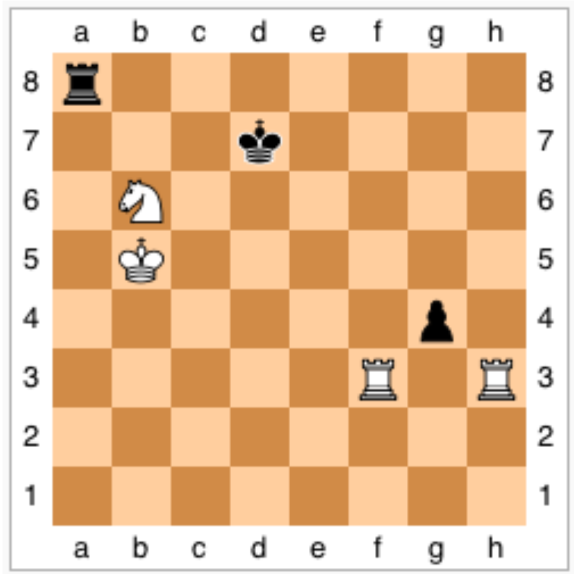
Learning from **one example**

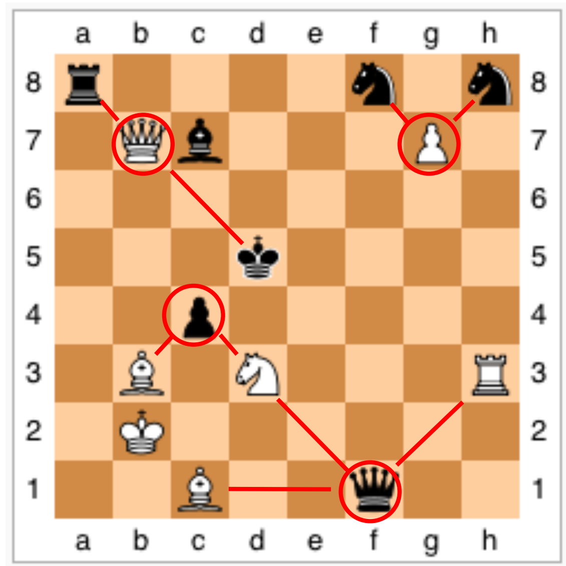
Explanation-Based Learning

1. From a **single example**

2. **Try to prove** the “fork”

3. **Generalize**



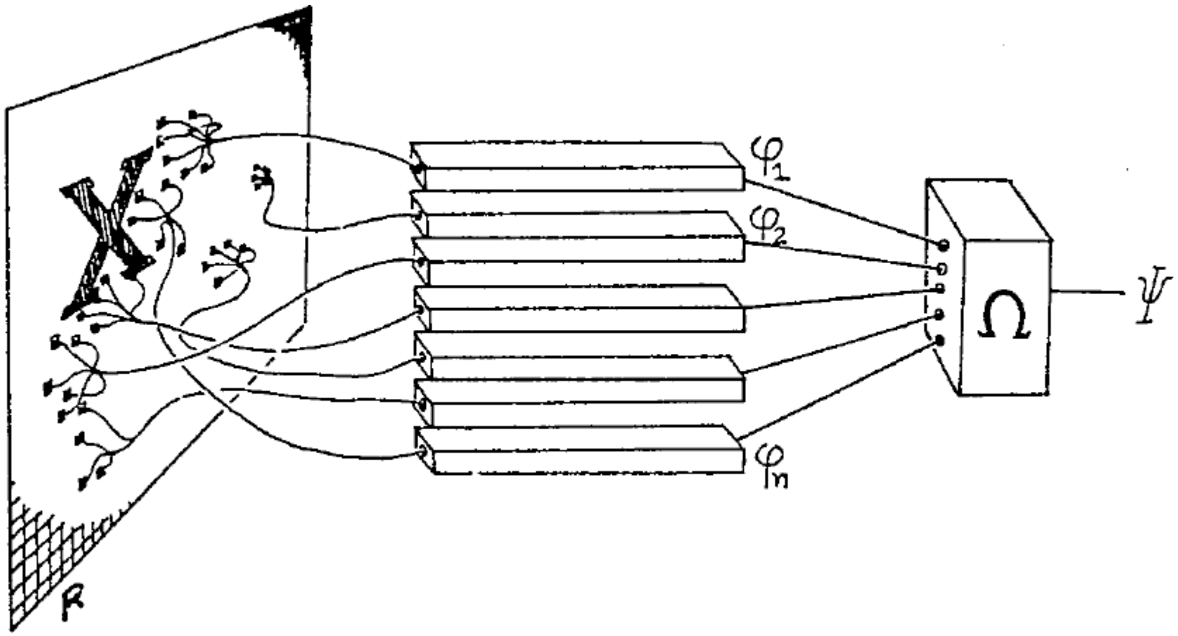


2- Apprendre à **prédire**

(Corrélations)

Premier connexionnisme : le perceptron

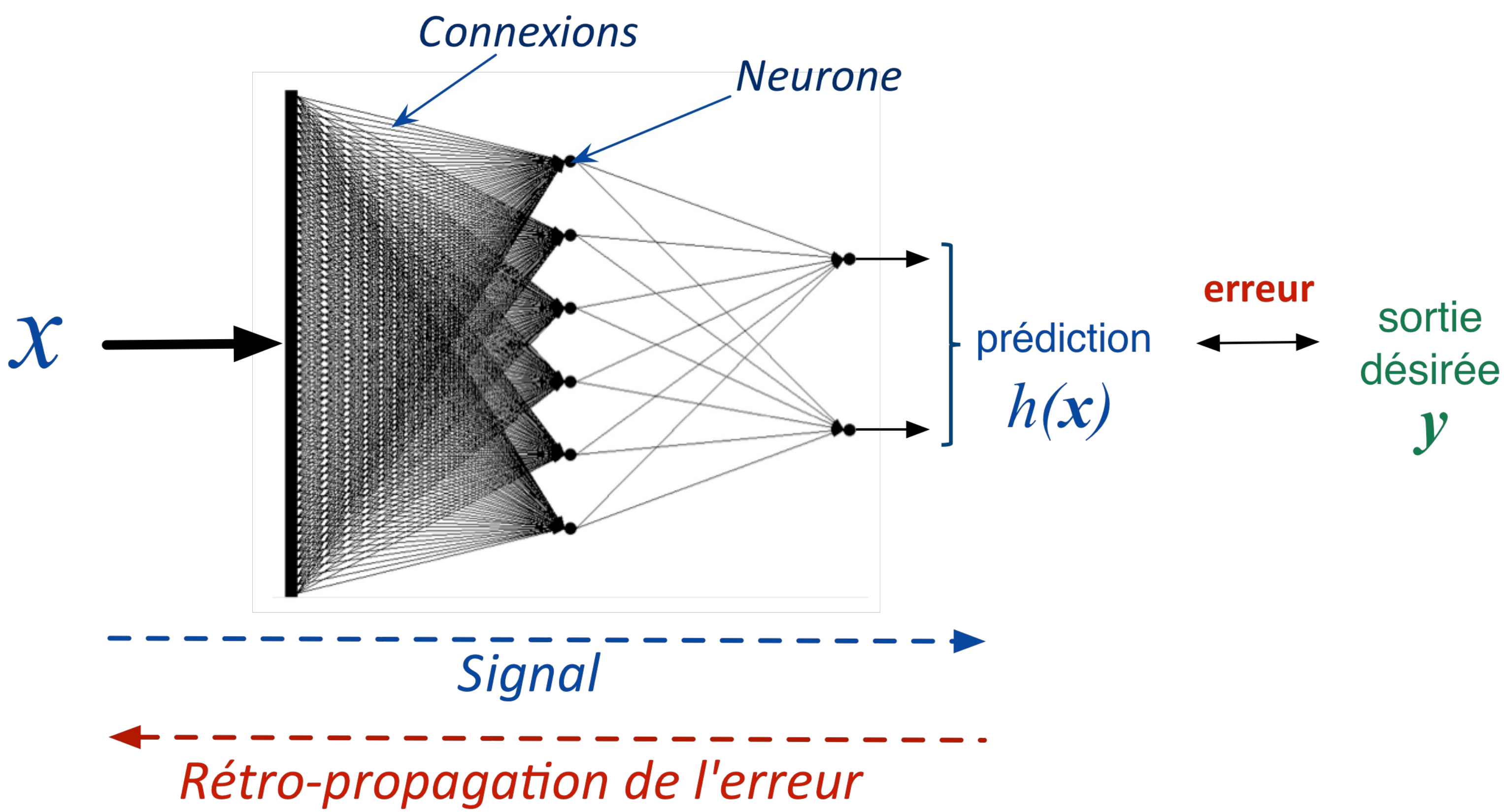
- Frank Rosenblatt (1958 – 1962)



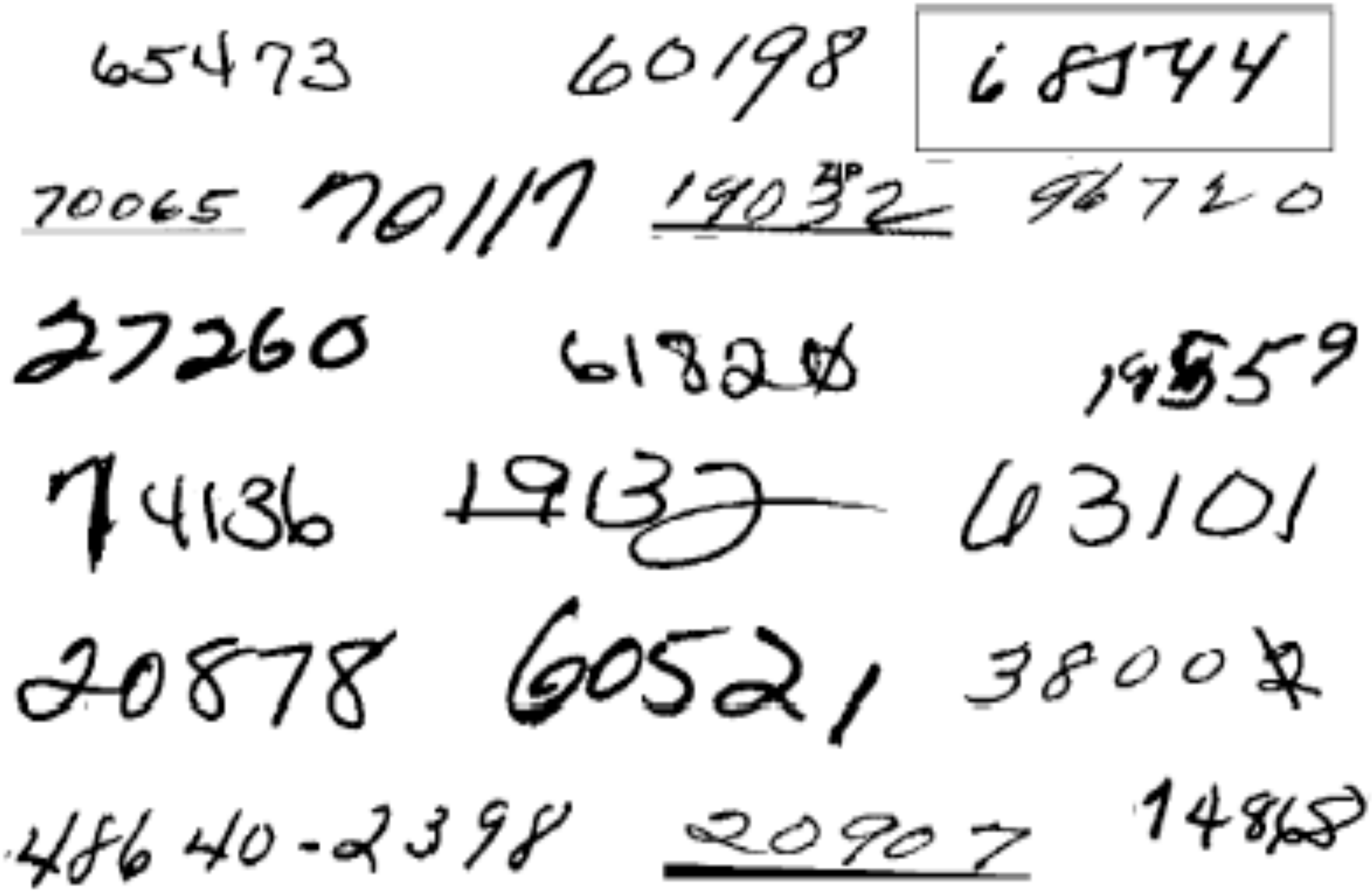
$$\Psi(\mathbf{x}) = \sum_{i=1}^{n} w_i \; \phi_i(\mathbf{x})$$

Learning with Multi-Layer Perceptrons

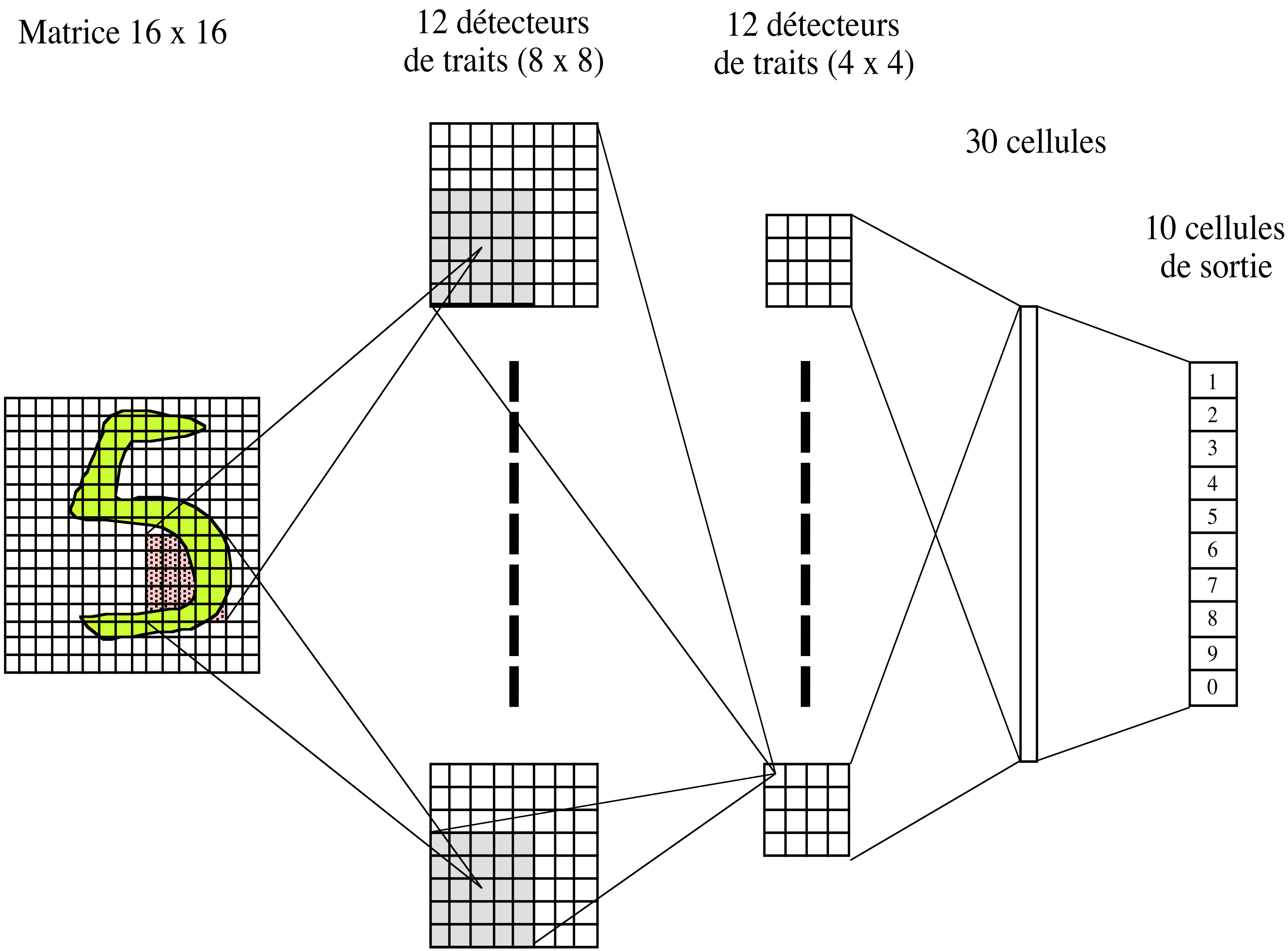
- Questions:
  - How to learn the **parameters** (weights of the connections) ?
  - How to set the **architecture** of the network?



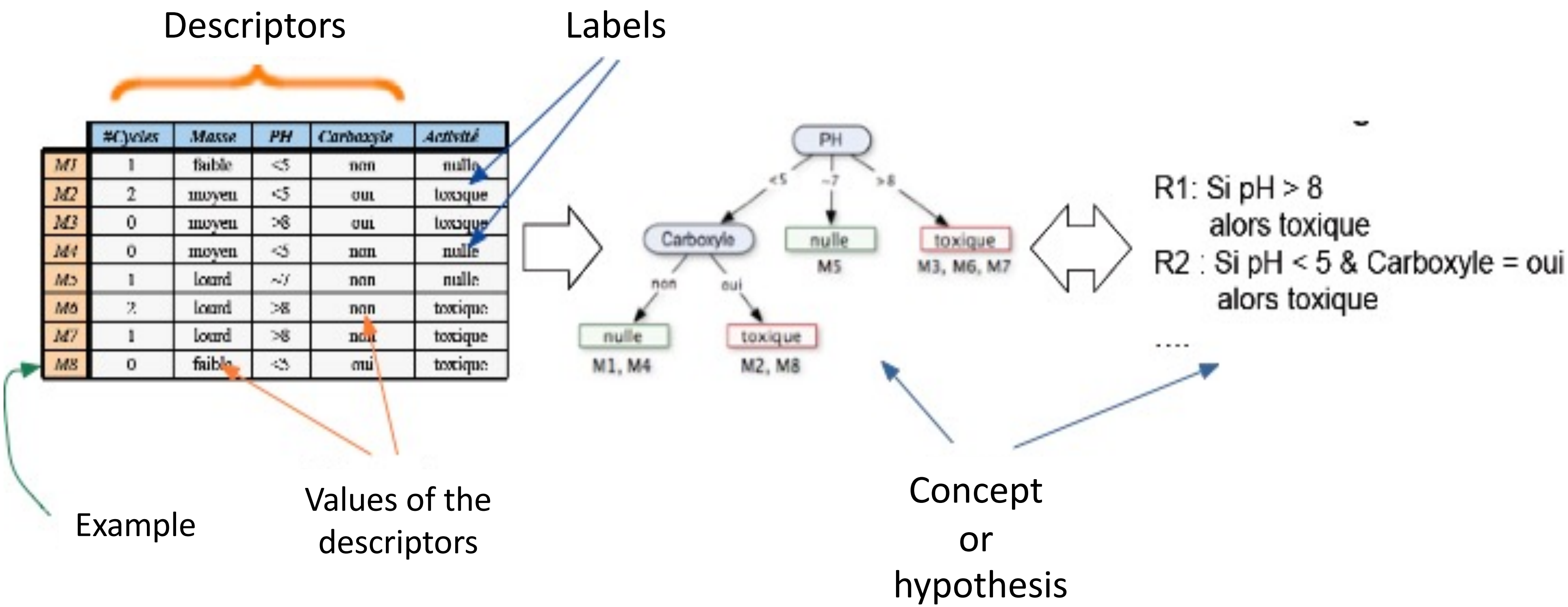
La base de données



Réseaux à convolution : Application aux codes postaux



Supervised Induction



	#Cycles	Masse	PH	Carboxyle	Activité
M1	1	faible	<5	non	nulle
M2	2	moyen	<5	oui	toxique
M3	0	moyen	>8	oui	toxique
M4	0	moyen	<5	non	nulle
M5	1	lourd	~7	non	nulle
M6	2	lourd	>8	non	toxique
M7	1	lourd	>8	non	toxique
M8	0	faible	<5	oui	toxique

55

Perspective historique : apprentissage automatique

1990s

- Espérance d'erreur de prédiction
  - Bases de données supposées i.i.d.
  - Données bruitées
  - Espérance du coût de prédiction
  - Théorie : quels principes inductifs ? Donc quel problème d'optimisation
  - Comment accélérer le processus d'optimisation : convexification
  - Nouvelles méthodes : SVM / Boosting / random forests / Lasso, …
  - Plus de raisonnement !!!

2000s

- Applications à des problèmes « réels »

2010s

2020s

Perspective historique : apprentissage automatique

1990s

- Espérance d'erreur de prédiction
  - Bases de données supposées i.i.d.
  - Données bruitées
  - Espérance du coût de prédiction
  - Théorie : quels principes inductifs ? Donc quel problème d'optimisation
  - Comment accélérer le processus d'optimisation : convexification
  - Nouvelles méthodes : SVM / Boosting / random forests / Lasso, …
  - Plus de raisonnement !!!

2000s

- Applications à des problèmes « réels »

2010s

- L'apprentissage « profond » : apprendre les (hiérarchies de) descripteurs
  - Performances en prédiction
  - Exhiber les descripteurs (et espaces « sémantiques »)
  - Transfert

2020s

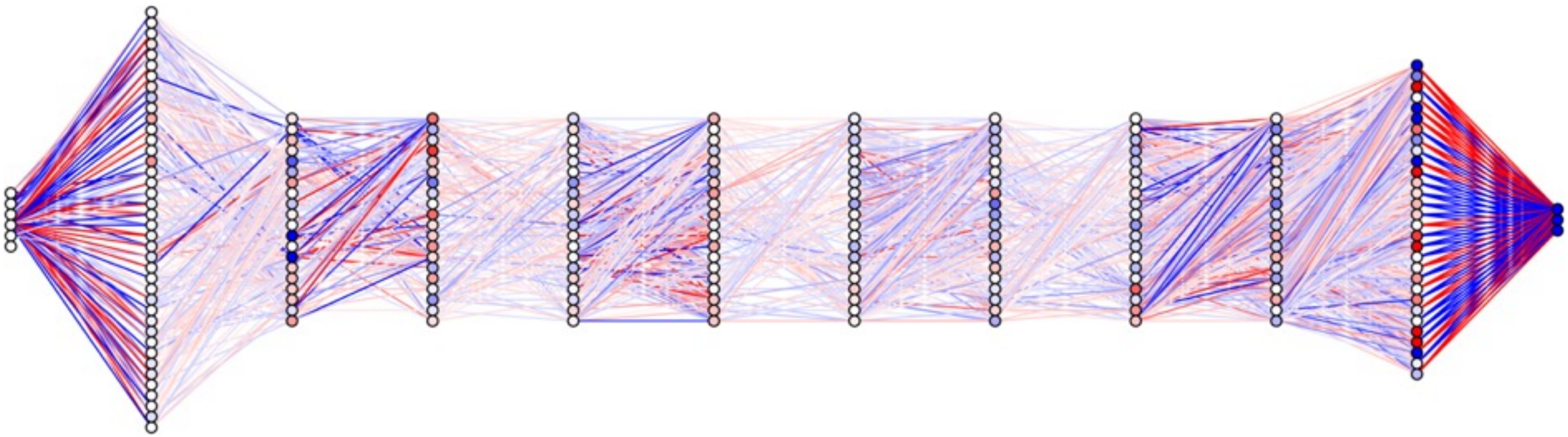
- … ?

L’apprentissage “**profond**”

Des réseaux de neurones artificiels

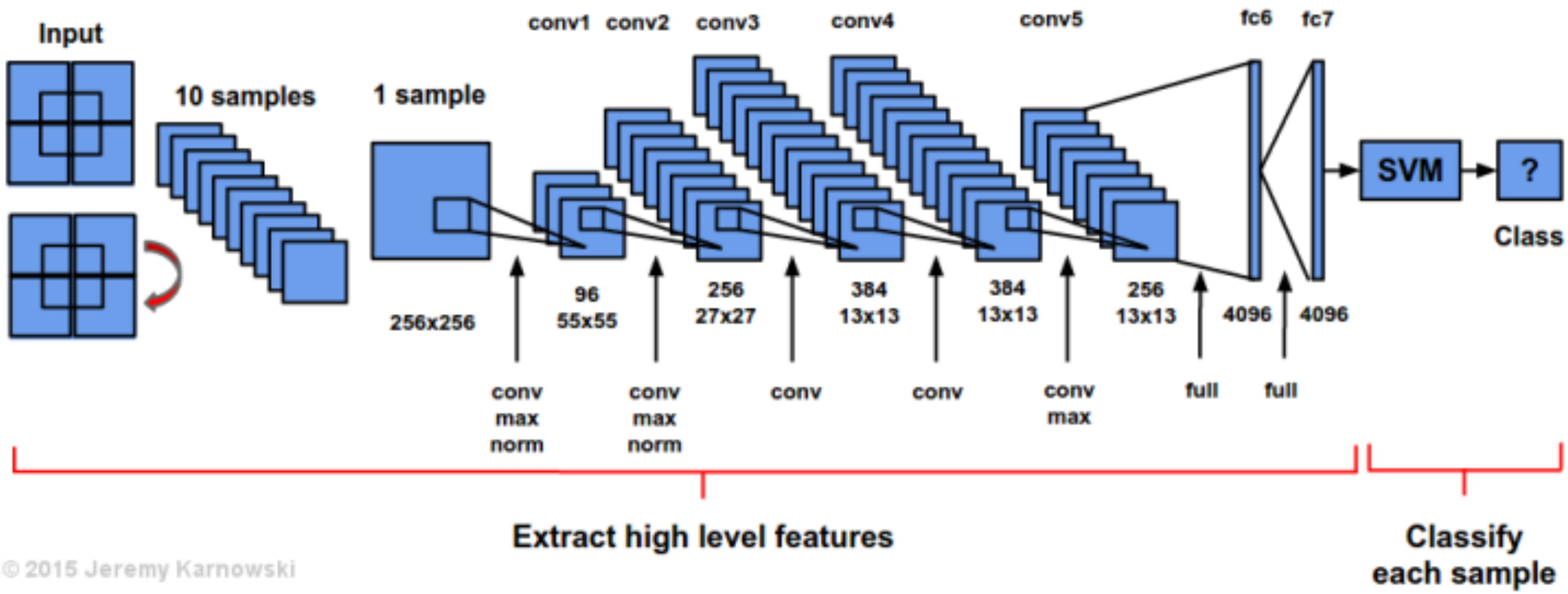
Les « réseaux de neurones **profonds** »

- Des réseaux de neurones artificiels
  - à grand nombre de couches (parfois > qqs 100)
  - et **très grand nombre de paramètres** (qqs 10^7 – 10^8 paramètres)



AlexNet

- Illustration



- Illustration

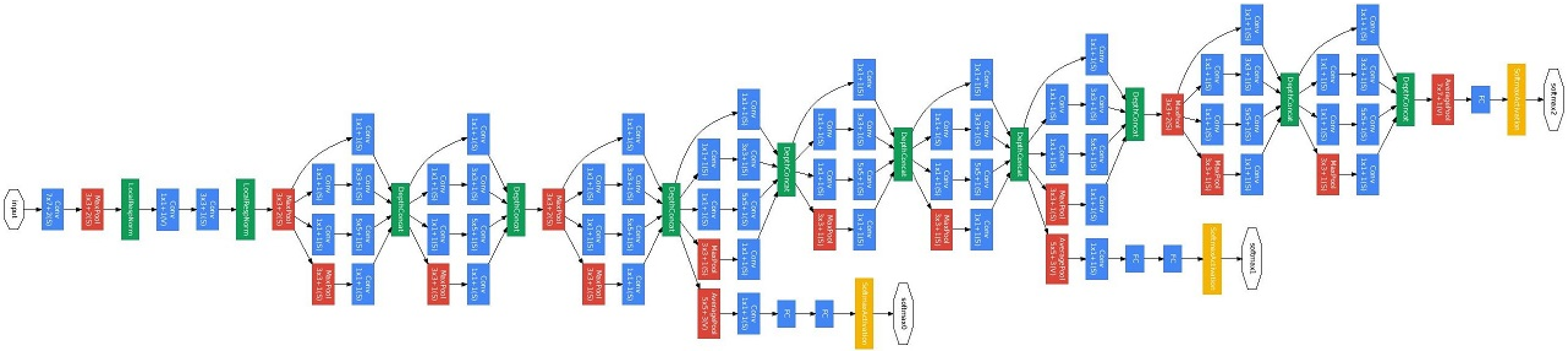
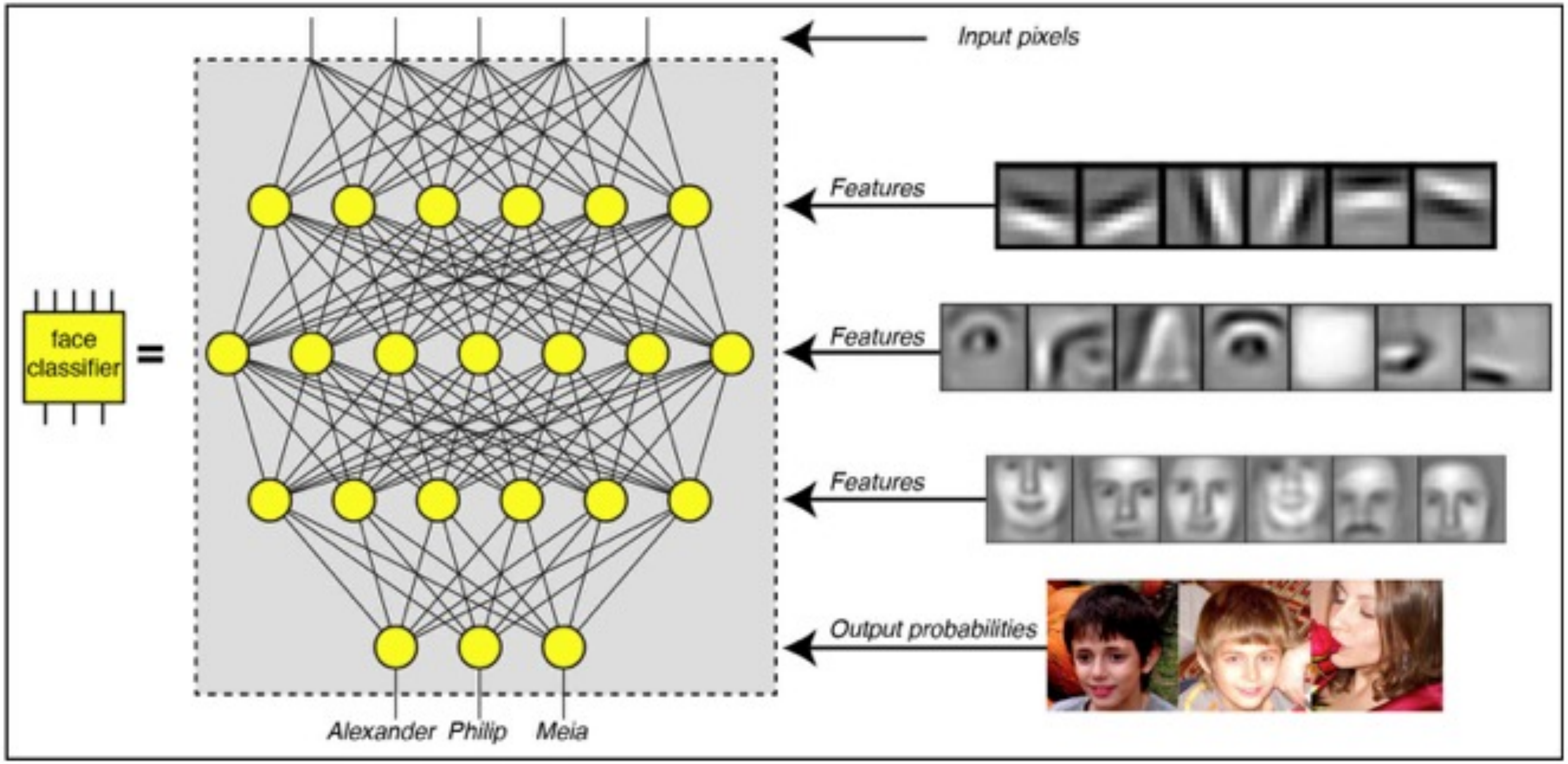
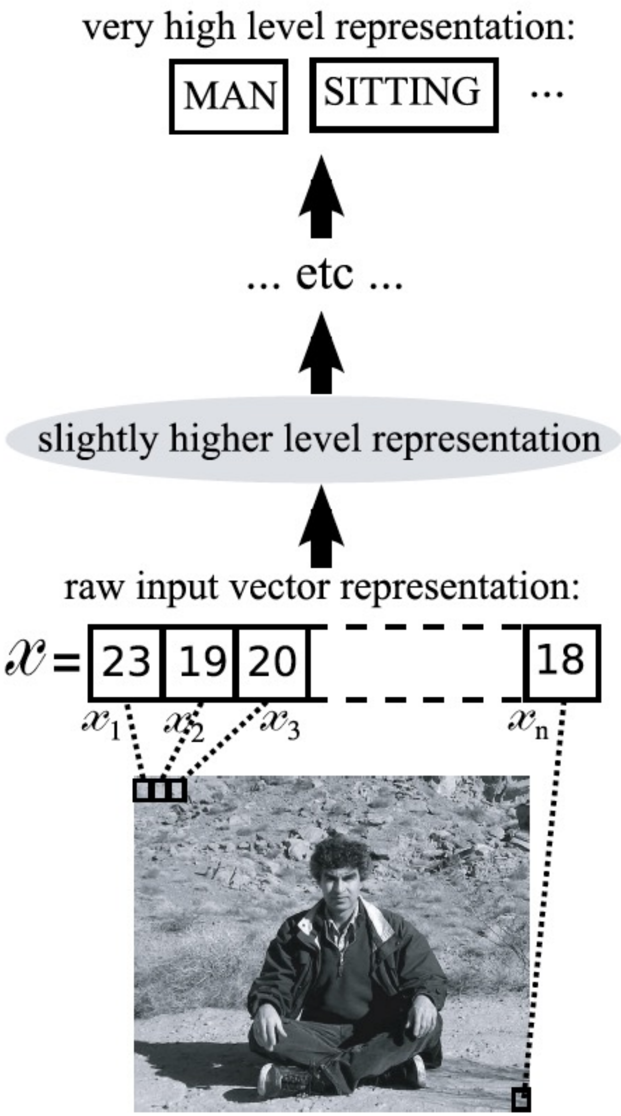


Image recognition

- Illustration



« Deep belief networks »



The ImageNet competition

- Over 15M labeled high resolution images
- Roughly 22K categories
- Collected from the Web and labeled by Amazon Mechanical Turk

Goal

- Image classification

Segmentation sémantique d'image

Model trained with a maximum range of 40m and EFS.

- Pour les véhicules autonomes

Le cas AlphaGo

- Un joueur « extraterrestre »
- Un jeu stupéfiant
- Révolutionne la manière de jouer
- Effervescence dans les écoles de go

What current Inductive Learning is **good at**

2. Discover prediction rules based on statistical correlations (PREDICTIVE learning)

   – Geared towards **minimizing prediction errors**

   – In **stationary** environments

   – **Statistical** correlations: needs lots of data and ...

Is this less of a car
because the context is wrong?

Outline

1. La perspective de l’Intelligence Artificielle
2. Notions de base en science des données
3. L’induction d’arbres de décision
4. L’induction d’arbres de régression
5. Les forêts aléatoires

What do you expect?

- Better **understand** your data

What do you expect?

- Better **understand** your data

- Be able to make **prediction**

What do you expect?

- Better **understand** your data
- Be able to make **prediction**
- Be able to make **prescription**

What do you expect?

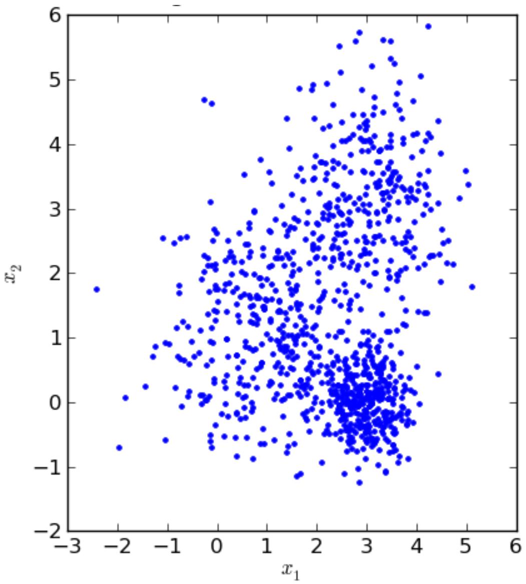
- Better **understand** your data

- Be able to make **prediction**

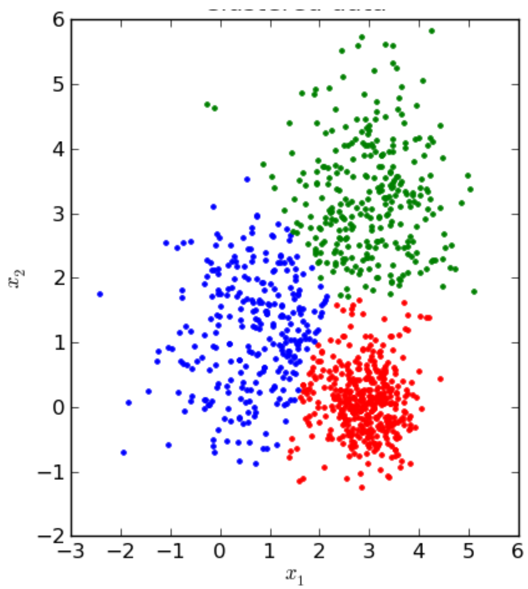
- Be able to make **prescription**

(1) **Understand** your data

- **Re-express** it
  - In a concise way
  - To be interpretable by an expert of the domain



Original data



Clustered data

Three groups of customers with such and such **characteristics** ...

(2) Make **predictions**

- Extrapolate your data to find **predictive correlations**
  - From a **training** set $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_m, y_m)\}$

(2) Make **predictions**

- Extrapolate your data to find predictive correlations
  - From a **training** set $S = \{(x_1, y_1), (x_2, y_2), \dots , (x_i , y_i), \dots , (x_m, y_m)\}$

$$S = \{(x_1, y_1), (x_2, y_2), \dots , (\overbrace{x_i , y_i}^{f}), \dots , (x_m, y_m)\}$$

(2) Make **predictions**

- Extrapolate your data to find predictive correlations
  - From a **training** set $S = \{(x_1, y_1), (x_2, y_2), \ldots , (x_i , y_i), \ldots , (x_m, y_m)\}$

$$S = \{(x_1, y_1), (x_2, y_2), \ldots , (\underbrace{\overbrace{x_i , y_i}^{f}}_{h}), \ldots , (x_m, y_m)\}$$

(2) Make **predictions**

- Extrapolate your data to find predictive correlations
  - From a **training** set $S = \{(x_1, y_1), (x_2, y_2), \dots , (x_i , y_i), \dots , (x_m, y_m)\}$

$$S = \{(x_1, y_1), (x_2, y_2), \dots , (\underbrace{\overbrace{x_i , y_i}^{f}}_{h}), \dots , (x_m, y_m)\}$$

New x --> y ?

(2) Make **predictions**

- Extrapolate your data to find predictive correlations
  - From a **training** set $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots, (x_m, y_m)\}$

$$S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_i \overset{f}{\curvearrowright} y_i), \ldots, (x_m, y_m)\}$$

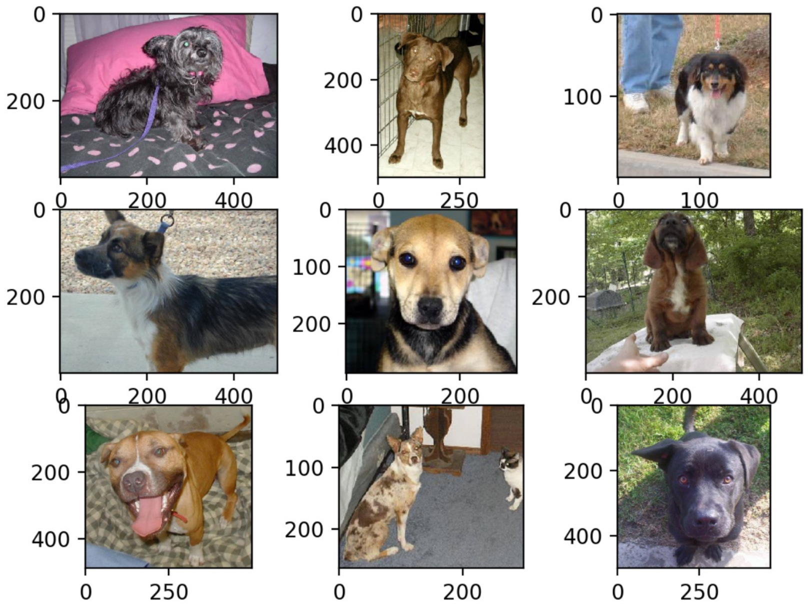
h

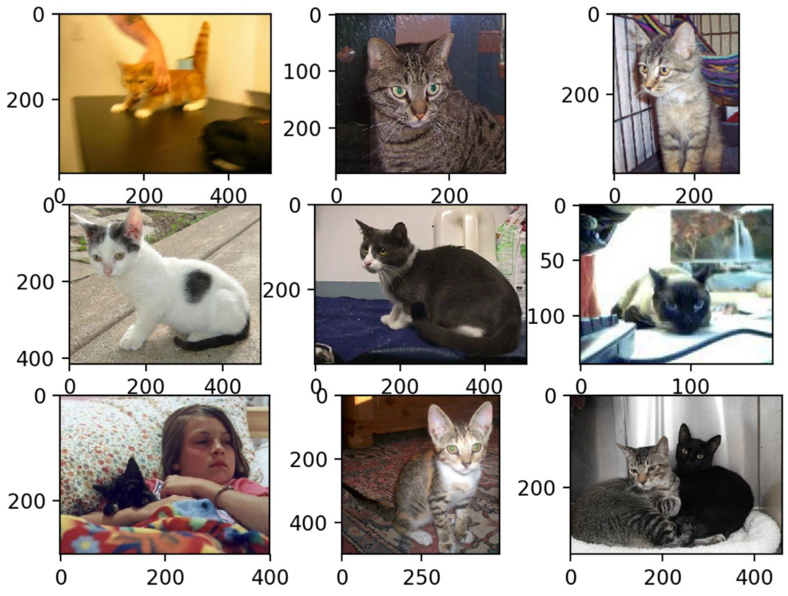
$$x - h \to y$$

(2) Machine Learning as ...

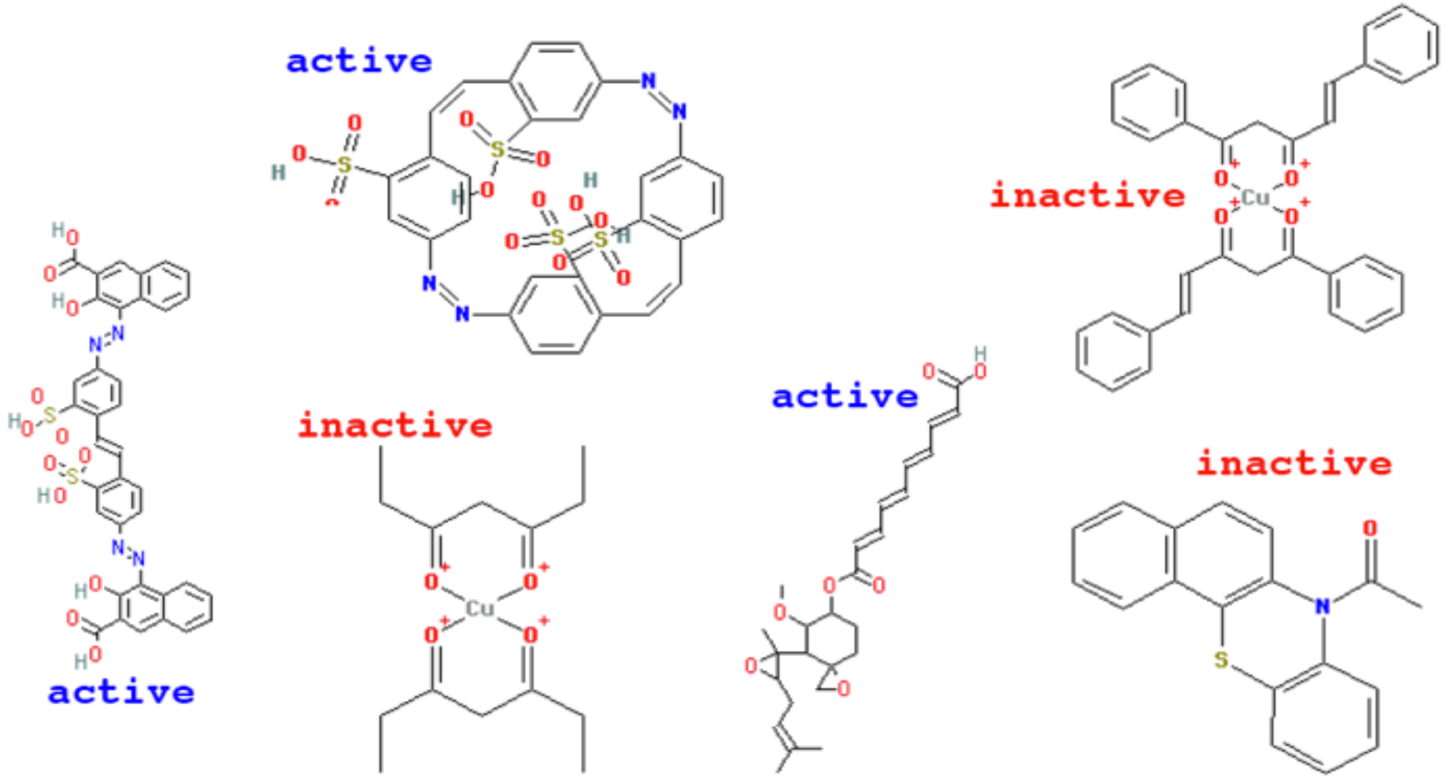
... Learning a function from an **input** space X to an **output** space Y

Cats vs. **dogs**





Supervised learning



NCI AIDS screen results (from http://cactus.nci.nih.gov).

(3) Make **prescriptions**

- Learning causal relationships
  - The barometer **allows the prediction** of tomorrow's weather
  - But tampering with its needle **will not** change the weather

Correlation **is not** causality

Discovering causal relationships

(generally) requires knowing **more than the data**

How do you **evaluate** the **results?**

How do you **evaluate** the **results?**

- **Descriptive** learning

How do you **evaluate** the **results?**

- **Descriptive** learning
  - **Validation** by the **expert**

How do you **evaluate** the **results?**

- **Descriptive** learning
  - **Validation** by the **expert**
  - The expert can be wrong
    - **Wants to see things** that are not really there
    - **Blind** to interesting but out of the blue patterns

How do you **evaluate** the **results?**

- **Descriptive** learning
  - **Validation** by the **expert**
  - The expert can be wrong
    - **Wants to see things** that are not really there
    - **Blind** to interesting but out of the blue patterns

Descriptive learning usually takes place in an **exploratory phase**

→ Be **very careful**

How do you **evaluate** the **results?**

- **Predictive** learning

How do you **evaluate** the **results?**

- **Predictive** learning
  - **Predictive** performance (on a test set)
    - E.g. error rate

How do you **evaluate** the **results?**

- **Predictive** learning
  - **Predictive** performance (on a test set)
    - E.g. error rate
  - But, this is not all there is to it
    - **Interpretability** of the **model**
    - Explanation/**justification** of the **prediction**
    - **Fruitfulness** wrt. the **domain theory**

How do you **evaluate** the **results?**

- **Predictive** learning
  - **Predictive** performance (on a test set)
    - E.g. error rate
  - But, this is not all there is to it. We want also
    - **Interpretability** of the **results**
    - **Interpretability** of the **model** and the **process**
    - Gaining a **better understanding of the world** when including the learned decision function in an existing theory

Often, we are **not** interested in prediction alone,

but in **understanding** the **prediction** and/or the **predictive model**

We are **thinking** beings

A **basic** principle

- Machine Learning “just” **reformulates** what has been given as input
- A conservation theorem:
  - **No information is “added**
  - Data + prior knowledge

A **basic** principle

- Machine Learning "just" reformulates what has been given as input
- A conservation theorem:
  - **No information is "added"**
  - Data + prior knowledge

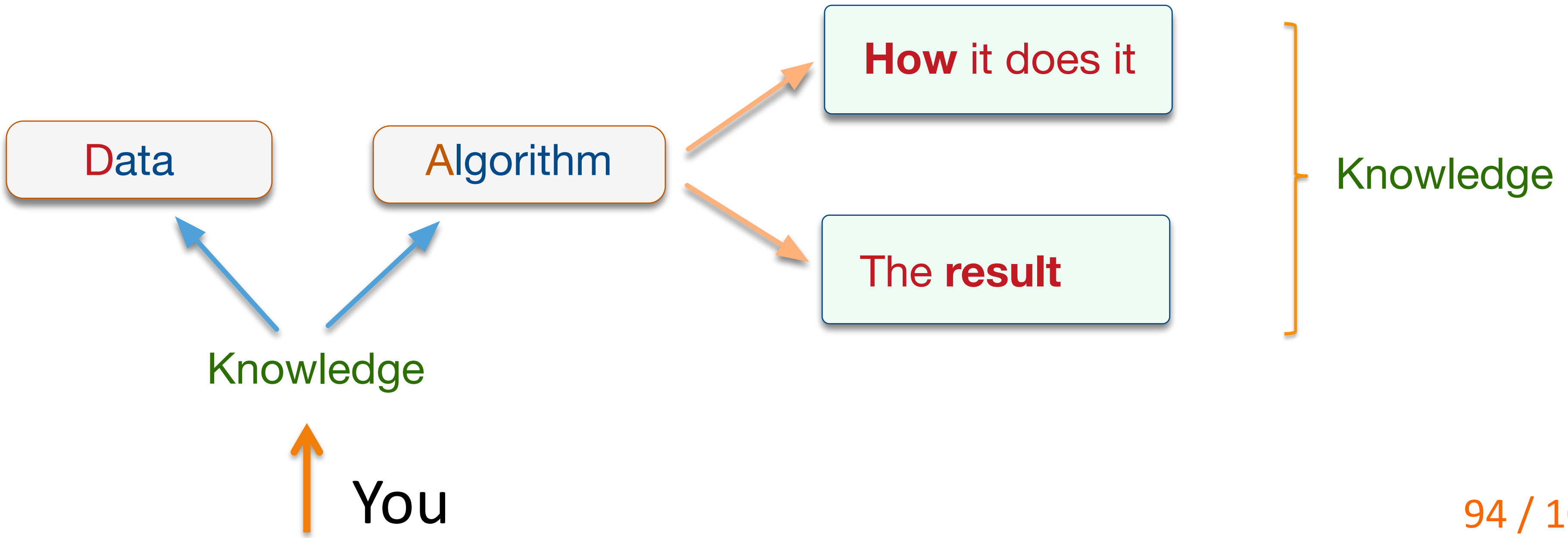
Little data + **lots** of prior knowledge

Big data + **less** prior knowledge

A basic principle

- Machine Learning "just" reformulates what has been given as input
- A conservation theorem:
  - **No information is "added"**
  - Data + prior knowledge

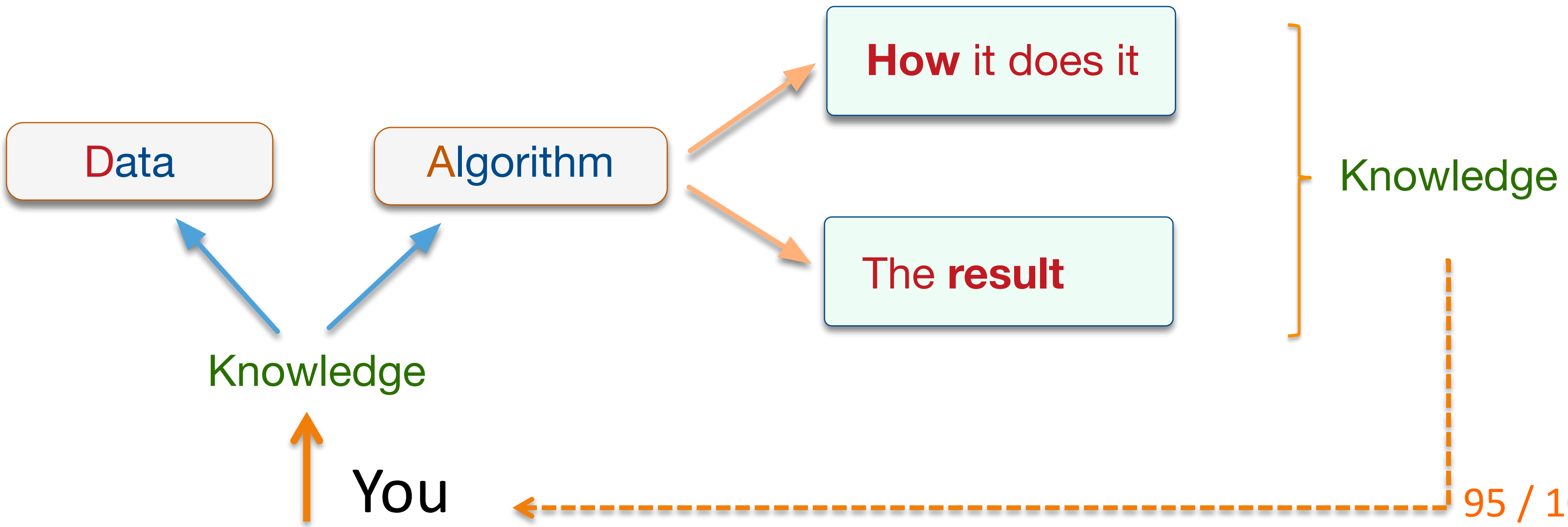
Little data + **lots** of prior knowledge
Big data + **less** prior knowledge



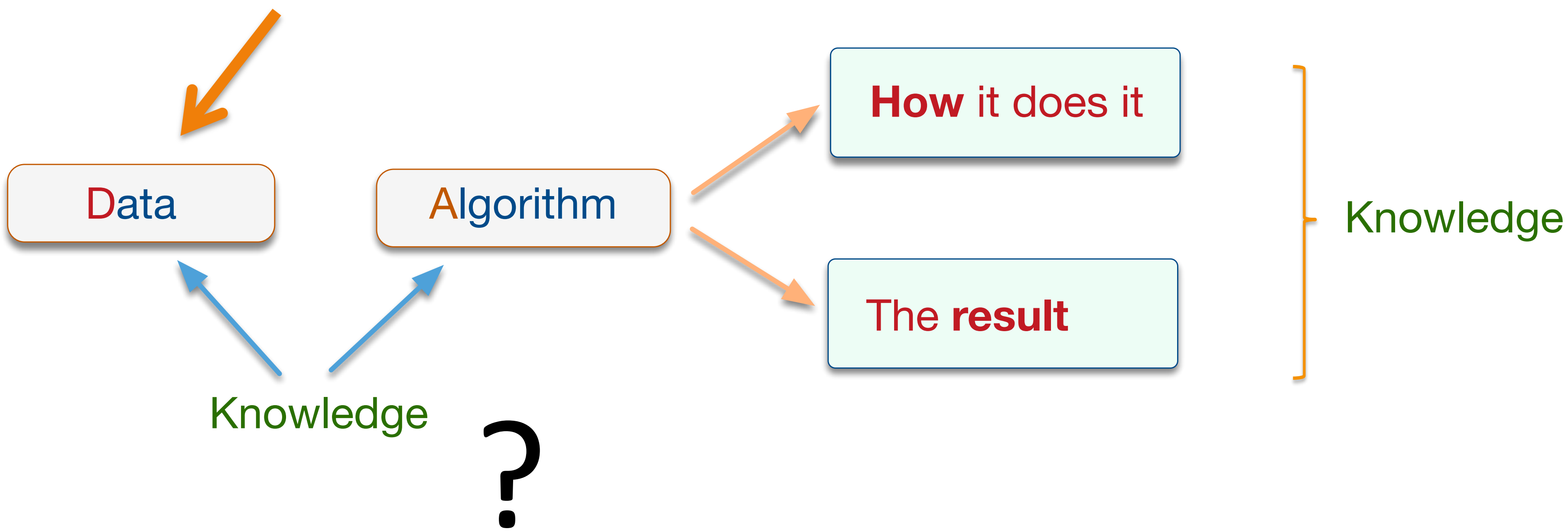
A basic principle

- Machine Learning "just" reformulates what has been given as input
- A conservation theorem:
  - **No information is "added"**
  - Data + prior knowledge

Little data + **lots** of prior knowledge
Big data + **less** prior knowledge

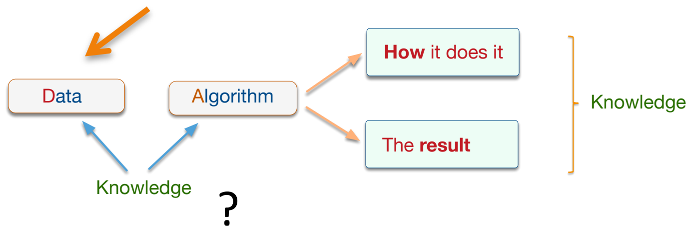


Prior knowledge



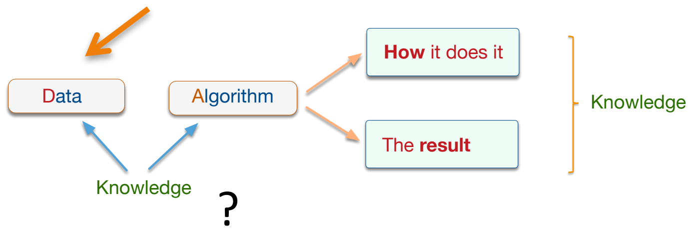
...

Knowledge as **input** to ML



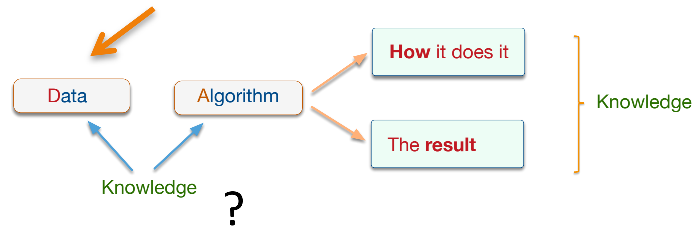
- Knowledge in the data
  - The **experimental apparatus**
  - Choice of **the descriptors** (the features)
  - **Enrichment** using ontologies
  - **Normalization** of the values
  - **Missing** values
  - Possibly **added data point**
    - With invariances in mind
  - …

Knowledge as **input** to ML



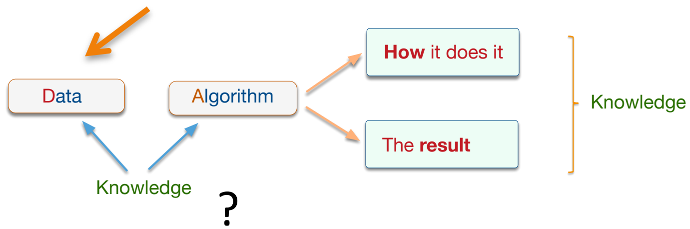
- Knowledge in the data
  - The **experimental apparatus** → With its own imperfection and biases
  - Choice of **the descriptors** (the features)
  - **Enrichment** using ontologies
  - **Normalization** of the values
  - **Missing** values
  - Possibly **added data point**
    - With invariances in mind
  - …

Knowledge as **input** to ML



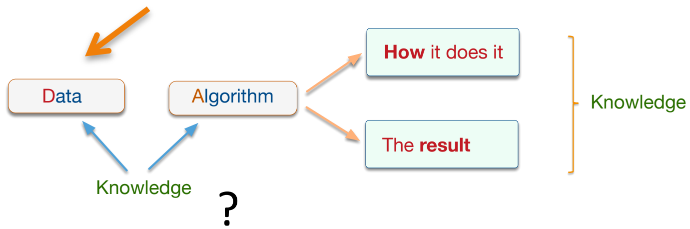
- Knowledge in the data
  - The **experimental apparatus**
  - Choice of **the descriptors** (the features) → Necessarily biased
  - **Enrichment** using ontologies
  - **Normalization** of the values
  - **Missing** values
  - Possibly **added data point**
    - With invariances in mind
  - …

Knowledge as **input** to ML



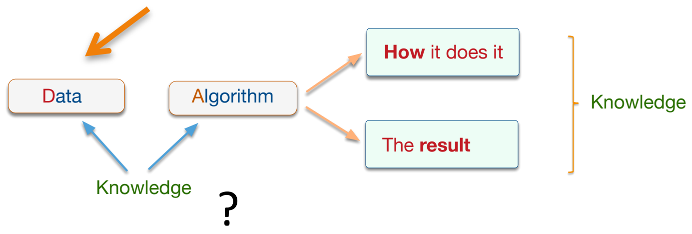
- Knowledge in the data
  - The **experimental apparatus**
  - Choice of **the descriptors** (the features)
  - **Enrichment** using ontologies → Prior (and biased) knowledge
  - **Normalization** of the values
  - **Missing** values
  - Possibly **added data point**
    - With invariances in mind
  - …

Knowledge as **input** to ML



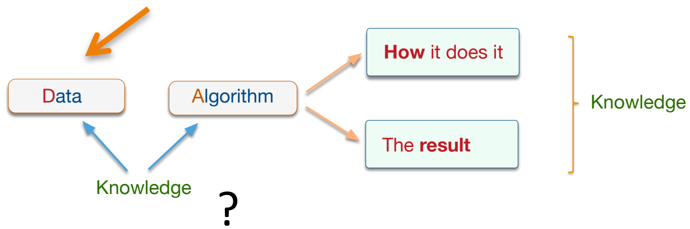
- Knowledge in the data
  - The **experimental apparatus**
  - Choice of **the descriptors** (the features)
  - **Enrichment** using ontologies
  - **Normalization** of the values → No perfect normalization
  - **Missing** values
  - Possibly **added data point**
    - With invariances in mind
  - …

Knowledge as **input** to ML



- Knowledge in the data
  - The **experimental apparatus**
  - Choice of **the descriptors** (the features)
  - **Enrichment** using ontologies
  - **Normalization** of the values
  - **Missing** values → What choice of imputation method?
  - Possibly **added data point**
    - With invariances in mind
  - …

Knowledge as **input** to ML

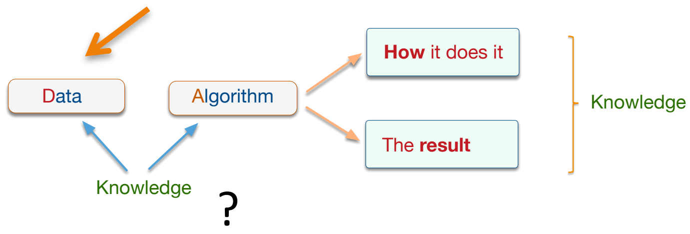


- Knowledge in the data
  - The **experimental apparatus**
  - Choice of **the descriptors** (the features)
  - **Enrichment** using ontologies
  - **Normalization** of the values
  - **Missing** values
  - Possibly **added data point**
    - With invariances in mind
  - …

How do you add points?

Choice of invariance (prior assumptions)

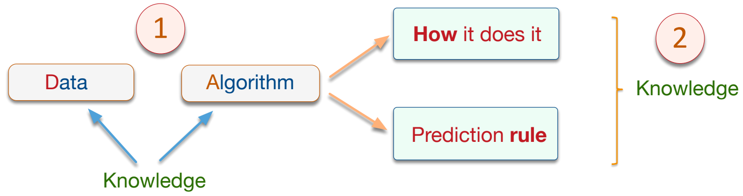
Knowledge as **input** to ML



- Knowledge in the data
  - The **experimental apparatus**
  - Choice of **the descriptors** (the features)
  - **Enrichment** using ontologies
  - **Normalization** of the values
  - **Missing** values
  - Possibly **added data point**
    - With invariances in mind
  - …

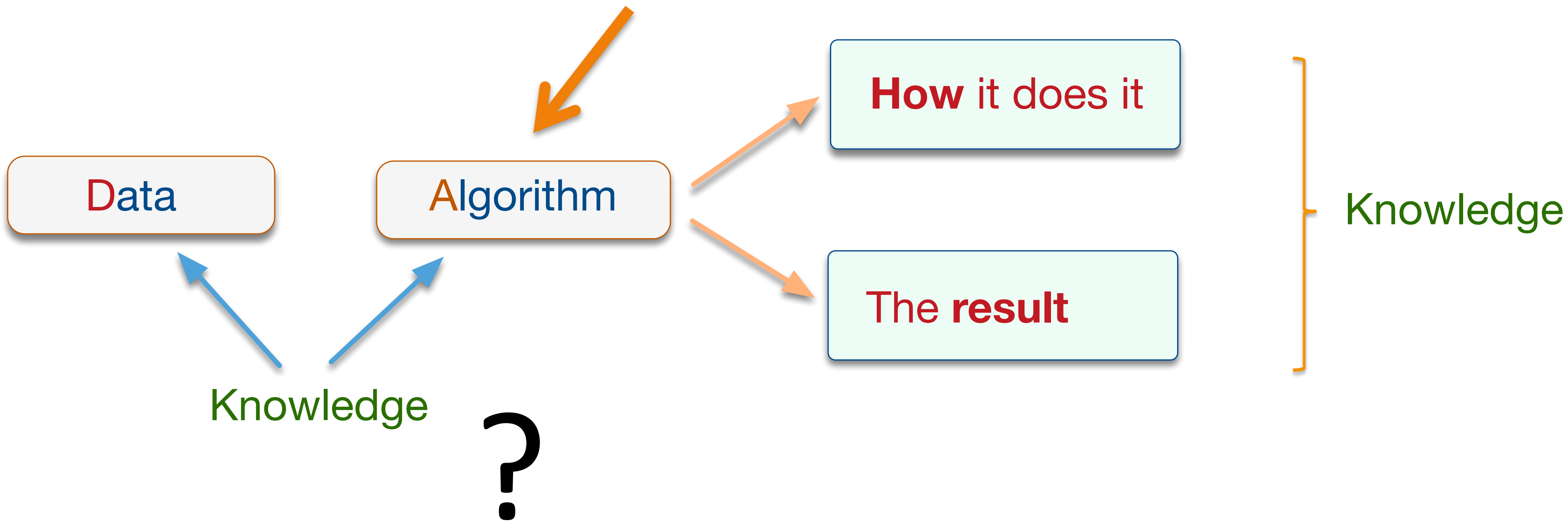
Prior assumptions everywhere

Knowledge as **input** to ML

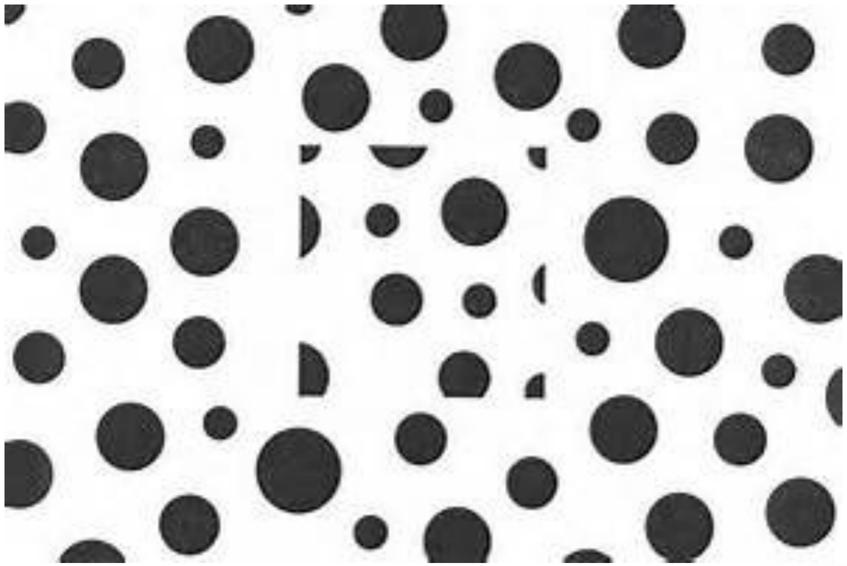


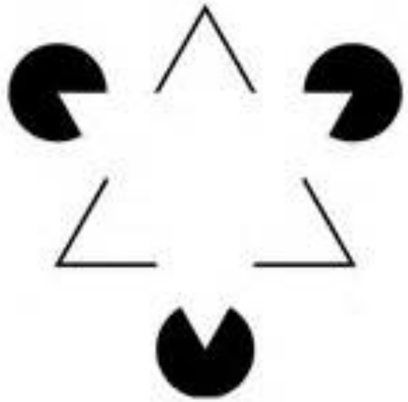
- Knowledge in the data

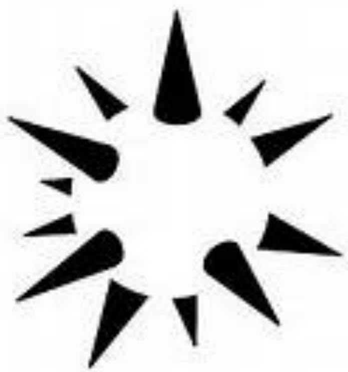
And sometime, there is not even that much **available** data!


How it does it
Data
Algorithm
The result
Knowledge
Knowledge
?

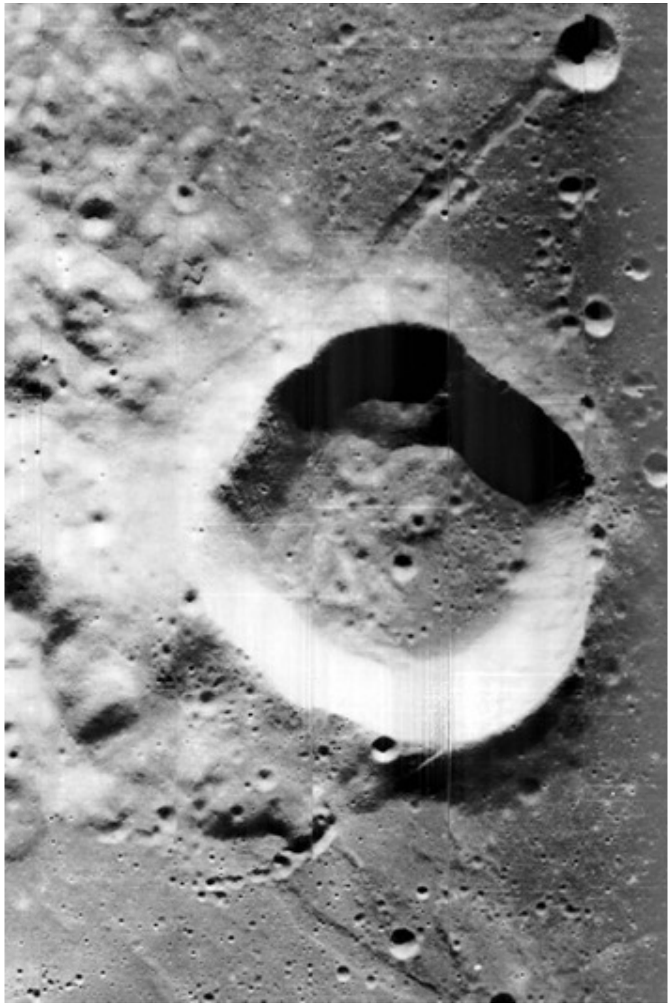
...

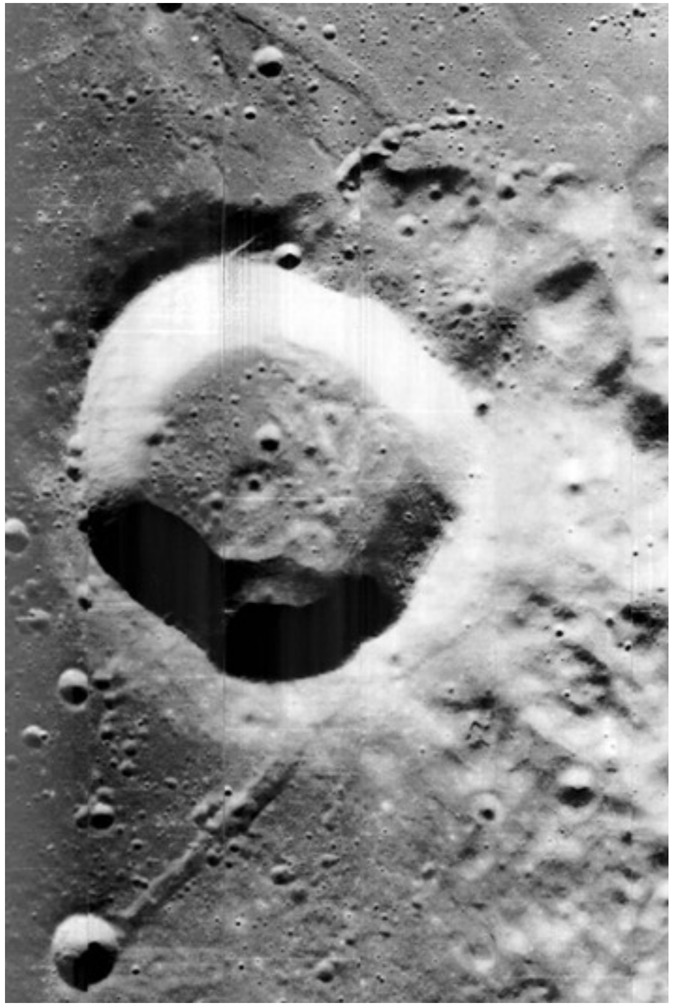






Induction et illusions

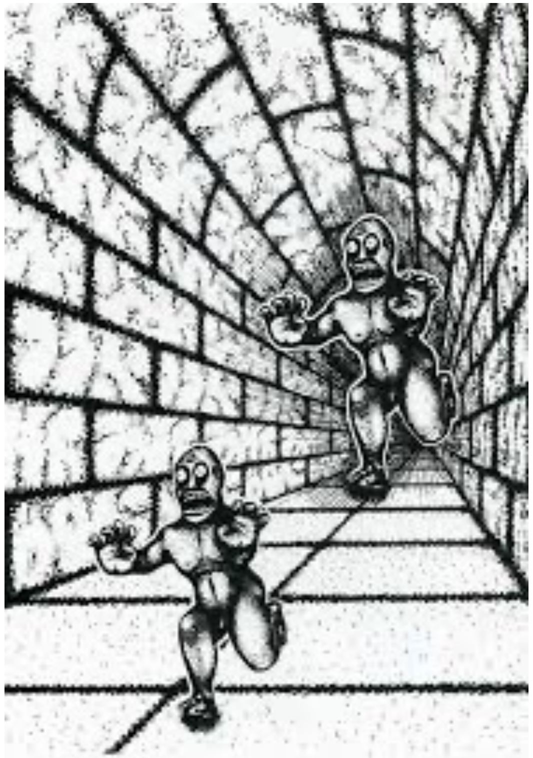


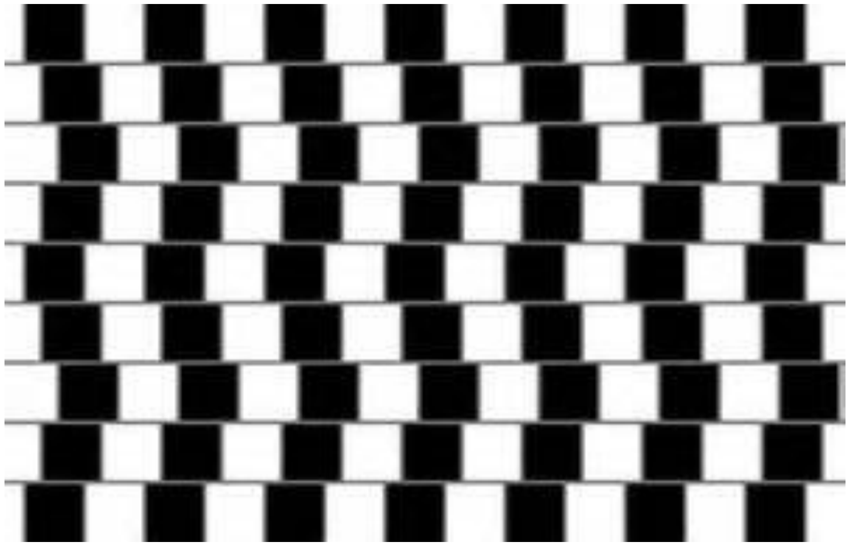


Cratère ou colline ?

Interpreting – completion of percepts

Induction and its illusions





What current Inductive Learning is **good at**

1. Identifying patterns in data (DESCRIPTIVE learning)
   - But **no guarantees** about the value of their value