

## ESR 1- Explainable and strong artificial intelligence in High Energy Physics

Type	No. Legs	Stinger	No. Eyes	Compound Eyes	Wings
Spider	8	✗	8	✗	0
Beetle	6	✗	2	✓	2
Bee	6	✓	5	✓	4
Fly	6	✗	5	✓	2

Person: *“Why is image J labelled as a Spider instead of a Beetle?”*

ExplAgent: *“Because the arthropod in image J has eight legs, consistent with those in the category Spider, while those in Beetle have six legs.”*

Person: *“Why did you infer that the arthropod in image J had eight legs instead of six?”*

ExplAgent: *“I counted the eight legs that I found, as I have just highlighted on the image now.”* (ExplAgent shows the image with the eight legs counted).

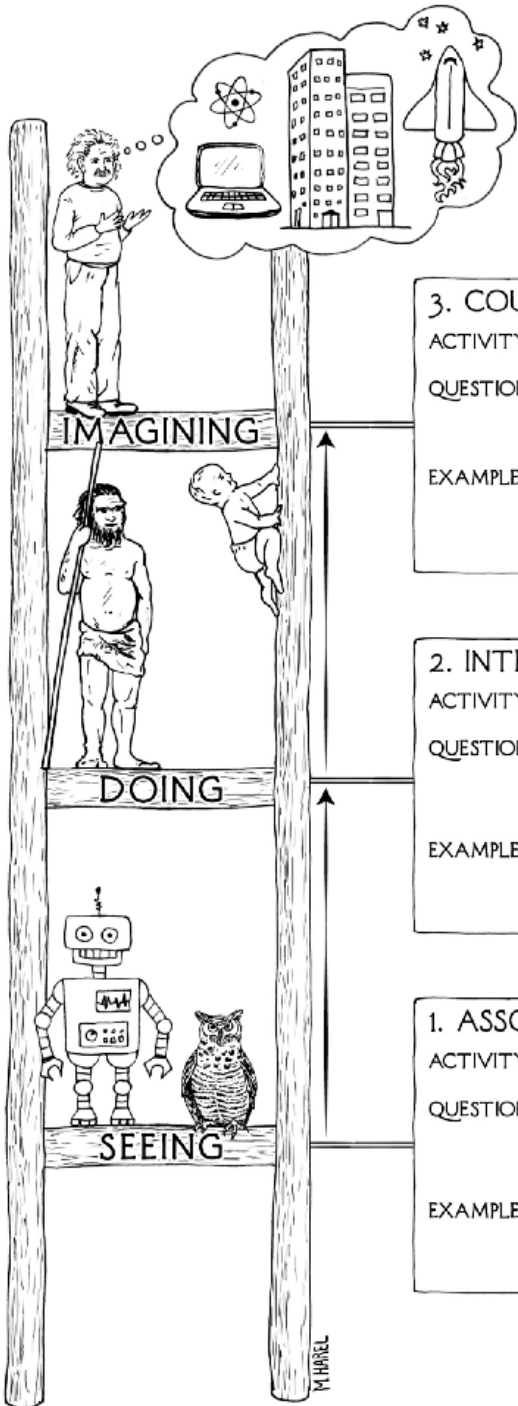
Person: *“How do you know that spiders have eight legs?”*

ExplAgent: *“Because in the training set I was trained on, almost all animals with eight legs were labelled as Spider.”*

Person: *“But an octopus can have eight legs too. Why did you not classify image J as an octopus?”*

ExplAgent: *“Because my function is only to classify arthropods.”*

# ESR 1- Explainable and strong artificial intelligence in High Energy Physics



### 3. COUNTERFACTUALS

ACTIVITY: Imagining, Retrospection, Understanding

QUESTIONS: *What if I had done ...? Why?*  
(Was it X that caused Y? What if X had not occurred? What if I had acted differently?)

EXAMPLES: Was it the aspirin that stopped my headache?  
Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?

### 2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do ...? How?*  
(What would Y be if I do X?  
How can I make Y happen?)

EXAMPLES: If I take aspirin, will my headache be cured?  
What if we ban cigarettes?

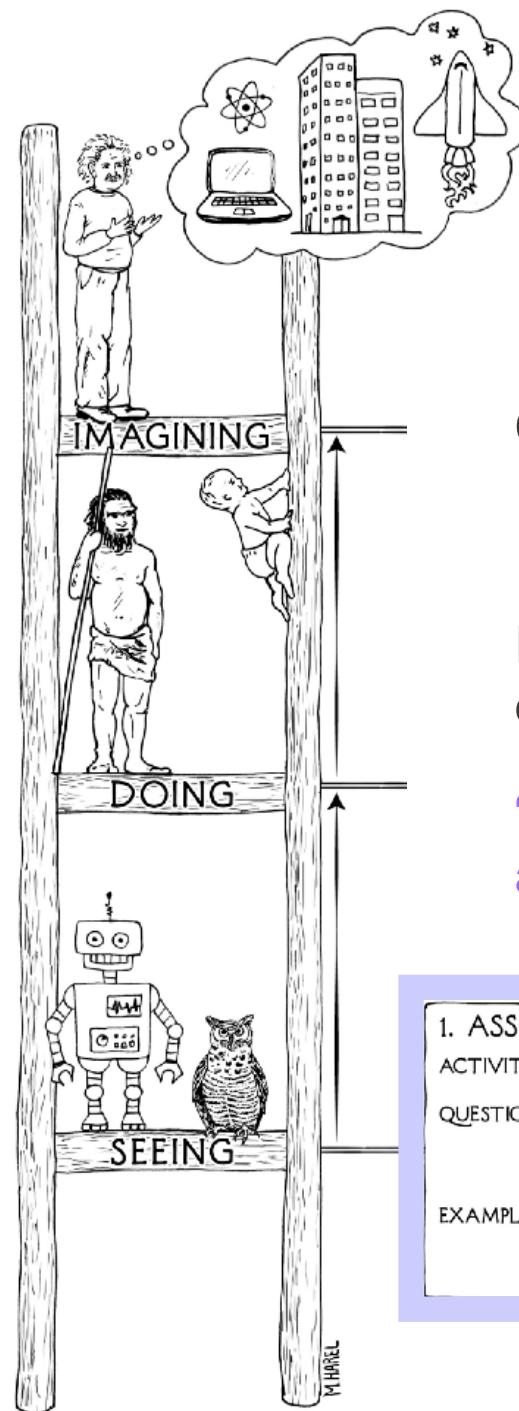
### 1. ASSOCIATION

ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see ...?*  
(How are the variables related?  
How would seeing X change my belief in Y?)

EXAMPLES: What does a symptom tell me about a disease?  
What does a survey tell us about the election results?

**Seeing;** we are looking for regularities in observations.



## “What if I see ...?”

Calls for predictions based on passive observations.

It is characterized by the question “What if I see ...?”

For instance, imagine a marketing director at a department store who asks,

“How likely is a customer who bought toothpaste to also buy dental floss?”

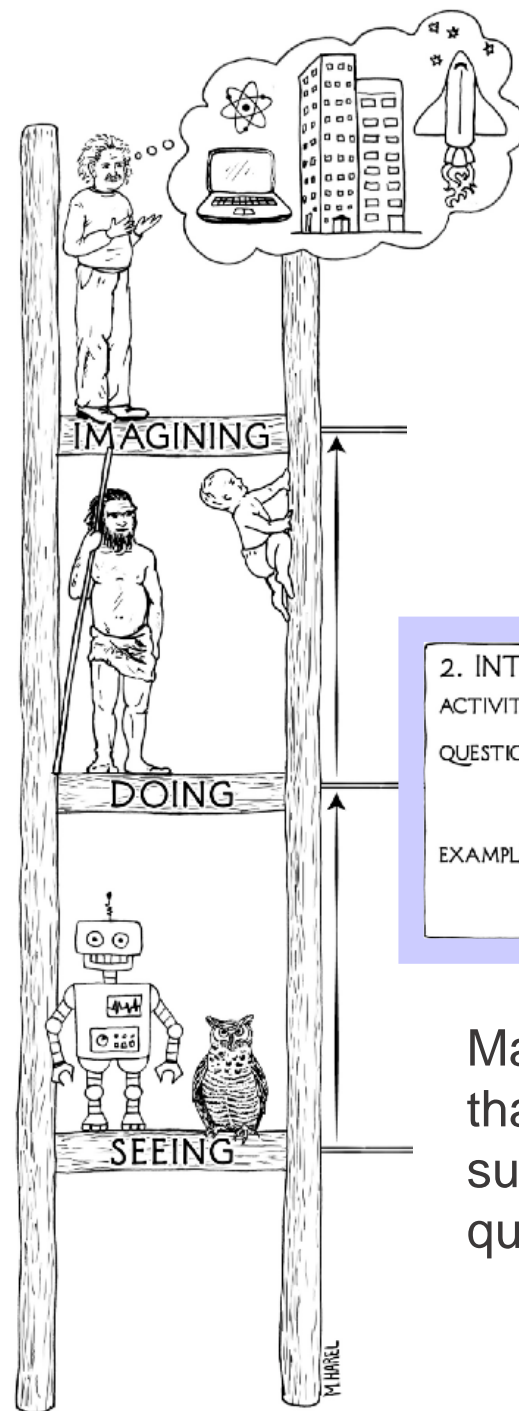
### 1. ASSOCIATION

ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see ...?*  
(How are the variables related?  
How would seeing X change my belief in Y?)

EXAMPLES: What does a symptom tell me about a disease?  
What does a survey tell us about the election results?

**Intervention;** ranks higher than association because it involves not just seeing but changing what is.



## “What if do ...?” & “How?”

We step up to the next level of causal queries when we begin to change the world. A typical question for this level is

“What will happen to our floss sales if we double the price of toothpaste?”

### 2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do ...? How?*  
(What would Y be if I do X?  
How can I make Y happen?)

EXAMPLES: If I take aspirin, will my headache be cured?  
What if we ban cigarettes?

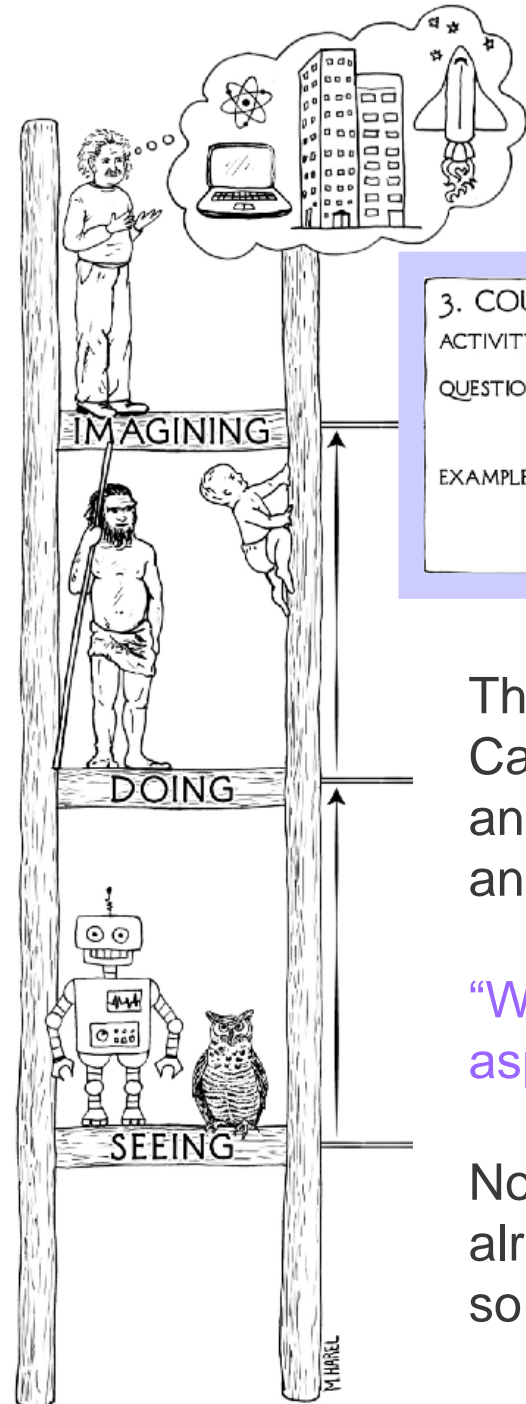
This already calls for a new kind of knowledge, absent from the data, which we find at rung two of the Ladder of Causation, **Intervention**.

Many scientists have been quite traumatized to learn that none of the methods they learned in statistics is sufficient even to articulate, let alone answer, a simple question like

“What happens if we double the price?”



**Counterfactuals;** ranks higher than intervention because it involves **imagining, retrospection and understanding.**



# “What if I had done ...?” & “Why?”

**3. COUNTERFACTUALS**  
ACTIVITY: Imagining, Retrospection, Understanding  
QUESTIONS: *What if I had done ...? Why?*  
(Was it X that caused Y? What if X had not occurred? What if I had acted differently?)  
EXAMPLES: Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?

- We might wonder, My headache is gone now, but
- Why?
  - Was it the aspirin I took?
  - The food I ate?
  - The good news I heard?

These queries take us to the top rung of the Ladder of Causation, the level of **Counterfactuals**, because to answer them we must **go back in time, change history,** and ask,

“What would have happened if I had not taken the aspirin?”

No experiment in the world can deny treatment to an already treated person and compare the two outcomes, so we must import a whole new kind of knowledge.

## **Some of the possible formulations for High Energy Physics (HEP):**

- **Anomaly detection**
- **Anomaly explanation**
- **Anomaly simulation (interventional reasoning)**
- **Anomaly suggestion (counterfactual reasoning)**
- **..... applications limited by my current ignorance about HEP,  
need and interest to collaborate and learn from HEP experts.**

## **Recommendation Systems (RS)**

- **Novelty detection**
- **Novelty explanation**
- **Novelty simulation (interventional reasoning)**
- **Novelty suggestion (counterfactual reasoning)**
- **..... applications limited by my current ignorance about RS,  
need and interest to collaborate and learn from RS experts.**