Notes on Knowledge Representation and Reasoning

Riccardo Lo Iacono

October 7, 2025

Contents.

- 1 - Introduction to Knowledge Representation and Reasoning.

Knowledge Representation and Reasoning (KRR) is a branch of Artificial Intelligence focused on understandig how knowledge, whose meaning will be soon clarified, can be represented symbolically and manipulated automatically by reasoning programs.

Formally, knowledge is defined as the relationship existing between a *knower* and a proposition; that is, the idea expressed by a simple declarative sentence (eg. "Mark will come to the party"). To represent and manipulate knowledge, which is our end goal, we have to understand what reasoning is. Let $p_1, p_2, \ldots, p_n$ be a sequence of premises; we define *reasoning* as the process which, from the premises $p_1, \ldots, p_n$ allows one to draw a conclusiion. Later on we discuss the logic formalization of reasoning.

Throughout this notes we use a "logic language", later discussed, known as First Order Logic (FOL). Similarly to any other language it has its own:

- *syntax*: the set of symbols and rules of the language;
- *semanic*: the meaning of syntactically correct sentences; and
- *pragmatics*: how contex and habits contribute to the meaning.

The next two sections cover all the type of reasoning. Precisely, we first discuss of all those aspects of the human reasoning that lead to an incorrect reasoning, we then focus on the three main types of correct reasoning.

- 1.1 - Sources of incorrect reasoning.

At the source of an incorrect (human) reasoning there are mainly three causes: *fallacies*, *biases* and *paradoxes*. Each of these groups a set of causes that we discuss briefly.

Remark. The lists we provide for each (fallacies, biases and paradoxes) is not exhaustive at all. More on these can be found easily online.

- 1.1.1 - Fallacies.

The concept of *fallacy* is straightforward. It is a argument that, despite sounding convincing, is invalid.

Example
Consider the sentece "A pencil isn't an animal. All animal feeds. Therefore a pencil doesn't feed". Though both premisies are true, and therefore the conclusion, the overall sentece is non-sensical.

We can distinguish several types of fallacies, some of which are listed below.

- Fallacy of composition (compositio): attributes to an entity as a whole some property valid for its parts.

  "The product is on sale at a price of 10,000 euros, payable in small installments of only 100 euros each. Therefore the total price is moderate."

- Fallacy of division (divisio): attributes to the individual parts a property that is valid for the whole entity

  "A truck is heavy. Therefore each of its component is heavy."

- Fallacy of relevance: derives a conclusion from premisies that are irrelevant.

  It has different forms. We list them below.

  - Argumentum ad hominem: discredits a thesis in order to deduce its falsity.
  - Tu quoque: the one proposing a thesis does the opposite of what he wants to prove.
  - Ad vericundum: the thesis is supposed to be true just because credit is given to those proposing it.
  - Ad populum: accepts a thesis simply because it's accepts by the majority.
  - Straw Man: creates a distorted version of a thesis, to refute it.
  - Ad ignorantiam: supports a thesis by refuting the arguments in favor of the opposite thesis.
  - Petitio principi: the thesis itself is hidden within the premisies.
  - False dichotomy: present two solution to the problem, as if these are the only ones.
  - Non causa pro causa: fix two peculiar characteristics and derive that the first causes the second one.

- 1.1.2 - biases.

We as humans are subject to biases, which are systematic pattern deviation from the norm or rationality in the mental process of judgement. As for fallacies, we can distinguish several forms of bias; and interestringly enough biases are at the base of some of the fallacies we discussed previously.

As done for fallacies, we list some of the most common types of biases, to give the reader an idea of what these are concretely.

- Confirmation bias: is a mental process that consists in selecting the information possessed in order to place greater attention, and therefore greater credibility, on those that confirm one's beliefs.
- Belief bias: is the tendency to judge the correctness of an argument based on the correctness of the conclusion.
- Availability bias (studied by Kahneman[1]and Tversky) is the tendency that leads us to make judgments on the basis of the most available examples and/or information, which most easily and vividly come to mind.

- 1.1.3 - paradoxes.

The concept of paradox is well known to everybody, though in the context of logic it's usually missleading; by this we refer to the fact that some of those that are generally considered paradoxes, are not. For instance, let us consider Epimenides's liar paradox:

"All Cretans always lie."

Since Epimenides was Cretan, he himself would be lying. Thus, the phrase "All Cretans always lie" cannot be true, cause it would contradict itself. Therefore, it's simply false.

More often then not paradoxes derive from the fact that semantic meaning is attributed to a sentence that refers to itself. This is what we call self-referentiality. This problem with self-referentiality was later understood by *Russell*.

- 1.2 - Correct forms of reasoning.

One can reach a correct reasoning through three approches: *deduction*, *abduction* and *induction*. We discuss each of these in the next few sections, but we focus maily on deduction in the rest of these notes.

We note that, in the case of formal reasoning, frequent mistakes usually concernes conditional statements, i.e., sentences of the form "If A, then B" or those that can somehow be transformed into one of this type. We call A and B, respectively, *antecedent* and *consequent*.

"If it rains, we always take the umbrella. Yesterday we took the umbrella, so yesterday it rained."

[1]For their result, Kahneman was praised with the nobel (Tversky was dead at that time). On the topic of biases, we suggest reading "Thinking, Fast and Slow" by Kahneman.

- 1.2.1 - Deduction.

Let us consider again the sentece:

> "If it rains, we always take the umbrella. Yesterday we took the umbrella, so yesterday it rained."

Is it necessary true? What if we took the umbrella to rapair it?

This kind of reasoning is called deduction; that is, deduction is the process that from a "rule" (the conditional) derives a conclusion.

The above is an example of what we call *affirming the consequent*; that is, we deduce that A is true knowing that B is true, when there exist a conditional statements involving A and B. As its counterpart we have the *denying the antecedent* fallacy, in which we deduce the falsity of B by knowing with certainty that A is not true.

Given a conditional sentece, the two correct ways to apply deduction are called

- modus ponens: we know that A is true and deduce that B is also true.
- mudus tollens: we know that B is false and deduce that A is also false.

- 1.2.2 - Abduction and induction.

For what we have said about deduction, is easy to understand that it with logical inferences that are certain. Abduction and induction on the other hand cover probable inferences. That is, deduction differs from both, abduction and induction, by the degree of trust we put in the inference. To see such differece, see how the sentence

> "If it rains, we always take the umbrella. Yesterday we took the umbrella, so yesterday it rained."

becomes the sentence

> "If it rains, we always take the umbrella. Yesterday we took the umbrella, so yesterday it probably rained."

in the abduction case, and

> "If it rains, we always take the umbrella. Tomorrow it will rain, so probably tomorrow we will take the umbrella."

in the induction one.

Acronyms.

FOL First Order Logic. 1

KRR Knowledge Representation and Reasoning. 1