What is the place of knowledge in Machine Learning?

Antoine Cornuéjols

AgroParisTech – INRA MIA 518

LINK research group

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

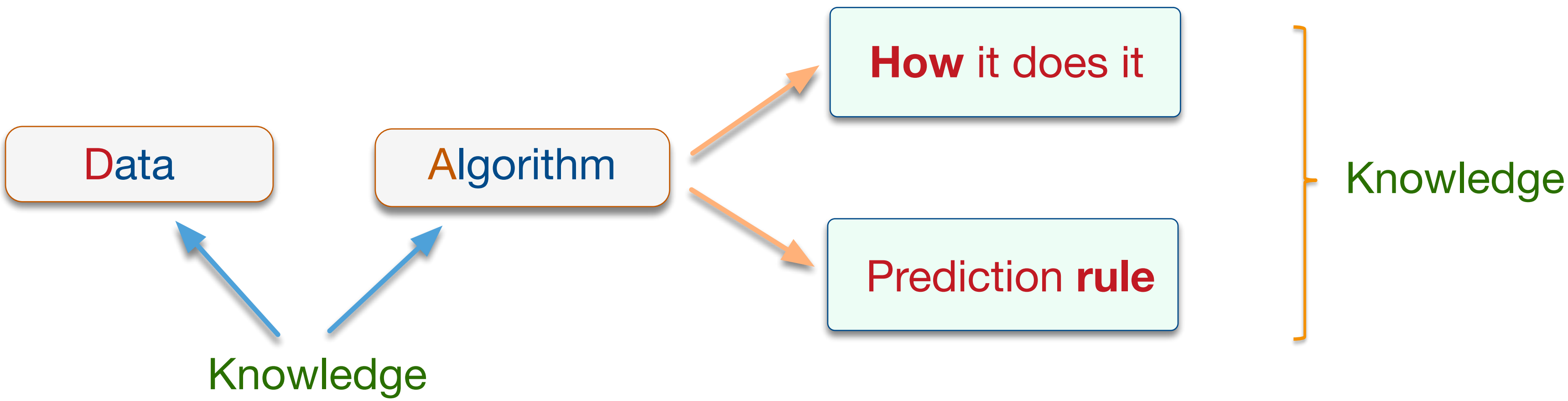
Little data + **lots** of prior knowledge
Big data + **less** prior knowledge

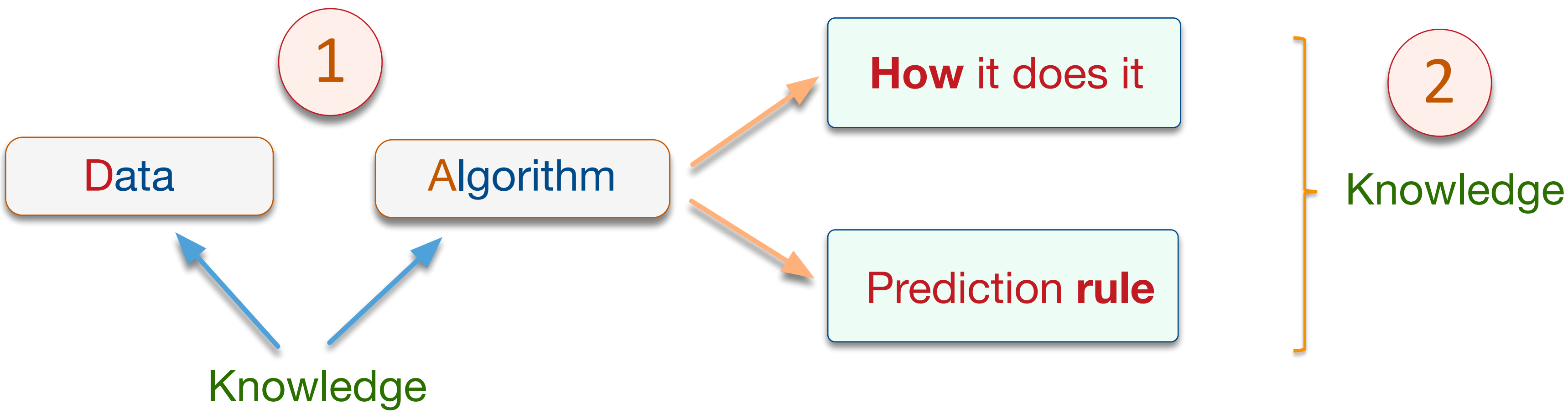
A basic principle

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

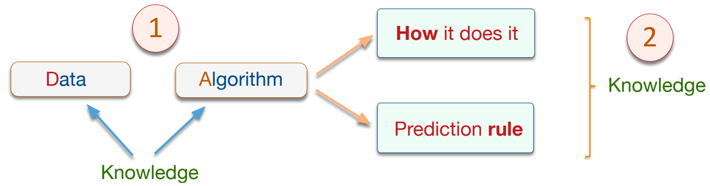
Little data + **lots** of prior knowledge

Big data + **less** prior knowledge



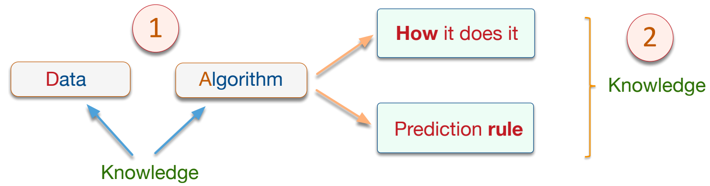

1
Data
Algorithm
Knowledge
How it does it
Prediction rule
2
Knowledge

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 learning algorithm
  - Constraints on the hypothesis space: **representation bias**

$$h^\star \;=\; \underset{h \in \mathcal{H}}{\mathrm{ArgMin}} \left[R_{\mathrm{Emp}}(h) \;+ \lambda \, \mathrm{reg}(h) \right]$$

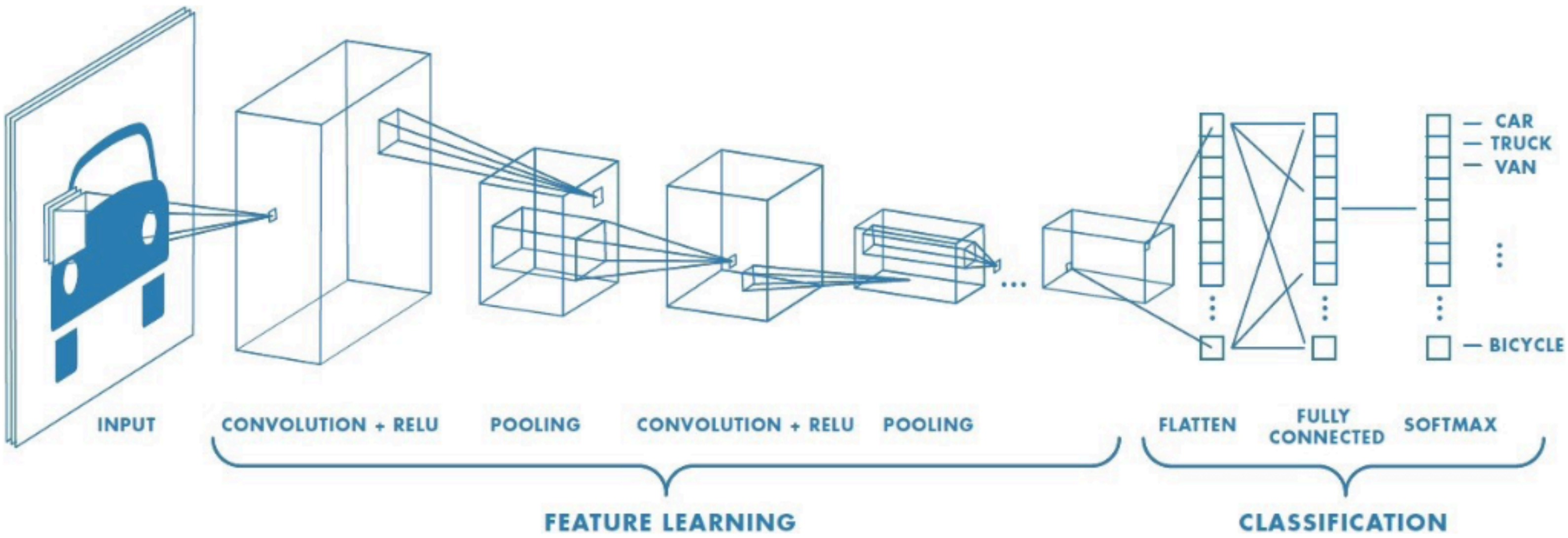
Looking for **sparse linear hypotheses**

$$h^\star \;=\; \underset{h \in \mathcal{H}}{\mathrm{ArgMin}} \left[\frac{1}{m} \sum_{i=1}^{m} \ell(h(\mathbf{x}_i), y_i) \;+ \lambda \, ||h||_1 \right]$$

Favors hypotheses with few non null coefficients

Knowledge as **input** to ML

- Convolutional Neural Networks
  - Knowledge embedded in the architecture of the network

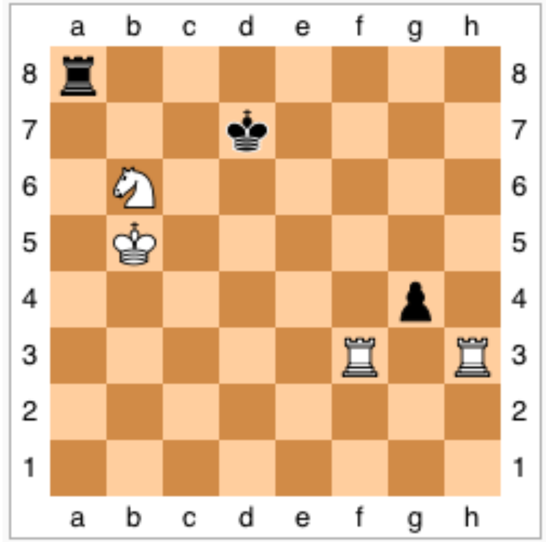


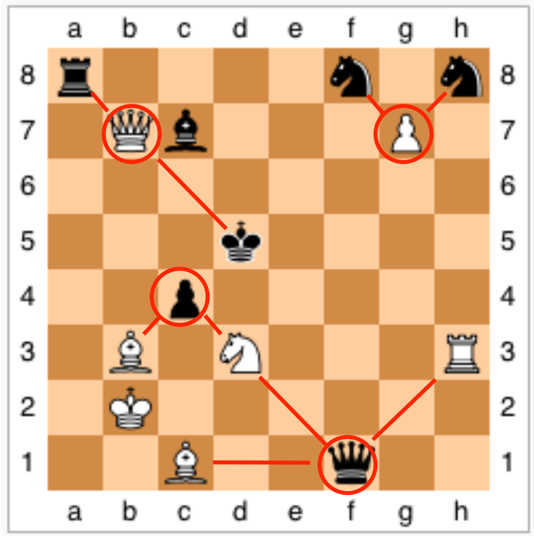
From https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Learning **from a single example**

Explanation-Based Learning

1. A **single example**
2. **Search for a proof** of a « fork »
3. **Generalization**





Explanation-Based Learning

Ex : **learn a concept** `stackable(Object1, Object2)`

- **Theory**:

```
(T1) : weight(X, W) :- volume(X, V), density(X, D), W is V*D.

(T2) : weight (X, 50) :- is-a(X, table).

(T3) : lighter(X, Y) :- weight (X, W1), weight(X, W2), W1 < W2.
```

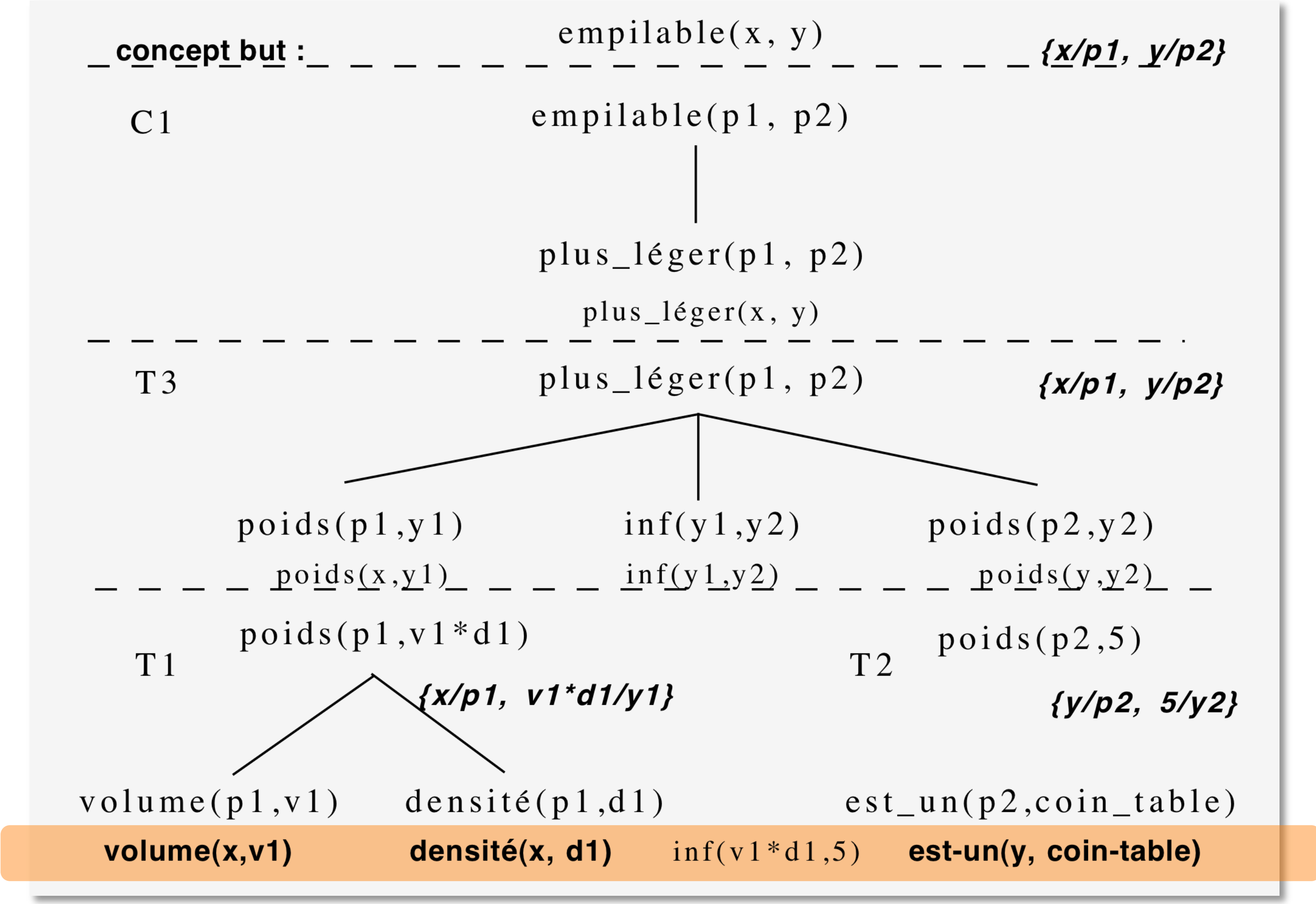
- **Operationality constraint**:
  - Concept to express with predicates *volume, density, color, …*

- **Positive example** (solution) :

```
on(obj1, obj2).
is_a(object1, box).
is_a(object2, table).
color(object1, red).
color(object2, blue).
matter(object2, wood).
```

```
volume(object1, 1).
volume(object2, 0.1).
owner(object1, frederic).
density(object1, 0.3).
material(object1, cardboard).
owner (object2, marc).
```

Explanation-Based Learning



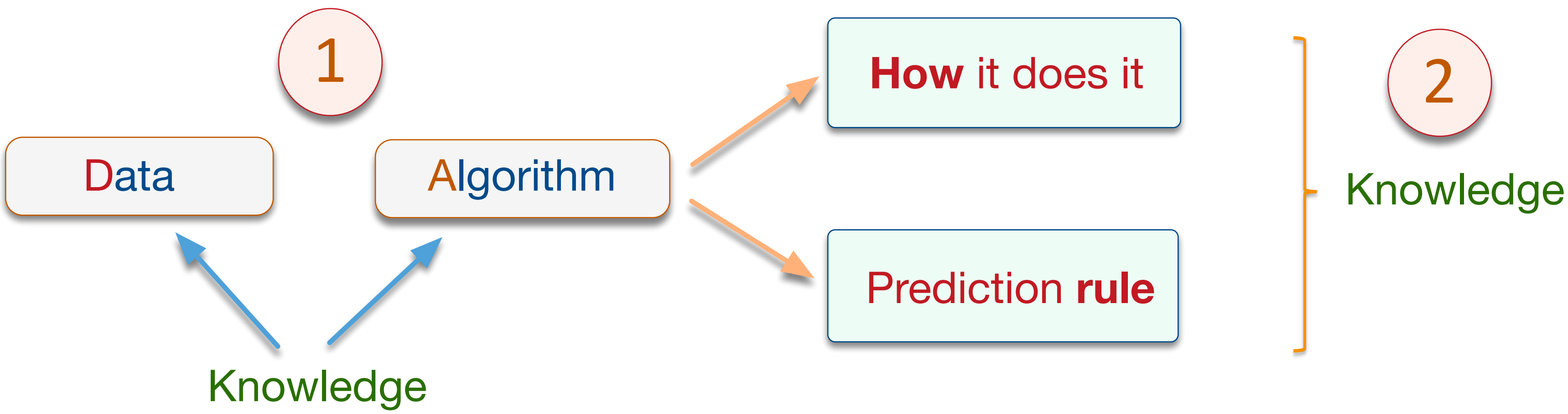
Arbre de preuve généralisé obtenu par régression du concept cible dans l'arbre de preuve en calculant à chaque étape les littéraux les plus généraux permettant cette étape.

Explanation-Based Learning

- Induction **from a single example**
  - … and a **strong domain theory**
- Language of **logics**
- **Operators** for reasoning (deduction, …)

Now used in « solvers » of SAT problems

because the “data” is clean


1
Data
Algorithm
Knowledge
How it does it
Prediction rule
2
Knowledge

What kind of knowledge can we **extract**?

- The hypothesis returned: a decision function

What kind of knowledge can we **extract**?

- The hypothesis returned: a decision function
  - **Recommending** a **movie**
  - **Recommending** a **life partner**
  - Written **character recognition**
  - Recognize **traffic signs**
  - Decide the **value of a position** in go
  - Predict the **risk of crime occurrence**
  - Decide if someone should **get a loan**
  - Decide to **hire or not** someone

What kind of knowledge can we **extract**?

- Just a decision function?

- Or included in a larger inference system?

E.g. in legal cases

→ Question **where** that decision **come from**

What kind of knowledge can we **extract**?

- **Interpretability** of the decision function
  - **Decision trees** seem readily interpretable
  - **Linear decision functions** are less so
  - **Random forests** are much less still
  - **SVM**
  - **Neural Networks**

Require a difficult analysis

What kind of knowledge can we **extract**?

- **Interpretability** of the learning process leading to a decision function
  - **Sensitivity analysis**
    - If this input value is changed, what happens
  - **Explanation-Based** Learning

What kind of knowledge can we **extract**?

- When interpretability is NOT needed?

What kind of knowledge can we **extract**?

- When interpretability is NOT needed?
  - When **low risk** associated with the decision
    - E.g. *recommendation for a movie*
  - When **good guarantees** on performance exist
    - E.g. *character recognition*

What kind of knowledge can we **extract**?

- When interpretability IS needed?

What kind of knowledge can we **extract**?

- When interpretability IS needed?

  1. With **high risk decisions**
     - *E.g. chirurgical operation*
     - *E.g. shutting down a nuclear plant*
     - *E.g. autonomous vehicle*

What kind of knowledge can we **extract**?

- When interpretability IS needed?

  1. With **high risk decisions**
     - *E.g. chirurgical operation*
     - *E.g. shutting down a nuclear plant*
     - *E.g. autonomous vehicle*

  2. Satisfying **curiosity** (what science is about)
     - *E.g. explain surprising results*
     - *E.g. when no easy explanation exists*
     - *E.g. when the decision function must be included in a larger inference system (a domain theory)*

What kind of knowledge can we **extract**?

- When interpretability IS needed?

What kind of knowledge can we **extract**?

- When interpretability IS needed?

  3. **Debugging**

     - *E.g. why is that decision wrong (counterfactual)*
     - *E.g. if a bicycle is recognized because it has two wheels, what if one is hidden behind side bags?*
     - *E.g. why the system seems gender biased?*

What kind of knowledge can we **extract**?

- When interpretability IS needed?

3. **Debugging**
   - *E.g. why is that decision wrong (counterfactual)*
   - *E.g. if a bicycle is recognized because it has two wheels, what if one is hidden behind side bags?*
   - *E.g. why the system seems gender biased?*

4. Interpretability **demands higher standard predictive systems**
   - *An interpretable system **can be manipulated***
     - *E.g. if someone knows that a loan is granted if you have more than 2 credit cards*

   - *In order **not to be manipulated**,*
     *the predictive system must use **causal factors***

What kind of knowledge can we **extract**?

- To **recognize cars**

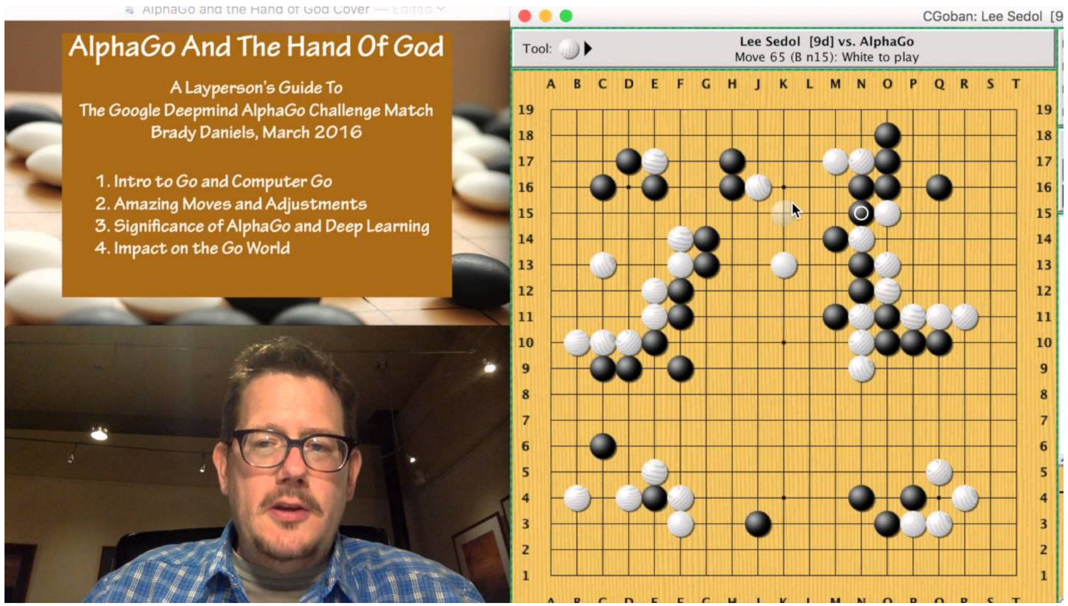
Is this less of a car
because the context is wrong?

What kind of knowledge can we **extract**?

- To decide the value of a position in go

  The "hand of God"

How to **revise** or **reconstruct** a theory of go?



- Why is Machine Learning currently **lacking**?
  - The **exclusive focus on predictive performance** leads to an incomplete learning problem formulation

- Why is Machine Learning currently **lacking**?
  - The **exclusive focus on predictive performance** leads to an incomplete learning problem formulation
  - We want also
    - **Interpretability** of the **results**
    - **Interpretability** of the **process**
    - Gaining a **better understanding of the world** when including the learned decision function in an existing theory

- Why is Machine Learning currently **lacking**?
  - The **exclusive focus on predictive performance** leads to an incomplete learning problem formulation
  - We want also
    - **Interpretability** of the **results**
    - **Interpretability** of the **process**
    - Gaining a **better understanding of the world** when including the learned decision function in an existing theory

Somehow, we have to change the inductive criterion used in Machine Learning