Fausto Giunchiglia

The Logic of Machine Intelligence

From theory to practice

September 24, 2023

Springer Nature

Contents

Part I
World and representation

Chapter 1
The world and the mind

We often confuse the world with our mental representation of the world itself. Is this a correct assumption?

1.1 Frames of mind

Following social theory, a *frame* is a schema of interpretation, a collection of anecdotes and stereotypes, that individuals rely on to understand and respond to events [5]. People build a series of mental representations of the world through biological and cultural influences. They then use these filters to make sense of the world. The choices people make are influenced by frames. Participation in a language community necessarily influences an individual's perception of the meanings attributed to words or phrases.

Example 1.1 (The car accident) In [8] *two experiments are reported in which subjects viewed films of automobile accidents and then answered questions about events occurring in the films. The question, "About how fast were the cars going when they smashed into each other?" elicited higher estimates of speed than questions which used the verbs collided, bumped, contucted, or hit in place of smashed. On a retest one week later, those subjects who received the verb smashed were more likely to say "yes" to the question, "Did you see any broken glass?", even though broken glass was not present in the film. These results are consistent with the view that the questions asked subsequent to an event can cause a reconstruction in one's memory of that event*" (Quote from the abstract of [8]).

Example 1.2 (The Asian disease problem) Tversky and Kahneman [6] demonstrated systematicity when the same problem is presented in different ways, for example in the *Asian disease problem*. Participants were asked to "imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume the exact scientific estimate of the consequences of the programs are as follows." The first group of participants was presented with the following choice. In a group of 600 people,

- Program A: "200 people will be saved";
- Program B: "there is a 1/3 probability that 600 people will be saved, and a 2/3 probability that no people will be saved"

72% of the participants preferred program A, 28%, opted for program B. The second group of participants was presented with a different choice. In a group of 600 people,

- Program C: "400 people will die";
- Program D: "there is a 1/3 probability that nobody will die, and a 2/3 probability that 600 people will die"

In this decision frame, 78% preferred program D, with the remaining 2% opting for program C. Programs A and C are identical, as are programs B and D. The change in the decision frame between the two groups of participants produced a preference reversal: when the programs were presented in terms of lives saved, the participants preferred the secure program, A (= C). When the programs were presented in terms of expected deaths, participants chose the gamble D (= B).[4].

1.2 Optical Illusions

An optical illusion, or visual illusion, occurs when the visual system creates a perception that seems different from the surrounding reality. The main categories of illusions are physical, physiological and cognitive, each with types such as ambiguities, distortions, paradoxes and fictions. Examples include the apparent curvature of a stick in water (physical distortion), the effect of adapting to movement (physiological paradox), and the residual impression of an image (physiological fiction). Pathological visual illusions result from pathological changes in physiological mechanisms and can lead to visual hallucinations. These illusions can be used in the monitoring and rehabilitation of psychological disorders such as phantom limb syndrome and schizophrenia.

Example 1.3 (Herman Grid) A demonstration of how our perception can deceive us is *Herman Grid* [1], which is an optical illusion in which a grid of white dots on a black background appears to create dark spots at the points of intersection. Although we know that there are no spots actually present, our perception tricks us into believing that they are present. This example demonstrates how our visual perception (and so our senses) can deviate from objective reality.

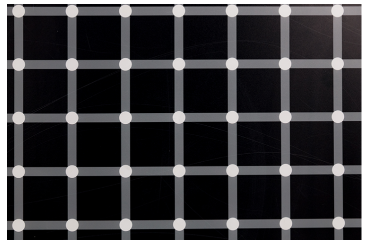


Fig. 1.1 Herman Grid

Example 1.4 (Kanizsa triangle) The illusion consists in the fact that looking at the image we hallucinate to see two triangles and 3 circles, but actually none of them is there.

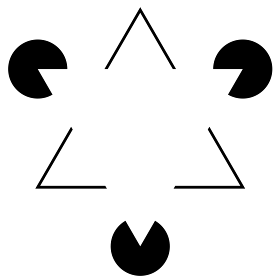


Fig. 1.2 Kanizsa Triangle

Example 1.5 (Pareidolia) Pareidolia is the tendency for perception to impose a meaningful interpretation on a nebulous stimulus, usually visual, so that one sees an object, pattern, or meaning where there is none. For example, we tend to see faces everywhere, even in the surface of the Moon.

Fig. 1.3 Pareidolia

Example 1.6 (My Wife and My Mother-in-Law) This is a famous optical illusion in which viewers can see either a young woman looking away or an old woman in profile, depending on how they interpret the drawing's lines. The illusion plays on our ability to switch between different perspectives.

Fig. 1.4 My Wife and My Mother-in-Law

Example 1.7 (Impossible Trident) Also known as the "blivet", this illusion depicts a three-pronged trident that mysteriously transforms into two cylindrical shafts at the other end. This illusion plays with our perception of three-dimensional space.

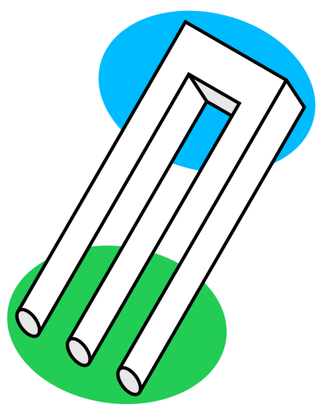


Fig. 1.5 Impossible Trident

Example 1.8 (Rubin's Vase) This is a classic example of figure-ground perception. Viewers can either see a vase in the center or two faces in profile facing each other.

The brain can switch between either interpretation but cannot see both at the same time.

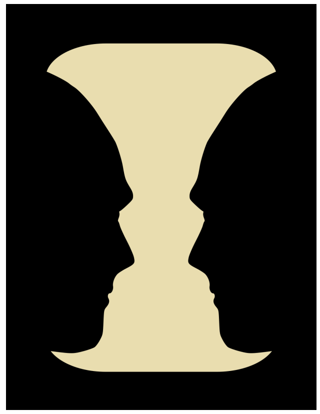


Fig. 1.6 Rubin's Vase

Example 1.9 (Penrose Triangle) This is an "impossible object" that cannot exist in three-dimensional space. It appears to be a solid object made of three straight beams of square cross-section, but its construction is impossible.

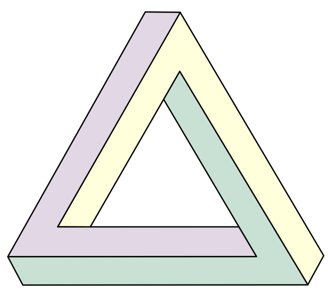


Fig. 1.7 Penrose Triangle

1.3 Mind Fallacies

A fallacy is reasoning that is logically invalid, or that undermines the logical validity of an argument. All forms of human communication can contain fallacies. The use of fallacies is common when the speaker's goal of achieving common agreement is more important to them than utilizing sound reasoning.

Fallacies can be classified depending on their structure (formal fallacies) or on their content (informal fallacies). A formal fallacy, also called a deductive fallacy or logical fallacy [2], represents a type of reasoning that loses validity due to a flaw in its logical structure. This flaw can be clearly represented within a standard logical system, such as propositional logic. In other words, it is a deductive argument that is invalid. Even though the premises of the argument might be true, the conclusion drawn from it is still false. Informal fallacies, the larger group, may then be subdivided into categories such as improper presumption, faulty generalization, and error in assigning causation and relevance, among others.

We provide below various examples of informal and formal fallacies.

Example 1.10 (Cognitive Bias) Cognitive biases [12] are an example of informal fallacies. They represent systematic patterns of deviation from the norm and rationality in the evaluation process. The Asian disease example, see above, is an instance of cognitive bias.

Example 1.11 (Misconceptions) Misconceptions are informal fallacies. A common misconception is a perspective or data that is often considered to be true but is actually false. Usually, such misunderstandings stem from entrenched traditions (such as gossipy tales), stereotypes, superstitions, fallacies, misinterpretations of science, or the spread of pseudoscience. Some of these misunderstandings are considered urban legends and often contribute to moral alarmism.

Example 1.12 (Cognitive Distortion) Cognitive distortions are an informal fallacy. They can be traced to "thinking fallacies," representing irrational or distorted ways through which we process information and perceive reality. Some of the main thinking fallacies involved include

- overgeneralization, which draws overly broad conclusions from a single negative event;
- mental filtering, which focuses attention only on the negative aspects of a situation;
- over-labeling, which assigns negative labels to oneself or others based on mistakes or failures;
- dichotomous thinking, which considers only extremes without acknowledging nuance;
- emotional reasoning makes one believe that one's feelings reflect objective reality;
- personalization leads one to interpret events as being directly related to oneself;
- Negative prediction involves predicting the worst without concrete evidence;
- Catastrophism makes one imagine the worst as the only possibility, ignoring alternatives, while sample selection draws general conclusions from a limited set of data or experiences.

Example 1.13 (Paradoxes) Paradoxes are examples of formal fallacies. Paradoxes are situations or statements that seem contradictory or contraintuitive, often challenging our normal thinking and expectations. They are intellectual puzzles that can cause confusion and amazement as they violate our common understanding of logic or the laws of reality.

Some paradoxes emerge from fallacious reasoning, where it appears that rules of thought are correctly applied, but the end result is nonetheless contradictory or nonsensical. These paradoxes teach us the importance of carefully examining the premises and inferences behind an argument.

Conversely, there are paradoxes that emerge from complex situations or situations that fall into categories of mathematical or philosophical problems. These paradoxes can challenge our intuition and reveal the limitations of our knowledge. In some cases, these paradoxes can highlight deep issues in the very structure of our rational thinking.

A particular type of paradoxes, called antinomies, is characterized by the presence of self-contradictions in situations where we would expect consistency. These paradoxes can be used to highlight the inherent challenges in dealing with concepts such as truth, description or infinity.

Example 1.14 (The Map - Territory confusion) The map-territory relationship [7] is a fundamental concept for understanding the fallacies of the human mind. Essentially, it points out that the mental representations we create, such as concept maps, models and interpretations, are not identical to the reality they seek to represent. This concept detects several distortions in the perception and interpretation of human reality. For example, people often generalize and make incorrect conclusions based on limited experiences, confirming their own biases and ignoring conflicting information. Cultural beliefs influence mental maps, leading to distorted perceptions. Cognitive distortions and overconfidence in representations can lead away from objective reality. In summary, understanding the map-territory relationship prompts us to be aware of discrepancies between our mental representations and actual reality, helping us to avoid wrong thinking traps and maintain a critical perspective.

1.4 So What?

In this chapter, we explored the inherent flaws in human thinking, such as fallacies, biases and misconceptions, which can affect our understanding of reality and decision making. However, we can adopt several strategies to overcome these challenges and promote more accurate, rational and logic-based thinking.

Logic is a crucial tool for avoiding fallacious reasoning. Formalizing thinking through logic provides us with a structured framework for evaluating arguments and drawing conclusions. The systematic approach of logic helps us recognize and foil fallacious reasoning. Learning to identify the premises, inferences and conclusions in an argument enables us to detect logical errors or inconsistencies. This is key in Computer Science and even more in Artificial Intelligence.

Chapter 2
Representations

The various types of fallacies described in Section 1.3 raise the issue of whether it is possible to deal with them. But why? To deal in which sense? Modulo some extreme cases, humans and humanity have been able to develop well and grow in time despite the pervasiveness of fallacies in human interactions with the world and with others. Two are the main reasons underlying this work. The first is that, because of the Web and social platforms, now humans are able to interact with people that are hardly known and with very different cultures and, furthermore, they get in contact with parts of the world that they never visited physically. The probability of misalignments and misunderstandings among people has grown immensely. The second is that, in this era where we want to build CS and AI systems which are more and more complex, more and more intelligent, and which pervasively interact with people in their everyday lives, we need to have systems which are robust, trustable, and whose behaviour we fully understand and also control.

The first step is to find a way to build representations of the world which are not ambiguous and which can be used as the basis for solving the interpretations problems highlighted in Section 1.3. This is the gaal of this section.

2.1 The Semantic Gap

Living organisms perceive reality, what we call the world, through the lenses of their perceptions organs. This process is not neutral. Different species and even different humans perceive the world differently. We talk of Semantic Gap relating to the impossibility for humans and machines to perceive the world as it really is, or even in the same way. The Semantic Gap is the source of the pervasive misalignment of the mental models of the world that humans, and also machines, build.

Intuition 2.1 (World) The **world** is what we perceive through the five senses and assume it exists. It is the spatio-temporal dimension in which humans live and interact with other humans and everything else around them.

Fig. 2.1 The 5 Senses: sight, hearing, touch, taste and smell.

Intuition 2.2 (Memory) When we perceive the world we create in our mind a **memory** of what we have perceived, the memory being itself a part of the world.

Intuition 2.3 (Mental Representations) Mental representations are a part of a person's memory. Mental representations are such that there is a correspondence between their contents and what is the case in the world they describe.

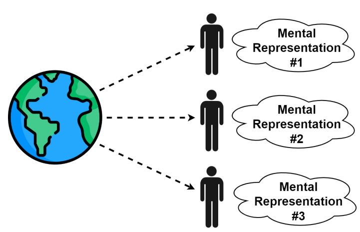


Fig. 2.2 Mental Representations.

Observation 2.1 (Mental representations) All humans have their own mental representations of the world. They are a fundamental mechanism enabling human knowledge, reasoning, action and communication.

Intuition 2.4 (Semantic gap) The **semantic gap** is the difference between the world and a human's mental representation of the world itself, what (s)he has perceived.

Observation 2.2 (Semantic gap) Most of the details of how perception and memory operate and how the different processes compose to generate memories is largely unknown. However we know that our memories are an encoding of what we perceive and that this encoding is partial and not faithfully representing what caused it.

2.2 Mental Representations

We have two types of mental representations.

Intuition 2.5 (Analogical mental representations) Analogical mental representations are mental representations that **depict** the world as we perceive it through the five senses.

Example 2.1 (Analogical mental Representations) We see an apple, we smell its fragrance, we taste it when eating.

Observation 2.3 (Analogical mental representations) Analogical mental representations enable us to acquire information about the world, directly from the world. They are used to act in the world, to learn from what has been previously perceived and to build an understanding of the world itself.

We describe analogical mental representations using languages. We use languages to build mental linguistic representations about the world, as represented in mental analogical representations.

Intuition 2.6 (Language) A **language** is any notation, generated by humans, agreed upon by humans, which allows to describe analogical representations, to reason about them, and to communicate about them to other humans.

Intuition 2.7 (Linguistic mental representations) Linguistic mental representations are mental representations that **describe** mental analogical representations using language.

Example 2.2 (Linguistic mental representations, language) The most important example of languages used in linguistic mental representations are the natural languages, e.g., Italian, and English, as memorized in our mind. Examples of linguistic mental representations are a poem and, in general, any piece of text describing the world that we remember.

Observation 2.4 (Linguistic mental representations) Linguistic mental representations are used to describe what is happening in analogical mental representations. They allow to communicate to other humans about our mental representations (and, thus, indirectly about the world), to learn from what has been previously described or perceived, and to reason in order to to derive unknown facts from what we already know.

Intuition 2.8 (Represent, depict, describe) To **represent** the world means anyone of two things: to depict it or to describe it.

Observation 2.5 (Analogical and linguistic mental representations) Analogical representations depict the world. Linguistic describe (the analogical representations of the world). They both represent the world.

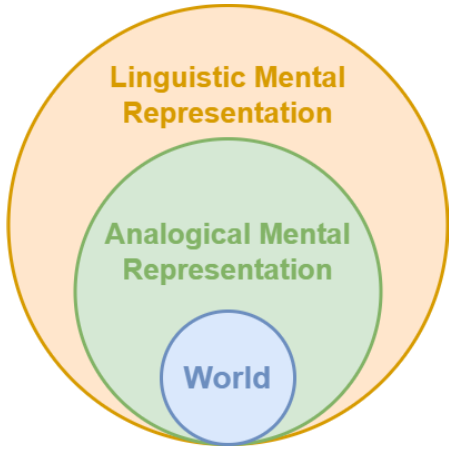


Fig. 2.3 Diagram of Mental Representations

Observation 2.6 (Partiality of mental representations) Because of the semantic gap, mental representations never describe the world completely. This has consequences. First, there are indefinitely many analogical mental representations that describe the same real world situation. Similarly, there is an indefinite number of linguistic mental representations for the same analogical representation.

Observation 2.7 (Number of mental representations) Because of partiality there are indefinitely many analogical mental representations that describe the same real world situation. Furthermore, there is an indefinite number of linguistic mental representations for the same analogical representation.

Observation 2.8 (Diversity of mental representations) Because of partiality, any two mental representations are necessarily different, depending on the spacetime coordinates under which they are generated, and the purpose of the person who generates them

Example 2.3 (Diversity of mental representations) Two people describing the same trip would do it so differently. For example, one person might have a partial and rough mental representation of the city, based mainly on a few famous tourist sites and the positive experiences he had during the trip. While the other person might have a different mental representation focused on other aspects of the city. She might remember the difficulties she encountered in finding the right way, some less pleasant experiences with locals or bad weather during the trip. Her mental representation might be more influenced by these less positive aspects.

Intuition 2.9 (Consistency and inconsistency of mental representations) We say that any two mental representations are **inconsistent** when it is impossible for those two mental presentations to represent the (same part of the) world, as he know it. **Consistency** means absence of inconsistency. Two consistent mental representations are still diverse but they are compatible in the sense that there is a (analogical representation of the) world which is described by both.

Example 2.4 (Inconsistency of mental representations) It is impossible to have two different objects (e.g., two cats) exactly in the same place in the same moment or the same object (a cat) in two different places in the same moment. Similarly, given an object, certain properties (e.g., being of color blue) prevent other properties from holding (e.g., being red), again in the same moment.

Observation 2.9 (Subjectivity of mental representations) Given the world they perceive, humans build one or more among the many possible mental analogical and linguistic representations of what they have perceived. Each individual has a unique and personal perspective on the world, influenced by different experiences, knowledge and viewpoints.

Observation 2.10 (Subjectivity vs. objectivity of mental representations) Humans may confuse the real world with their mental representations. A consequence is the assumption that (their mental representation of) the world is the same for everybody. Would this be the case would all be living in the same (mental representation of the) world. Because of subjectivity, this assumption turns out to be wrong.

Observation 2.11 (Subjectivity, inconsistency and objectivity) Two subjective mental representations may be (mutually) inconsistent. The presence of inconsistency provides evidence of the subjectivity of the mental representations involved.

2.3 Representations

The subjectivity and heterogeneity of mental representations raises some important questions. Is it possible to guarantee that the mental representations of different people are the same? Or, at least, that they are not mutually inconsistent and also similar enough in some key features, in particular those which are relevant to the problem to be solved? How do we enforce or at least facilitate the construction of similar mental representations

Intuition 2.10 (Representations) A **representation** is a part of the world, developed by the mind of a human, that represents that human's mental representation, and is made accessible, via one of the five senses, to other humans.

As for mental representations, we have two types of representations.

Intuition 2.11 (Analogical Representations) Analogical representations depict analogical mental representations.

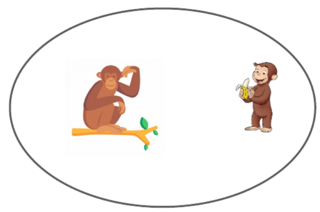
Intuition 2.12 (Linguistic Representations) Linguistic representations describe linguistic mental representations.

Example 2.5 (Representations) The following are examples of representations:

1. Any written natural language text is a linguistic representation, which can be generated on multiple media, for instance, paper, media, projection on a screen;
2. Any spoken natural language stream is a linguistic representation, which can be registered on transcribed on paper;
3. All forms of art, e.g., drawings, statues, paintings, music, monuments, are analogical representations.

Example 2.6 (Linguistic and analogical mental representations)

- There is a tree
- There is a banana
- The monkey is eating a banana
- The monkey is sitting on a tree
- The monkey is scratching his head



Observation 2.12 (Partiality, number, diversity, (in)consistency, subjectivity and objectivity) Observations 2.6 on the partiality, 2.7 on the number, 2.8 on the diversity and 2.9 on the inconsistency of mental representations apply also to representations. Not being in the mind of people, representations cannot be said to be subjective or objective. The question is about the mental representations they generate.

Observation 2.13 (From mental representations to representations to mental representations) The process and consequences of generating representations is well represented in the analogical representation in Figure 2.4. That is: the representation is generated by a single person starting from his/her mental representation and in turn it generates new mental representations in the minds of the people looking at it.

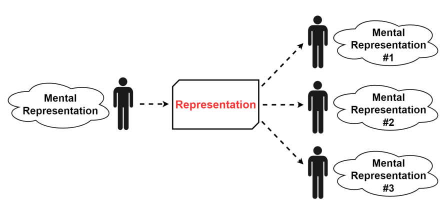


Fig. 2.4 From Mental Representations to Representations and back.

Observation 2.14 (From mental representations to representations) Representations, by their own nature and purpose, are such that there is a correspondence between their contents and those of the mental representations they describe. This is why people generate them.

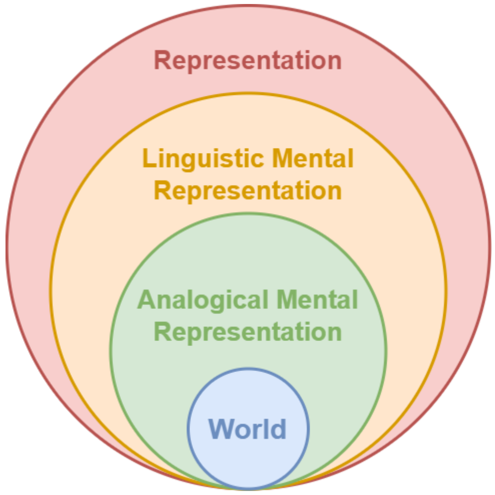


Fig. 2.5 Diagram of Representations

Observation 2.15 (From representations to mental representations) There is no guarantee that a representation generates similar subjective mental representations. Think for instance of the many different interpretations, impressions, feelings that a piece of art generates.

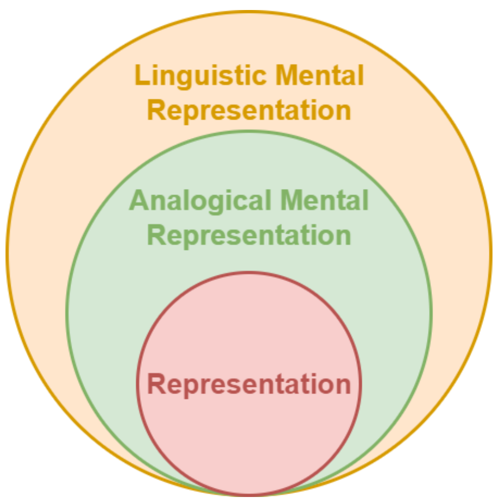


Fig. 2.6 Diagram of Representations

2.4 Exercises

Exercise 2.1 (Linguistic and analogical mental representations) Create a linguistic representation for the analogical mental representation in Figure 2.1.

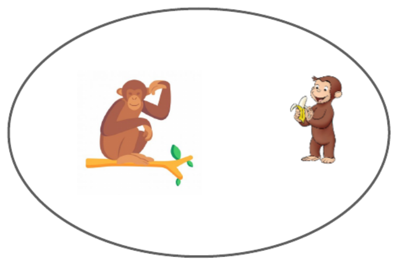


Fig. 2.7 Monkey and banana.

Exercise 2.2 (Linguistic and analogical mental representations) Create an analogical representation for this linguistic mental representation. The phrases are written in Tswana, a language spoken in southern Africa.

- Mongwe le mongwe o tshela mo
- lafatsheng la gagawe go ya ka kitso
- ya gagwe le go ya ka seo a se lemogang.
- Maitemogelo a tlhola seo, puo e letlelela
- go arologana tlhaloso
- le batho ba bangwe, ka moo

Chapter 3
Models and assertional theories

Observation 2.15 may suggest that there is no solution to the problem of subjectivity of mental representations. However this is not the case. The key observation is that representations are built by humans with the specific purpose of making mental representations of the same representation converge as much as possible, minimizing in particular the probability of inconsistencies. The question to be answered is how to build such representations.

3.1 Models

The starting point is analogical mental representations, as our representations start from here. Consider the following example.

Example 3.1 (What is in an analogical representation) Consider the analogical representation depicted in the image in Figure 3.1. We can see three people, that we can assume have names Paolo, Stefania and Sofia, that they are friends, various dogs, the fact that they are one at the right of the other, and of course much more.

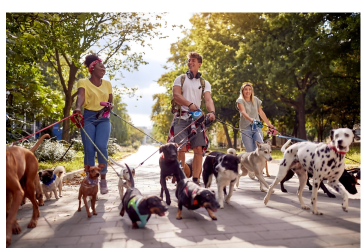


Fig. 3.1 An analogical representation of an everyday situation.

Observation 3.1 (Analogical representations as sets of facts) Any analogical representation, for instance that in Figure 3.1, always depicts various objects, e.g., Sofia, which belong to certain classes, e.g., "Sofia is a person", with certain properties articulated at various levels of complexity, e.g., "Sofia has blond hair", which are doing things, e.g., "Sofia is walking", and are engaged in certain relations with other objects, e.g. "Sofia is a friend Paolo and she is now interacting with her dogs". Despite their heterogeneity, all the statements above share the fact that they describe a certain state of affairs in the world. We call these statements facts. Any analogical representation can be thought of as a set of facts. We call analogical representations described as sets of facts, models

Intuition 3.1 (Fact) A **fact** `f` is something happening at certain spacetime coordinates.

Definition 3.1 (Model) A **model** `M` is a set of facts $\mathtt{F} = \{\mathtt{f}\}$

$$\mathtt{M} = \{\mathtt{f}\} \tag{3.1}$$

Observation 3.2 (Facts and models) Facts are the atomic, not further decomposable, elements of a model. Note that, contrary to models, facts are taken as a primitive notion and therefore cannot be formally defined

Example 3.2 (The facts of a model represented in Figure 3.1) A model, one among many others, of the situation represented in Figure 3.1 could for instance contain the following facts :

```
Sofia is a person    Paolo is a man
Rocky is a dog       Sofia is near Paolo
Sofia has blond hair Sofia is a friend of Paolo
Rocky is an animal   Rocky is the dog of Sofia
...
```

Observation 3.3 (The subjectivity of facts) Facts are what is observed and is also described, e.g., to third parties. The problem is that, just because of what discussed in Section 2.2 and, specifically what facts are is subjective and hidden in the minds of people who perceive them. How many more and/or different facts from those listed in Example 3.2 could you think of? Indefinitely many! Notice that any fact can be decomposed in any set of simpler facts if this is the current focus of the observer. So, for instance, instead of focusing on Sofia I could focus on her hair, or legs or . . .

Observation 3.4 (Mutually (in)consistent facts in a model) The model of Example 3.2 could be extended to assert the fact that Sofia is a woman or that Paolo is a person. We would however have problems extending it by adding the fact that Paolo is a woman, or that Sofia is a dog, as we would have two mutually inconsistent facts, something that we know cannot happen in the world as we perceive it. See also Observation 2.9. A model cannot contain facts which, at least intuitively. are mutually inconsistent. Beyond this simple example, the issue is how to formalize this intuition and then how to be able to detect it by reasoning about models.

Observation 3.5 (Facts and assertions) A fact, to be a fact, must be linguistically described as such. It is not by chance that in Example 3.2 we pointed to facts via a set of natural language descriptions. We call such descriptions, assertions. The simplest way to think of an assertion is as a declarative natural language sentence articulated in terms of a *subject* being in some more or less complex *relation* with an *object* (as in, e.g., "Stefania is walking with the dogs towards the city center"), or of a *subject* holding a certain more or less complex property (as in, e.g., "Stefania has blond long hair").

3.2 Assertional theories

Observation 3.6 (Assertions and assertional theories) Assertions are indivisible, we say atomic, descriptions of fact. Assertional theories are descriptions of models

Intuition 3.2 (Assertion) An **assertion** a is an atomic linguistic representation of some fact $\mathtt{f}$.

Definition 3.2 (Assertional theory) An **assertional theory** $\mathcal{T}_\mathcal{A}$ is a set of assertions $\mathcal{T}_\mathcal{A} = \{a\}$

$$\mathcal{T}_\mathcal{A} = \{a\} \tag{3.2}$$

We need to state that an assertion is the description of a specific fact and, more in general, that an assertional theory describes a model.

Example 3.3 (An assertional theory of the model represented in Figure 3.1) An assertional theory, one among many others, describing the facts from Example 3.2 in natural language could be, for instance:

Sofia è una persona	Paolo è un uomo	Rocky è un cane
Sofia è vicina a Paolo	Rocky è il cane di Sofia	Sofia è un'amica di Paolo
Rocky è un animale	Sofia ha i capelli biondi . . .	

As can be seen, the assertional theory is in Italian, since being in natural language it can also be expressed in this way, and it could equally be expressed in English.

3.3 Interpretation functions

Definition 3.3 (Interpretation function) Let $\mathcal{I}_A$ be an **interpretation function** of an assertional theory, defined as

$$\mathcal{I}_A : \mathcal{T}_\mathcal{A} \rightarrow \mathtt{M}. \tag{3.3}$$

We say that a fact $\mathtt{f} \in \mathtt{M}$ is the **interpretation** of $a \in \mathcal{I}_A$, and write

$$\mathtt{f} = \mathcal{I}_A(a) = a^{\mathcal{I}_A} \tag{3.4}$$

to mean that a is a linguistic description of $\mathtt{f}$. We say that $\mathtt{f}$ is the **interpretation of** a, or, equivalently, that a **denotes** $\mathtt{f}$.

Observation 3.7 (Interpretation function, polysemy) $\mathcal{I}_A$ is assumed to be a function, that is, for any fact there is only one assertion describing it. In fact, we must guarantee that, if two facts f_1 and f_2 are different then they cannot both be the result of the interpretation of the same assertion a, i.e., it cannot be that if $\mathcal{I}_A(a) = f_1$ then also $\mathcal{I}_A(a) = f_2$. This phenomenon, called *polysemy* is pervasive in natural languages and it is one of the main sources of misunderstandings and, therefore, of the construction of diverging mental representations of the same representation. The polysemy of assertions arises directly from the polysemy of words. As examples: the proper name *Java* has three meanings, that is, it is a programming language, a type of coffee beans, and an island. The word *car* ha various meanings. For instance it may mean automobile or a car part of of a train. General words, such as *to do* have more than ten meanings. Polysemy is common to most words, in particular with those words which are most commonly used (people tend to give words their own specific meaning) and it is one of the major complications (not the only one) which arise when building natural language understanding systems.

Observation 3.8 (The non ambiguity of interpretation functions) As from Section 1.3 linguistic descriptions are ambiguous. As from Observation 3.7, one of the main reasons is the polysemy of words. However this ambiguity is in the mind of the listener/reader. The speaker/writer can be assumed to always have in mind the unique analogical representation (s)he is describing. The notion of interpretation function enforces this assumption forcing the speaker/writer to be explicit about the intended meaning.

Observation 3.9 (Interpretation function, synonymity) Two assertions are synonyms when they have the same meaning, that is, the interpretation of two different assertions a_1 and a_2, may denote the same fact f, i.e., $\mathcal{I}_A(a_1) = \mathcal{I}_A(a_2) = f$. Synonymous words are again pervasive in natural languages, in particular with the most common entities. People, and entities in general, have multiple names, e.g., name, surname, name plus surname, nicknames, which are synonymous. Multiple languages generate multiple names of the same entity (e.g., Great Britain, Gran Bretagna). There are also synonymous nouns, for instance *car* and *automobile*. Notice how the word *car* is both polysemous and synonymous. This is again quite common. In general synonymity is not a problem. However, in relational databases synonymity is not allowed, essentially for efficiency reasons. Databases are developed based on the *unique name assumption*, that is, in databases, different strings and assertions always mean different things.

Example 3.4 (An interpretation function providing an interpretation of the assertional theory describing Figure 3.1) . The natural interpretation function which interprets the sentences in Example 3.3 (left) to the facts in Example 3.2 (right) is

$\mathcal{I}_A$(Sofia è una persona) = `Sofia is a person`
$\mathcal{I}_A$(Paolo è un uomo) = `Paolo is a man`
$\mathcal{I}_A$(Rocky is a dog) = `Rocky is a dog`
$\mathcal{I}_A$(Sofia is near Paolo) = `Sofia is near Paolo`
$\mathcal{I}_A$(Rocky è il cane di Sofia) = `Rocky is the dog of Sofia`
$\mathcal{I}_A$(Sofia è un'amica di Paolo) = `Sofia is a friend of Paolo`
$\mathcal{I}_A$(Rocky è un animale) = `Rocky is an animal`
$\mathcal{I}_A$(Sofia ha i capelli biondi = `Sofia has blond hair`
...

Observation 3.10 (Assertions and facts, subjectivity) The problem of the subjectivity of representations remains, this being an unavoidable fact of life. However the notions of fact, assertion and interpretation function give leverage. First, facts are assumed to be unequivocally described, via interpretation functions, by assertions where, in turn, are linguistic representations and, as such, can be shared. Second, assertions, though subjectively selected by humans, are assumed to be atomic, that is, to provide the minimal possible level of details at which a model can be described.

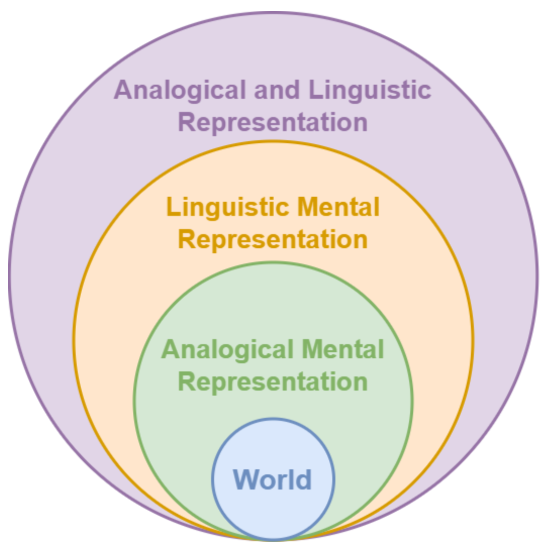


Fig. 3.2 Diagram of Representations

Chapter 4
Formal models and assertional theories

In order to avoid fallacious reasoning we need to represent models and assertional theories in a unambiguous, that is formal, way. Four are the features of of interest to us:

- **Formality**: It should be a logical language, that is, with well defined syntax and semantics;
- **Universality**: it should be able to represent all types of facts;
- **Intuitiveness**: it should allow for assertions whose basic elements (entity names, concepts and properties) as well as their structure (that is how the basic elements are connected together to build assertions) should be, on one side, intuitive to people while, on the other side, have a direct map to the structure and organization of the reference domain;
- **Computational efficiency**: $\mathcal{L}_A$ should allow for a fast and efficient inference engine, exploiting the inherent efficiency of the data structures used to memorize the world model.

Observation 4.1 (Types of assertional languages) ER and UML models are intuitive but not universal, they represent only knowledge facts. DBs are computationally efficient but represent only data facts. Natural language is universal but its semantics are not formally defined and it is not computationally efficient. The latter weakness extends to well defined subsets of natural languages. As an instance of this case, logical languages are universal with well defined syntax and semantics but they are not intuitive to understand and also computationally not efficient (see also Section **??**).

In Section 4.1, we introduce some basic definitions of set theory, useful in order to define `D`, while, in Section 4.2, we introduce some basic definitions of graph theory useful in order to define $\mathcal{L}_A$.

4.1 Set theory

4.1.1 Basic definitions

We can define sets in two ways

- **Listing:** The set is described by listing all its elements (for instance, $A = \{a, e, i, o, u\}$).
- **Abstraction:** The set is described through a property of its elements (for instance, $A = \{x|x \text{ is a vowel of the Latin alphabet}\}$).

We have the following basic definitions.

Definition 4.1 (Empty Set) $\emptyset$ is the set containing no elements.

Definition 4.2 (Membership) $\boldsymbol{a \in A}$, element $\boldsymbol{a}$ belongs to the set $\boldsymbol{A}$.

Definition 4.3 (Non-membership) $\boldsymbol{a \notin A}$, element $\boldsymbol{a}$ doesn't belong to the set $\boldsymbol{A}$.

Definition 4.4 (Equality) $\boldsymbol{A = B}$, if and only if $\boldsymbol{A}$ and $\boldsymbol{B}$ contain the same elements.

Definition 4.5 (Inequality) $\boldsymbol{A \neq B}$, if and only if it is not true that $\boldsymbol{A = B}$.

Definition 4.6 (Subset) $\boldsymbol{A \subseteq B}$, if and only if all elements in $\boldsymbol{A}$ also belong to $\boldsymbol{B}$.

Definition 4.7 (Proper Subset) $\boldsymbol{A \subset B}$, if and only if $\boldsymbol{A \subseteq B}$ and $\boldsymbol{A \neq B}$.

Definition 4.8 (Universal Set) The universal set is the set of all elements or members of all related sets and is denoted by the letter $\boldsymbol{\mathcal{U}}$.

We use **Venn diagrams** to represent sets. Venn diagrams consist of overlapping or intersecting circles representing sets and their relationships. Each circle represents a specific set, and the area where the circles overlap represents the elements shared between the corresponding sets. An element that does not belong to a set is represented as a dot outside the circle representing the set.

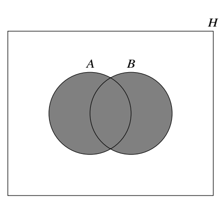


Fig. 4.1 Union set operation

Definition 4.9 (Union) Given two sets $\boldsymbol{A}$ and $\boldsymbol{B}$, the union of $\boldsymbol{A}$ and $\boldsymbol{B}$ is defined as the set containing the elements belonging to $\boldsymbol{A}$ or to $\boldsymbol{B}$ or to both, and is denoted with $\boldsymbol{A \cup B}$.

Definition 4.10 (Intersection) Given two sets $\boldsymbol{A}$ and $\boldsymbol{B}$, the intersection of $\boldsymbol{A}$ and $\boldsymbol{B}$ is defined as the set containing the elements that belong both to $\boldsymbol{A}$ and $\boldsymbol{B}$, and is denoted with $\boldsymbol{A} \cap \boldsymbol{B}$.

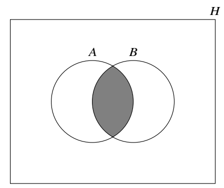


Fig. 4.2 Intersection set operation

Definition 4.11 (Difference) Given two sets $\boldsymbol{A}$ and $\boldsymbol{B}$, the difference of $\boldsymbol{A}$ and $\boldsymbol{B}$ is defined as the set containing all the elements which are members of $\boldsymbol{A}$, but not members of $\boldsymbol{B}$, and is denoted with $\boldsymbol{A} \setminus \boldsymbol{B}$.

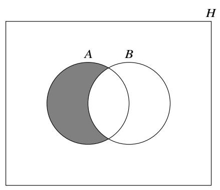


Fig. 4.3 Difference set operation

Definition 4.12 (Complement) Given a universal set $\boldsymbol{U}$ and a set $\boldsymbol{A}$, where $\boldsymbol{A} \subseteq \boldsymbol{U}$, the complement of $\boldsymbol{A}$ in $\boldsymbol{U}$ is defined as the set containing all the elements in $\boldsymbol{U}$ not belonging to $\boldsymbol{A}$, and is denoted with $\boldsymbol{A^c}$ or $\boldsymbol{U} \setminus \boldsymbol{A}$.

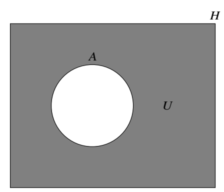


Fig. 4.4 Complement set operation

Theorem 4.1 (Properties of Operations)

- ***With same set***
  - $A \cap A = A$
  - $A \cup A = A$
- ***Commutative***
  - $A \cap B = B \cap A$
  - $A \cup B = B \cup A$
- ***Empty set***
  - $A \cap \emptyset = \emptyset$
  - $A \cup \emptyset = A$
- ***Associative***
  - $(A \cap B) \cap C = A \cap (B \cap C)$
  - $(A \cup B) \cup = A \cup (B \cup C)$
- ***Distributive***
  - $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$*;*

 – $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

- ***De Morgan laws***
  – $\overline{A \cup B} = \overline{A} \cap \overline{B}$
  – $\overline{A \cap B} = \overline{A} \cup \overline{B}$

4.1.2 Relations

Definition 4.13 (Cartesian product) Given two sets $\boldsymbol{A}$ and $\boldsymbol{B}$, the Cartesian product of $\boldsymbol{A}$ and $\boldsymbol{B}$ is defined as the set of ordered couples $\boldsymbol{(a, b)}$ where $\boldsymbol{a \in A}$ and $\boldsymbol{b \in B}$, formally:

$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\}$$

Example 4.1 (Cartesian product) Given $A = \{1, 2, 3\}$ and $B = \{a, b\}$, then

$$A \times B = \{(1, a), (1, b), (2, a), (2, b), (3, a), (3, b)\}$$
$$and$$
$$B \times A = \{(a, 1), (a, 2), (a, 3), (b, 1), (b, 2), (b, 3)\}$$

Definition 4.14 (Relation) A relation R from the set A to the set B is a subset of the Cartesian product of A and B: $R \subseteq A \times B$.
If $(x, y) \in R$, then we will write xRy and we say 'x is R-related to y'.

Proposition 4.1 *A binary relation on a set A is a subset $R \subseteq A \times A$.*

Given a relation R from A to B:

- the **domain** of R is the set $Dom(R) = \{a \in A | \text{there exists a } b \in B, aRb\}$
- the **co-domain** of R is the set $Cod(R) = \{b \in B | \text{there exists an } a \in A, aRb\}$

Example 4.2 Given $A = \{1, 2, 3, 4\}$, $B = \{a, b, d, e, r, t\}$ and aRb iff in the Italian name of a there is the letter b, then $R = \{(2, d), (2, e), (3, e), (3, r), (3, t), (4, a), (4, r), (4, t)\}$

Example 4.3 Given $A = \{3, 5, 7\}$, $B = \{2, 4, 6, 8, 10, 12\}$ and aRb iff a iff a is a divisor of b, then $R = \{(3, 6), (3, 12), (5, 10)\}$

Definition 4.15 (Inverse relation) Let R be a relation from A to B. The inverse relation of R is the relation $R^{-1} \subseteq B \times A$ where

$$R^{-1} = \{(b, a) | (a, b) \in R\}$$

Definition 4.16 (Relation Properties) Let R be a binary relation A.R is:

- **reflexive** iff aRa for all $a \in A$
- **symmetric** iff aRb implies bRa for all $a, b \in A$
- **transitive** iff aRb and bRc imply aRc for all $a, b, c \in A$

- **anti-symmetric** iff aRb and bRa imply $a = b$ for all $a, b \in A$

Definition 4.17 (Equivalence relation) Let R be a binary relation on a set A. R is an equivalence relation iff it satisfies all the following properties:

- reflexive
- symmetric
- transitive

Remark 4.1 An equivalence relation is usually denoted with ∼ or ≡

Definition 4.18 (Set partition) Let A be a set, a partition of A is a family F of non-empty subsets of A so that:

- the subsets are pairwise disjoint
- the union of all subsets is the set A

Remark 4.2 Each element of A belongs to exactly one subset in F

Definition 4.19 (Equivalence class) Let A be a set and ≡ an equivalence relation on A, given an x ∈ A we define equivalence class X the set of elements x' ∈ A s.t. x' ≡ x, formally:

$$X = \{x' | x' \equiv x\}$$

Remark 4.3 Any element x is sufficient to obtain the equivalence class X, which is denoted also with [x].

x ≡ x' implies [x]=[x']=X

Definition 4.20 (Quotient set) We define quotient set of A with respect to an equivalence relation ≡ as the set of equivalence classes defined by ≡ on A, and denote it with A / ≡.

Theorem 4.2 *Given an equivalence relation ≡ on A, the equivalence classes defined by ≡ on A are a partition of A. Similarly, given a partition on A, the relation R defined as xRx' iff x and x' belong to the same subset, is an equivalence relation on A.*

Example 4.4 (Parallelism relation) Two straight lines in a plane are parallel if they do not have any point in common or if they coincide.
The parallelism relation || is an equivalence relation since it is:

- reflexive: $r||r$
- symmetric: $r||s$ implies $s||r$
- transitive $r||s$ and $s||t$ imply $r||t$

We can thus obtain a partition in equivalence classes: intuitively, each class represent a direction in the plane.

Order relation:

Definition 4.21 (Order) Let A be a set and R be a binary relation on A.
R is an order (**partial**), usually denoted with $\leq$, if it satisfies the following properties:

- reflexive $a \leq a$
- anti-symmetric $a \leq b$ and $b \leq a$ imply $a = b$
- transitive $a \leq b$ and $b \leq c$ imply $a \leq c$

If the relation holds for all $a, b \in A$ then it is a **total order**.

A relation is a **strict order**, denoted with "<", if it satisfies the following properties:

- transitive $a < b$ and $b < c$ imply $a < c$
- for all $a, b \in A$ either $a < b$ or $b < a$ or $a = b$

4.1.3 Functions

Definition 4.22 (Functions) Given two sets A and B, a function f from A to B is a relation that associates to each element a in A exactly one element b in B. Denoted with:

$$f : A \rightarrow B$$

The domain of f is the whole set A.
The image of each element a in A is the element b in B s.t. $b = f(a)$.
The co-domain of f (or image of f) is a subset of B defined as follows:

$$Im_f = \{b \in B | \text{ there exists an } a \in A \text{ s.t. } b = f(a)\}$$

Remark 4.4 It can be the case that the same element in B is the image of several elements in A.

Classes of functions:

Definition 4.23 (Surjective function) A function $f : A \rightarrow B$ is surjective if each element in B is image of some elements in A :

$$\text{for each } b \in B \text{ there exists an } a \in A \text{ s.t. } f(a) = b$$

Definition 4.24 (Injective function) A function $f : A \rightarrow B$ is injective if distinct elements in A have distinct images in B :

$$\text{for each } b \in Im_f \text{ there exists a unique } a \in A \text{ s.t. } f(a) = b$$

Definition 4.25 (Bijective function) A function $f : A \rightarrow B$ is bijective if it is injective and surjective:

$$\text{for each } b \in B \text{ there exists a unique } a \in A \text{ s.t. } f(a) = b$$

Definition 4.26 (Inverse function) If $f : A \longrightarrow B$ is bijective we can define its inverse function:

$$f^{-1} : B \rightarrow A$$

Remark 4.5 For each function f there is a inverse relation. This relation is a function iff f is bijective.

Example 4.5 (Inverse function) Example of two different inverse functions:

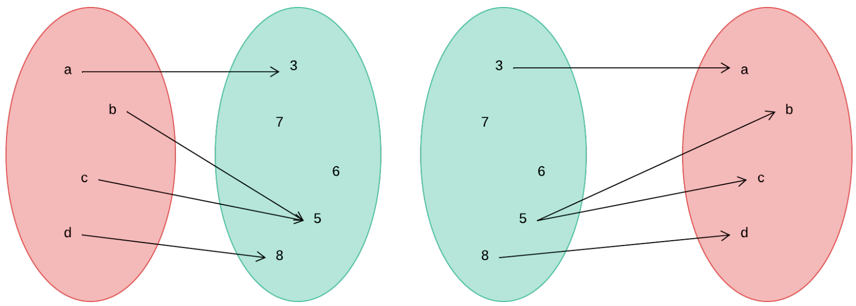


Fig. 4.5 Inverse of not bijective function

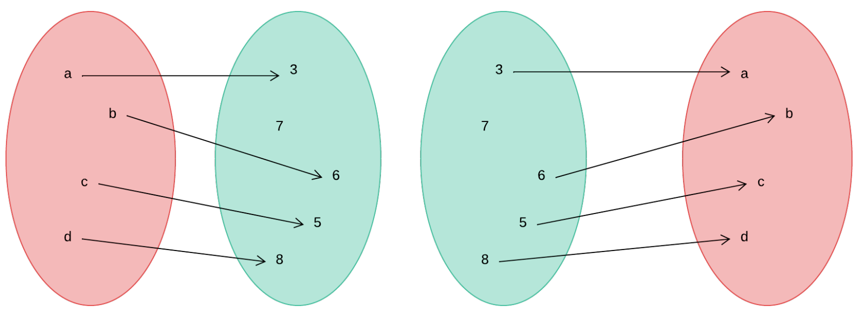


Fig. 4.6 Inverse of bijective function

Definition 4.27 (Composite function) Let $f : A \longrightarrow B$ and $g : B \longrightarrow C$ be functions. The composition of f and g is the function $g \circ f : A \longrightarrow C$ obtained by applying f and then g:

- $(g \circ f)(a) = g(f(a))$ for each $a \in A$
- $g \circ f = \{(a, g(f(a))|a \in A)\}$

4.2 Graph theory

4.2.1 Basic Notions

Definition 4.28 (Graph) A **graph** G is an ordered pair $G =< V, E >$, where V is the set of **vertices** (or **nodes**) and E is the set of **edges** (or **links**). Edges are pairs of vertices.

Definition 4.29 (Order) The **order** of a graph is the number of vertices of the graph.

Definition 4.30 (Size) The **size** of a graph is the number of edges in the graph.

Definition 4.31 (Degree) The **degree** of a vertex is the number of edges incident on that vertex.

Definition 4.32 (Directed graph) A **directed graph** is a graph where edges are ordered pairs of distinct vertices (x, y). x and y are called the **end points**, where x is the **tail** and y is the **head**.

From now on we concentrate on directed graphs.

Definition 4.33 (Leaf, intermediate node) In a directed graph, a **leaf** is a node with no outgoing nodes. A node which is not a leaf is an **intermediate node**.

Definition 4.34 (Path) A **path**, also called a **linear graph**, is a graph where the vertices can be ordered in a sequence $v_1, v_2, \dots, v_n$, where the edges correspond to the pairs of consecutive vertices $\{v_i, v_{i+1}\}$ for $i = 1, 2, \dots, n - 1$.

TODO FIGURE

Definition 4.35 (Cycle, cyclic graph) A **cycle**, also called a **circular graph**, is a path in which only the first and last vertices are equal. A **cyclic graph** is a graph which contains a cycle.

TODO FIGURE

Definition 4.36 (Tree, rooted tree, root, leaf, intermediate nodes) A **tree** is an undirected graph in which any two vertices are connected by exactly one path. A **polytree**, or **or directed tree**, or **oriented tree**, is a directed acyclic graph whose underlying undirected graph is a tree. A **rooted tree** is a tree in which one vertex has been designated the root. A **root** is a node with no incoming nodes.

TODO FIGURE

Definition 4.37 (Forest, polyforest, directed forest, oriented forest) A **forest** is an undirected graph. A **polyforest**, or **directed forest**, or **oriented forest**, is a directed acyclic graph whose underlying undirected graph is a forest.

TODO FIGURE

Definition 4.38 (Directed acyclic graph (DAG)) A **directed acyclic graph** (**DAG**) is a directed graph that does not contain any cycles.

TODO FIGURE

4.2.2 Labeled Graphs

From now on we concentrate on labeled directed graphs.

Definition 4.39 (Labeled Graph) A labeled graph is a type of graph where each vertex and edge is assigned a label.

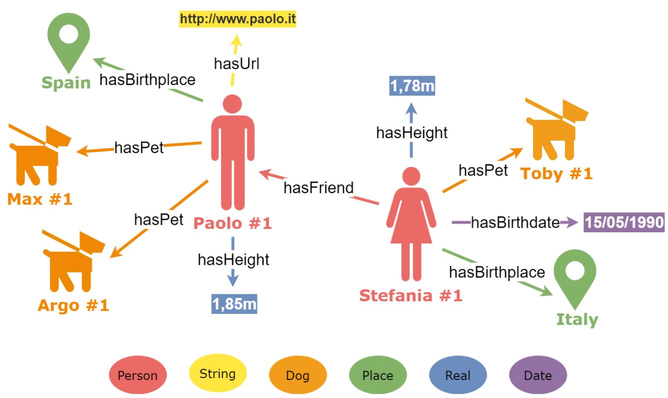


Fig. 4.7 Labeled Graph

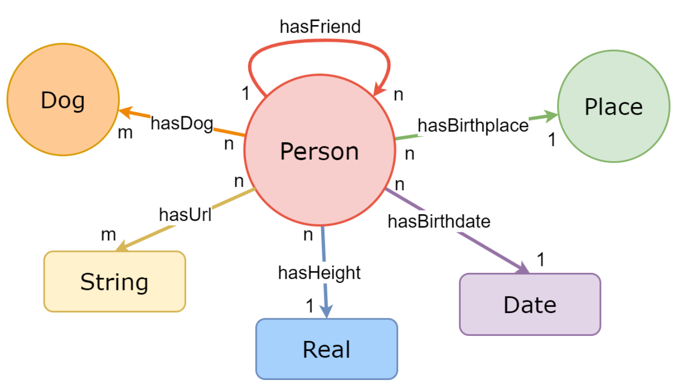


Fig. 4.8 Labeled Graph

Solutions

Exercises of Chapter 2

Solution **2.1 (Linguistic and analogical mental representations).** We create the following linguistic representation to describe the analogical one:

- In(tree, lab)
- In (monkey1, lab)
- In(monkey2, lab)
- Eating(monkey1, banana)
- SittingOn(monkey2, tree)
- Scratching(monkey2, hisHead)

Solution **2.2 (Linguistic and analogical mental representations).** The analogical representation is the same as the previous exercise:

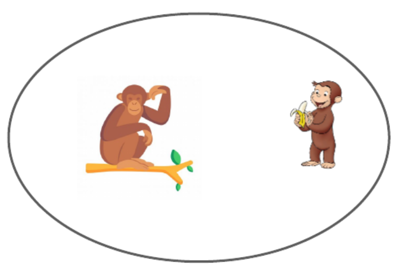


Fig. 4.9 Monkey and banana

Exercises of Chapter 4

Solution **??** **(Linguistic and analogical mental representations).** We can create an analogical representation using set theory in this way:

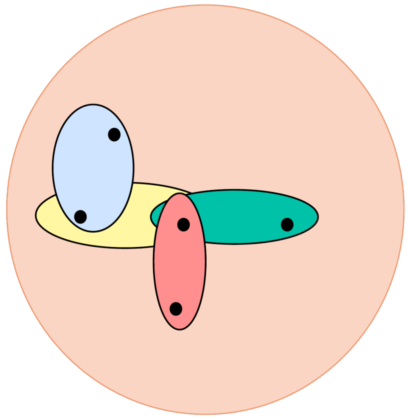


Fig. 4.10 Monkey and banana

Solution **??** **(Linguistic and analogical mental representations).** This time we use a labeled set theory diagram:

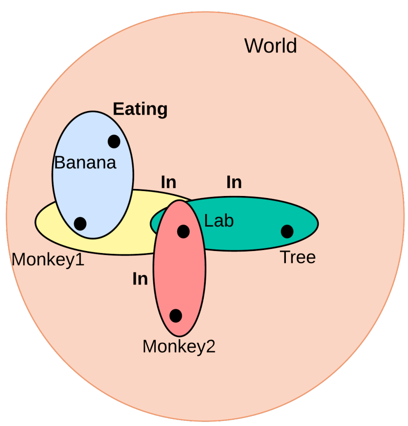


Fig. 4.11 Monkey and banana

Solution **??** **(Linguistic and analogical mental representations).** We can create an analogical representation using knowledge graphs in this way:

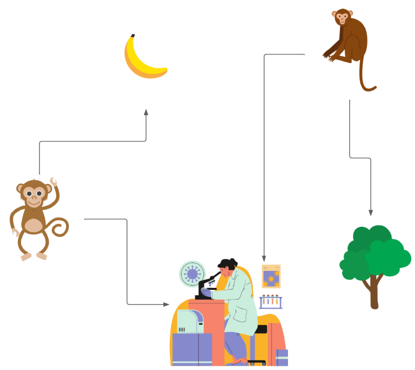


Fig. 4.12 Monkey and banana

Solution **??** **(Linguistic and analogical mental representations).** This time we use a labeled knowledge graph:

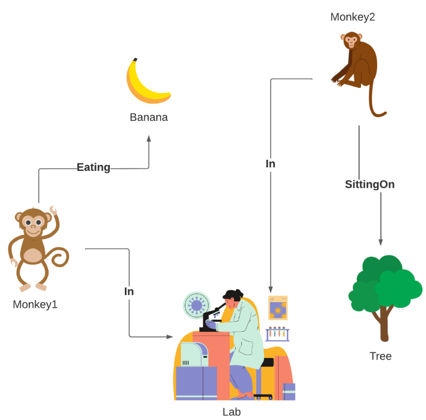


Fig. 4.13 Monkey and banana

References

1. Lingelbach B., Block B., Hatzky B., and Reisinger E. The hermann grid illusion – retinal or cortical? *Perception*, 1985.
2. Stephen F. Barker. *The Elements of Logic - Chapter 6*. New York, NY: McGraw-Hill, sixth edition, 2003.
3. S. Morris Engel. *With Good Reason: An Introduction to Informal Fallacies*. New York: St. Martin's Press, 1982.
4. Robert M Entman. Framing: Toward clarification of a fractured paradigm. *Journal of communication*, 43(4):51–58, 1993.
5. Erving Goffman. *Frame analysis: An essay on the organization of experience.* Harvard University Press, 1974.
6. Daniel Kahneman and Amos Tversky. The psychology of preferences. *Scientific American*, 246(1):160–173, 1982.
7. Alfred Korzybski. *Science and Sanity. An Introduction to Non-Aristotelian Systems and General Semantics*. The International Non-Aristotelian Library Pub., 1933.
8. Elizabeth F Loftus and John C Palmer. Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of verbal learning and verbal behavior*, 13(5):585–589, 1974.
9. Elliott Mendelson. *Introduction to Mathematical Logic*. Chapman & Hall, sixth edition, 2015.
10. Maarten de Rijke Patrick Blackburn. *Modal Logic*. Cambridge University Press, first edition, 2001.
11. Wolfgang Rautenberg. *A Concise Introduction to Mathematical Logic*. Springer New York, second edition, 2006.
12. Hoffrage U. *Cognitive Illusions: a handbook on fallacies and biases in thinking, judgement and memory*. Psychology Press, 2004.