Cognición y Lenguaje

Bárbara Eizaga Rebollar
CONTENTS:

1. The academic work 4
   1.1. How to write a literature review? Content, form and quotations 4
   1.2. Review sample 15

2. Cognition and Language 19
   2.1. Introduction: epistemology 19
   2.2. Types of thinking 21
   2.3. Deductive thinking 24
   2.4. Inductive thinking 33
   TEST 1: Are we rational solving tasks? 35

3. Cognitivism 44
   3.1. Introduction: Cognition and mind 44
   3.2. Summary of the ontological theories 45
   3.3. Historical overview of cognitivism 47
      3.3.1. First revolution: conductism and mind 47
      3.3.2. Second revolution: cognition and mind 49
   3.4. Human cognition: representation and computation 50
   TEST 2: The Game of Life 51

4. Knowledge representation in the mind 55
   4.1. Processable elements and their working 55
   4.2. Evolutionary problems: intentionality 59
   TEST 3: Intentions 63

5. Intelligent systems: computers and mind 64
   5.1. Introduction 64
   5.2. Mind architectures 64
   5.3. Symbolic architectures 65
      5.3.1. Origins of the symbolic cognitivism: Alan Turing 65
      5.3.2. Fundamental concepts of the computational theory 67
      5.3.3. Information processing in serial systems 68
      5.3.4. Artificial intelligence and computer simulation of behavior 69
      5.3.5. Can machines think? 69
      5.3.6. The computational paradox and the intentionality problem 70
   TEST 4: Dialog with machines 71
6. Connectionist architectures

6.1. The architecture of the connectionist mind
6.1.1. The brain metaphor of the mind
6.1.2. Structure of the connectionist approaches

6.2. Networks
6.2.1. Monolayer networks: the perceptron
6.2.2. Multilayer networks
6.2.3. Feedback networks: Visual recognition
6.2.4. WTA networks (Winner-take-all): Word recognition

6.3. Connectionist vs. Symbolic models
6.4. Advantages and disadvantages of the connectionist models

TEST 5: Artificial networks: Computation and learning algorithm

7. Information processing: selective control and motor activity

7.1. Previous setting: homunculi observing the world?
7.2. Is all processing controlled?
7.3. Selective control
7.4. Attentional control
7.4.1. Overt and covert attentional orientation
7.4.2. Stimuli- and goal-driven attentional control

7.5. Information processing: the sight
7.5.1. Saccadic activity, visual fixations and pupillary activity
7.5.2. Cognitive aspects related to visual recognition

7.6. Registros sensoriales en el procesamiento de la información
7.6.1. The visual sensory register: basic properties

TEST 6: Information processing control and eye movements

References
1. The academic work

1.1. How to Write a Literature Review

While the main focus of an academic research paper is to support your own argument, the focus of a literature review is to summarize and synthesize the arguments and ideas of others. The academic research paper also covers a range of sources, but it is usually a select number of sources, because the emphasis is on the argument. Likewise, a literature review can also have an “argument,” but it is not as important as covering a number of sources. In short, an academic research paper and a literature review contain some of the same elements. In fact, many academic research papers will contain a literature review section. But it is the aspect of the study (the argument or the sources) that is emphasized that determines what type of document it is.

1. Introduction:

A review describes, analyzes and evaluates. The review conveys an opinion, supporting it with evidence from the book. A literature review surveys scholarly articles, books and other sources (e.g. dissertations, conference proceedings) relevant to a particular issue, area of research, or theory, providing a description, summary, and critical evaluation of each work. The purpose is to offer an overview of significant literature published on a topic.

Before reading, consider:

- Title - What does it suggest?
- Preface or Introduction - Provides important information about the author's intentions or the scope of the book. Can you identify any limitations? Has the author ignored important aspects of the subject?
- Table of Contents - Shows how the book's organized -- main ideas, how they're developed (chronologically, topically, etc.)

Points to ponder as you read the entire book:

- Read the Book. This may seem obvious, but reading the book is key to a good review.
- Understand the book. If you don't understand the book, you cannot write a good review.
- As you read, make notes of different things about the book to quote in your review. When you have sufficient notes you are ready to write the review.

2. Components

The development of the literature review requires four stages:

- Problem formulation—which topic or field is being examined and what are its component issues?
- Literature search—finding materials relevant to the subject being explored.
- Data evaluation—determining which literature makes a significant contribution to the understanding of the topic.
- Analysis and interpretation—discussing the findings and conclusions of pertinent literature.

Literature reviews should comprise the following elements:

- Include title, author, place, publisher, publication date, edition, pages, special features (maps, etc.), ISBN.
In the first section, give a brief history of the author (previous works, awards, etc.)
In the second section, write an overview of the subject or theory under consideration, along with the objectives of the literature review. Divide the works under review into categories (e.g. those in support of a particular position, those against, and those offering alternative theses entirely).
Next, use your notes to evaluate the book by saying whether you agree or disagree with the author's point of view, whether you can follow the author's thesis, whether concepts are well-defined and ideas are well-developed, what areas are covered, how accurate the index and information in the book are, what sources used the author, whether important information was omitted, whether footnotes clarified or extended the points made in the text and what the book accomplished. Finally, compare the book to others by this author, or book in this field by other authors; explain how each work is similar to and how it varies from the others. (Use the books listed in the bibliography.) If this is the best book you have ever read, say so -- and why. If it is merely another nice book, say so.
Use the final paragraph as a summary of the whole review. Your conclusion should summarize, perhaps include a final assessment. Do not introduce new material at this point. Give your opinion of the book (i.e., which pieces are best considered in their argument, are most convincing of their opinions, and make the greatest contribution to the understanding and development of their area of research) and finish by recommending (or not) the book. State who would enjoy this book. At the end, include your full name and, optionally, your e-mail address.

In assessing each piece, consideration should be given to:

- Provenance—What are the author's credentials? Are the author's arguments supported by evidence (e.g. primary historical material, case studies, narratives, statistics, recent scientific findings)?
- Is contrary data considered or is certain pertinent information ignored to prove the author's point?
- Persuasiveness—Which of the author's theses are most/least convincing?
- Value—Are the author's arguments and conclusions convincing? Does the work ultimately contribute in any significant way to an understanding of the subject?

3. Definition and Use/Purpose

The purpose of a literature review is to:

- Place each work in the context of its contribution to the understanding of the subject under review.
- Describe the relationship of each work to the others under consideration.
- Identify new ways to interpret, and shed light on any gaps in, previous research.
- Resolve conflicts among seemingly contradictory previous studies.
- Identify areas of prior scholarship to prevent duplication of effort.
- Point the way forward for further research.
- Place one's original work (in the case of theses or dissertations) in the context of existing literature.

4. Strategies for writing the literature review

Find a focus

A literature review is usually organized around ideas, not the sources themselves as an annotated bibliography would be organized. This means that you will not just simply list your sources and go into detail about each one of them, one at a time. No. As you read widely but selectively in your topic area, consider instead what themes or issues connect your sources
together. Do they present one or different solutions? Is there an aspect of the field that is 
missing? How well do they present the material and do they portray it according to an 
appropriate theory? Do they reveal a trend in the field? A raging debate? Pick one of these 
themes to focus the organization of your review.

Construct a working thesis statement

Then use the focus you have found to construct a thesis statement (i.e., it tells the reader what 
to expect from the rest of the paper). Literature reviews have thesis statements as well. 
However, your thesis statement will not necessarily argue for a position or an opinion; rather it 
will argue for a particular perspective on the material. Some sample thesis statements for 
literature reviews are as follows:

1. The current trend in treatment for congestive heart failure combines surgery and 
   medicine.
2. More and more cultural studies scholars are accepting popular media as a subject worthy 
of academic consideration.

Consider organization

You have a focus and you have narrowed it down to a thesis statement. Now what is the most 
effective way of presenting the information? What are the most important topics, subtopics, 
etc., that your review needs to include? And in what order should you present them? Develop 
an organization for your review at both a global and local level:

First, cover the basic categories

Just like most academic papers, literature reviews must also contain at least three basic 
elements: an introduction or background information section; the body of the review containing 
the discussion of sources; and, finally, a conclusion and/or recommendations section to end the 
paper.

Introduction: Gives a quick idea of the topic of the literature review, such as the central theme 
or organizational pattern.

Body: Contains your discussion of sources and is organized either chronologically, thematically, 
or methodologically (see below for more information on each).

Conclusions/Recommendations: Discuss what you have drawn from reviewing literature so far. 
Where might the discussion proceed?

Organizing the body

Once you have the basic categories in place, then you must consider how you will present the 
sources themselves within the body of your paper. Create an organizational method to focus 
this section even further.

To help you come up with an overall organizational framework for your review, consider the 
following scenario and then three typical ways of organizing the sources into a review:

You've decided to focus your literature review on materials dealing with sperm whales. This is 
because you've just finished reading Moby Dick, and you wonder if that whale's portrayal is 
really real. You start with some articles about the physiology of sperm whales in biology 
journals written in the 1980's. But these articles refer to some British biological studies 
performed on whales in the early 18th century. So you check those out. Then you look up a
book written in 1968 with information on how sperm whales have been portrayed in other forms of art, such as in Alaskan poetry, in French painting, or on whale bone, as the whale hunters in the late 19th century used to do. This makes you wonder about American whaling methods during the time portrayed in Moby Dick, so you find some academic articles published in the last five years on how accurately Herman Melville portrayed the whaling scene in his novel.

**Chronological**

If your review follows the chronological method, you could write about the materials above according to when they were published. For instance, first you would talk about the British biological studies of the 18th century, then about Moby Dick, published in 1851, then the book on sperm whales in other art (1968), and finally the biology articles (1980s) and the recent articles on American whaling of the 19th century. But there is relatively no continuity among subjects here. And notice that even though the sources on sperm whales in other art and on American whaling are written recently, they are about other subjects/objects that were created much earlier. Thus, the review loses its chronological focus.

**By publication**

Order your sources by publication chronology, then, only if the order demonstrates a more important trend. For instance, you could order a review of literature on biological studies of sperm whales if the progression revealed a change in dissection practices of the researchers who wrote and/or conducted the studies.

**By trend**

A better way to organize the above sources chronologically is to examine the sources under another trend, such as the history of whaling. Then your review would have subsections according to eras within this period. For instance, the review might examine whaling from pre-1600-1699, 1700-1799, and 1800-1899. Under this method, you would combine the recent studies on American whaling in the 19th century with Moby Dick itself in the 1800-1899 category, even though the authors wrote a century apart.

**Thematic**

Thematic reviews of literature are organized around a topic or issue, rather than the progression of time. However, progression of time may still be an important factor in a thematic review. For instance, the sperm whale review could focus on the development of the harpoon for whale hunting. While the study focuses on one topic, harpoon technology, it will still be organized chronologically. The only difference here between a “chronological” and a “thematic” approach is what is emphasized the most: the development of the harpoon or the harpoon technology.

But more authentic thematic reviews tend to break away from chronological order. For instance, a thematic review of material on sperm whales might examine how they are portrayed as “evil” in cultural documents. The subsections might include how they are personified, how their proportions are exaggerated, and their behaviours misunderstood. A review organized in this manner would shift between time periods within each section according to the point made.

**Methodological**

A methodological approach differs from the two above in that the focusing factor usually does not have to do with the content of the material. Instead, it focuses on the “methods” of the researcher or writer. For the sperm whale project, one methodological approach would be to look at cultural differences between the portrayal of whales in American, British, and French art
work. Or the review might focus on the economic impact of whaling on a community. A methodological scope will influence either the types of documents in the review or the way in which these documents are discussed.

Once you have decided on the organizational method for the body of the review, the sections you need to include in the paper should be easy to figure out. They should arise out of your organizational strategy. In other words, a chronological review would have subsections for each vital time period. A thematic review would have subtopics based upon factors that relate to the theme or issue.

Sometimes, though, you might need to add additional sections that are necessary for your study, but do not fit in the organizational strategy of the body. What other sections you include in the body is up to you. Put in only what is necessary. Here are a few other sections you might want to consider:

*Current Situation*: Information necessary to understand the topic or focus of the literature review.

*History*: The chronological progression of the field, the literature, or an idea that is necessary to understand the literature review, if the body of the literature review is not already a chronology.

*Methods and/or Standards*: The criteria you used to select the sources in your literature review or the way in which you present your information. For instance, you might explain that your review includes only peer-reviewed articles and journals.

*Questions for Further Research*: What questions about the field has the review sparked? How will you further your research as a result of the review?

Use caution when paraphrasing

When paraphrasing a source that is not your own, be sure to represent the author's information or opinions accurately and in your own words. For example:

> However, other studies have shown that even gender-neutral antecedents are more likely to produce masculine images than feminine ones (Gastil, 1990). Hamilton (1988) asked students to complete sentences that required them to fill in pronouns that agreed with gender-neutral antecedents such as "writer," "pedestrian," and "persons." The students were asked to describe any image they had when writing the sentence. Hamilton found that people imagined 3.3 men to each woman in the masculine "generic" condition and 1.5 men per woman in the unbiased condition. Thus, while ambient sexism accounted for some of the masculine bias, sexist language amplified the effect. (Source: Erika Falk and Jordan Mills, "Why Sexist Language Affects Persuasion: The Role of Homophily, Intended Audience, and Offense," Women and Language19:2.

In the preceding example, Falk and Mills either directly refer in the text to the author of their source, such as Hamilton, or they provide ample notation in the text when the ideas they are mentioning are not their own, for example, Gastill's. But do not copy, avoid plagiarism. When in doubt, provide a citation.

Revise, revise, revise

Draft in hand? Now you are ready to revise. Spending a lot of time revising is a wise idea, because your main objective is to present the material, not the argument. So check over your review again to make sure it follows the assignment and/or your outline. Then, just as you would for most other academic forms of writing, rewrite or rework the language of your review so that you have presented your information in the most concise manner possible. Be sure to
use terminology familiar to your audience; get rid of unnecessary jargon or slang. Finally, double check that you have documented your sources and formatted the review appropriately for your discipline.

What does it mean to revise?

Revision is an ongoing process of rethinking the paper: reconsidering your arguments, reviewing your evidence, refining your purpose, reorganizing your presentation, reviving stale prose.

Style: Citations

In-text Citations

In MLA style, an in-text citation generally consists of the author's last name and the page number of the reference. When multiple elements are used in a parenthetical citation, they are separated by a space. In all cases except for block quotations, parenthetical citations are placed immediately before the final punctuation of the sentence that cites the work.

For Exact Quotes:

Rule: When a quotation runs no more than four lines, put it in quotes, incorporate the author into the text, and provide the page number in a parenthetical citation.

Example: "He was obeyed," writes Joseph Conrad of the company manager in Heart of Darkness, "yet he inspired neither love nor fear, nor even respect" (87).

Rule: When the author's name does not appear in the signal phrase, place the author's name and the page number(s) in the parenthetical citation.

Example: "If the existence of a signing ape was unsettling for linguists, it was also startling news for animal behaviourists" (Davis 26).

Rule: When a quotation runs more than four lines, do not use quotation marks, but indent it one inch from the main body of your text. Double space the block quote. Incorporate the author's name or the title of the piece into a signal phrase preceding the quote. Finally, provide the page number(s) of the excerpt, in parentheses, immediately following the final punctuation of the quotation.
Example: At the conclusion of *Lord of the Flies*, Ralph and the other boys realize the horror of their actions:

> The tears began to flow and sobs shook him. He gave himself up to them now for the first time on the island; great, shuddering spasms of grief that seemed to wrench his whole body. His voice rose under the black smoke before the burning wreckage of the island; and infected by that emotion, the other little boys began to shake and sob too. (186)

For Paraphrased Ideas:

Rule: When paraphrasing where the author(s) is clearly identified in your text, provide only the page number in the parenthetical citation.

Example: Others, like Jakobson and Smith, hold the opinion that children who attend preschool are better socially adjusted than those who do not (156).

Rule: When paraphrasing where author(s) is not clearly identified in your text, provide author(s) and page number in the citation.

Example: Between 1968 and 1988, television coverage of presidential elections changed dramatically (Hallin 5).

Citing from Indirect Sources:

Rule: When quoting a reference that is not originally from the source you have, after the reference use the phrase "qtd. in" (quoted in) and the author(s) of the source you have, the volume of the source (if more than one), and the page number. An indirect source may be documented in the Works Cited page.

Example: Samuel Johnson admitted that Edmund Burke was an "extraordinary man" (qtd. in Boswell 2: 450).

When do I use those three dots (…)?

Whenever you want to leave out material from within a quotation, you need to use an ellipsis, which is a series of three periods, each of which should be preceded and followed by a space. There are a few rules to follow when using ellipses:

1. **Be sure not to change the meaning of the quotation by omitting material.**

Take a look at the following example:
"The Writing Center is located on the UNC campus and serves the entire UNC community."

"The Writing Center . . . serves the entire UNC community."

The reader's understanding of the Writing Center's mission to serve the UNC community is not affected by omitting the information about its location.

2. **Use punctuation marks in combination with ellipses when removing material from the end of sentences or clauses.**

For example, if you take material from the end of a sentence, keep the period in as usual.

"The boys ran to school, forgetting their lunches and books. Even though they were out of breath, they made it on time."

"The boys ran to school. . . . Even though they were out of breath, they made it on time."

Likewise, if you excerpt material at the end of clause that ends in a comma, retain the comma.

"The red car came to a screeching halt that was heard by nearby pedestrians, but no one was hurt."

"The red car came to a screeching halt . . . , but no one was hurt."

3. **Including supplemental information that your reader needs to understand the quotation.**

For example, if you were quoting someone's nickname, you might want to let your reader know the full name of that person in brackets.

"The principal of the school told Billy [William Smith] that his contract would be terminated."

Similarly, if a quotation referenced an event with which the reader might be unfamiliar, you could identify that event in brackets.

"We completely revised our political strategies after the strike [of 1934]."

3. **Indicating the use of nonstandard grammar or spelling.**

In rare situations, you may quote from a text that has non-standard grammar, spelling, or word choice. In such cases, you may want to insert [sic], which means "thus" or "so" in Latin. Using [sic] alerts your reader to the fact that this non-standard language is not the result of a typo on your part. Always italicize "sic" and enclose it in brackets. There is no need to put a period at the end. Here's an example of when you might use [sic]:

Twelve-year-old Betsy Smith wrote in her diary, "Father is afraid that he will be guilty of beach [sic] of contract."

Here [sic] indicates that the original author wrote "beach of contract," not breach of contract, which is the accepted terminology.
Print Sources

Below you will find examples of how to cite print resources in a works-cited page. Although the examples on this page are presented single-spaced for clarity, please note that in MLA style the works-cited page is double-spaced.

Book


Edited Book


Translation (Book)


Work from an Anthology


Article from a Scholarly Journal


Article from a Newspaper


Newspaper Article without an Author


Article from a Magazine


Online and Nonprint Sources

Citations for electronic resources are, by and large, the same as for their print equivalents, with the addition of subscription information for databases, access dates, and URLs. Once again, examples are presented here single-spaced for clarity, but should appear double-spaced in a works-cited page.

Website

Blog Post


Newspaper Article without an Author, Retrieved from an Online Database


Full-text Article, Retrieved from an Online Database


Article from a Scholarly Journal (Online)


Sample Works-Cited Page

The citation examples provided in this tutorial are listed here as they would be on a works-cited page in a paper. The title "Works Cited" indicates that the list you provide contains only the works you actually cite in your paper. If you wish to also include in your list works that you consult but do not cite, give your page the broader title "Works Consulted."

Citations beginning with names and those beginning with titles are to be alphabetized together. Numbers in titles are treated as though they have been spelled out. For names, alphabetize based on the letters that come before the comma separating the last name from the first, and disregard any spaces or other punctuation in the last name. For titles, ignore articles such as "a" and "the" (and equivalents in other languages) for alphabetization purposes.


1.2. Review sample


Noam Chomsky is one of the most influential personalities in 20th century linguistics. His influence ranges from syntactic theory to the nature of the human mind and extends well into psychology and philosophy in general. In addition he is well known as a political activist. In this book, Neil Smith attempts to give an overview of all of these aspects and to show the coherence between Chomsky's views in the different areas.

**Synopsis**

The text is divided into an introduction and five chapters. The introduction is an intuitive evaluation of Chomsky's general intellectual achievement in terms of a comparison with other famous people, Descartes, Darwin, Freud, Einstein, and Picasso. The five chapters each discuss one aspect of Chomsky's thought.

**Chapter 1 is devoted to the nature of language and its relationship to the human mind.**

Chomsky's contribution in this respect can be summarized as creating the conditions for establishing linguistics as a science. In order to study language in a scientific way, the aim must be an explanatory account rather than an exhaustive description. Certain idealizations and departures from commonsense notions have to be made, but this is no different from what happens in physics.

Language is the knowledge component of the individual speaker rather than a corpus of utterances or a collection of grammatical sentences. Language can also be studied at species level in terms of the universal human ability for language acquisition.

**Chapter 2 deals with the development of Chomsky's linguistic theory.**

Originally, he distinguished surface structure and deep structure. Formulating rules for generating deep structure and transformations for deriving surface structure is not enough for an explanatory account, however. Therefore, Chomsky started looking for more general principles from an early date. The historical development towards Government and Binding, Principles and Parameters, and Minimalism shows a gradual increase in explanatory power.

**Chapter 3 discusses Chomsky's influence on psychology.**

The psychological reality of a linguistic theory should be interpreted in such a way that the entire body of available evidence is considered in order to produce a theory and this theory is then taken to be psychologically real. Evidence includes data from language processing, language acquisition, and language pathology. Essential for explanation is the concept of causation, a concept which is inherently non-observable. Connectionist approaches, which model the mind as a neural network, fail to account for the difference in type between the knowledge involved in language and in chess.

**Chapter 4 deals with the reception of Chomsky's ideas in philosophy.**

The philosophical tradition with its emphasis on truth conditions and public language rejects many of the key notions of Chomskyan linguistics. It considers language as a set of sentences used by a speech community for communication. Chomsky considers it an individual property and a species property, but not one that can be attributed to a community. Moreover, for Chomsky communication always depends on inferences and is not the primary function of language. The use of the label "realism" in this context is confusing, because it is claimed both by Chomsky and by his opponents (Quine, Montague).
Chapter 5 summarizes some of the more general trends in Chomsky's political positions.

Drawing his inspiration from a type of anarchism which rejects power as an argument to maintain the political situation as it is, Chomsky fiercely criticizes American foreign policy, e.g. in Vietnam and East Timor, American domestic policy, and manipulation of information in the media. His argumentation is always based on extensive documentary evidence. His radical rejection of authority also implies that he refuses to give advice on political matters.

Evaluation

The influence and controversial nature of Chomsky's positions in different areas has generated a substantial body of secondary literature on Chomsky. While explicitly distinguishing his enterprise from Barsky's (1997) biography, Smith presents us with a book which shares at least one essential property with biographies, namely that the link holding the chapters together is that they are all on the same person. One could call this book a thematically ordered, intellectual biography.

It is well known that the genre of biography is prone to degenerate into hagiography. By identifying with the subject of their biography, authors run the risk of indulging in uncritical admiration, the more so because sympathetic interest or admiration is likely to play a role in the choice of their subject in the first place. If, then, Smith (p. ix) states about Chomsky "It has been a privilege to work in his shadow" and thanks him for, among other things, sending him "some sixty pages of comments and suggestions" in reply to the pre-final version of the manuscript, we can hardly expect a critical account of Chomsky's ideas. This expectation is borne out.

Given the controversial nature of Chomsky's ideas, it is difficult to find works which convey them in the sense intended by Chomsky while keeping a certain critical distance. Most of the books with purely negative criticism of Chomsky's work are based on severe misunderstandings which make them unsuitable for gaining a general impression of his ideas. Conversely, Smith tends to focus so much on the description of Chomsky's ideas that the role of controversial discussion in their growth and development remains underrated. Thus in chapter 2 one almost gets the impression that the whole development of Chomskyan linguistics was brought about by Chomsky single-handedly. In chapter 4 it is sometimes difficult to form an idea of Chomsky's opponents as having a coherent system, because their objections are waived rather casually.

Perhaps the most useful evaluation of a book such as Smith's is a comparison with a number of other books which are intended to fulfil similar purposes. There are not so many books which combine overviews of Chomsky's scientific and political ideas. Salkie (1990) offers more detail in the latter, but considerably less on the former and is of course slightly dated. One could also look at a collection such as Chomsky (1997) which includes articles in both of these areas, but it is not a systematic overview. The strength of Smith's book is rather in his systematic exposition of Chomsky's scientific ideas than in the summary of his political ideas. As such it competes for instance with Botha (1989). Botha's book is more limited in scope in the sense that it concentrates on the underlying view of language and the human mind, while leaving out references to the actual theory. As a consequence it does not become outdated so quickly, because Chomsky's meta-theory is much more constant. A more recent competing book is Uriagereka (1998). This introduces both the metatheory and the current theory of Chomsky's Minimalist Programme. In this case, a major difference between the two is the length, Uriagereka's book being approximately three times as long as Smith's.

A comparison of the three books shows two disadvantages of Smith's book. First, it does not give such a good impression of the controversy triggered by Chomsky's positions. A remarkable feature of both Botha's and Uriagereka's books is that they introduce a special presentation technique in order to create a context of discussion. Smith uses plain academic prose which gives the impression of a soliloquy rather than a discussion. By taking the objections more seriously, Botha and Uriagereka actually reinforce the impression of the coherence of
Chomskyan ideas more than Smith does. As a second disadvantage, Smith is far less systematic in his presentation. In fact, it is difficult to get an overview of the structure of his book. Below the level of the chapters (roughly forty pages each), there are only unnumbered subheadings in two fonts which are not consistently used. In chapter 1, some of the subheadings occur in the table of contents, others which look the same in the text do not. In chapter 4 the heading "Controversies" is followed by nine lines of text and a heading of exactly the same type font. Apparently, some of the subsequent headings are meant to introduce further subsections of "Controversies", but it is not clear how many.

Obviously, Smith did not intend to give an overview of Chomskyan linguistic theory. Thus it would not be fair to compare the relevant sections in chapter 2 with a textbook or with historical overviews such as Newmeyer (1986). In order to give an impression of what the theory is like, Smith apparently tries to strike a balance between clarity of presentation, brevity of expression, and giving a real sense of the discussion. A typical example of the resulting type of presentation is found on p. 53ff., where Smith first introduces a number of sentences and contrastive sentence pairs used by Chomsky to illustrate his points and then explains their relevance. While this may still be seen as a creative technique to approximate the solution to an unsolvable problem, this can hardly be said of the presentation of the binding theory on p. 69ff. Two pages are reserved for the binding theory and most of it is devoted to picture-nouns and similar problem cases. I strongly suspect that few readers will learn anything from these pages: either they already know or they do not understand. For a general impression of Chomskyan theory, some of his own writings are probably more suitable.

Given the academic style of Smith's book, it is obvious that it is not meant for the same readership as such popular presentations as Pinker (1994) and Jackendoff (1993). It also has a much broader scope. However, if the style of Smith's book is considered, the question arises, why not read the original accounts such as Chomsky (1986) right away. Although Chomsky's scientific works have the reputation of being inaccessible, Smith hardly does anything to make the ideas easier to understand. The only concession to non-academic style seems to be the lack of footnote markers in the text. Instead, notes are given at the end of the book listed by page and with references to topics and quotes.

Finally, this book is an interesting example of British and American English mixed. Written by a British author in a British academic style, it uses American spelling throughout. The British and American perspectives of the author and the subject are also mixed in examples and chapter 5 has a subsection "The critique of (American) foreign policy" followed by "The critique of domestic policy".

**Conclusion**

This book gives a brief overview of Chomsky's ideas in the areas in which they are influential. It was written by an admirer. For general readers used to an academic style it might give a sense of Chomsky's reasoning and theorizing, but it is probably hard to understand much about the details given. Readers who tend to disagree with Chomsky will hardly be convinced by this book, because the author's attitude towards his subject is not just sympathetic but rather uncritical.

**References**


Reviewed by Pius Hacken (Universität Basel)
2. Cognition and language

2.1. Introduction: epistemology

Epistemology, also known as theory of knowledge, is a branch of philosophy, as ethics or logic. But unlike these theories, which are mostly normative (i.e., they are interested in the rules of moral behavior or good logical reasoning), epistemology is speculative; that is to say, its main interest is understanding knowledge as it is, not as it should be. Besides being part of philosophy, epistemology is the most important domain of the theory of science, an intellectual discipline including other studies such as history or sociology of science. Regardless of how we look at it, it is an abstract domain, or has been until now, in the sense that its principles are recognized and clarified by thinking, not by experimentation or observation. Its main goal is to understand the conditions in which human knowledge is possible, as well as the limits within which it can occur; that is to say, it judges its validity and its reach. Thus, it has to do with both empirical knowledge and rational knowledge.

At this point, several terminological terms should be pointed out. Epistemology comes from the Greek term, episteme, which means knowledge, and the suffix –logia, which means study of. Therefore, it seems intuitively logical that this study of knowledge has to do with what cognitians try to do: studying knowledge. It is important not to mistake epistemology with other two important studies interesting in knowledge as well: logic (the normative science aforementioned) and psychology (scientific theory dealing with mental processes and behaviour, mostly experimental). Logic studies specifically reasoning and tries to establish the formal conditions of knowledge; whereas psychology examines the cognoscent/human being in its psychophysical conditions, studying the mechanisms which allow the intellectual functioning of the different animal species, especially the Homo sapiens. Epistemology, on the other hand, is concerned with the nature and scope of knowledge as a bipolar relationship between an individual and an object, mainly with the elucidation of the correspondence between a thought and the reality it refers to. Some psychologists, as Jean Piaget, use the term epistemology to refer to the phenomenon of the genetics of knowledge in individual minds (genetic epistemology), including thus the history of scientific ideas as well as the study of development in individuals. This is, of course, a complete different sense of epistemology and should be distinguished from the one we have mentioned: the study of the nature and validity of knowledge. However, there is an important relationship between these two senses of the word, because the development of the individual minds generally occurs in an analogous way to the mankind’s discover of ideas. Usually, the individual development of the minds is motivated by the desire to get true knowledge.

Within philosophy, epistemology is very close to the study known as philosophy of mind or philosophical anthropology. This discipline seeks to present plausible hypotheses on understanding behaviour of humans as both creatures of their social environments and creators of their own values. Hence, it analyzes the relationship between physical and mental phenomena and, in general, the conditions that allow the existence of beings with mental life. This discipline has a long history in the Western tradition, dating back to the Greek sophists and Socrates. Throughout history, philosophy has provoked the most intense intellectual discussions trying to elucidate the question about the nature of the human being as opposed to other animals or living beings. In these discussions, the question at issue lies in the opposition between two very different trends: the ontological dualism, a set of views about the relationship between mind and matter, which claims that mental phenomena are, in some respects, non-physical whereas biological phenomena are the physical objects and processes; and the materialism or ontological monism, which holds that all things are composed of material and all phenomena (including consciousness) are the result of material interactions; in other words, matter is the only substance. As a theory, materialism is a form of physicalism which explains some effects deriving from some causes, although it is impossible to state that there are certain causes in reality. There may always be new features modifying those causes and this is what scientific thinking deals with: finding the causes and establishing the links between them and
their effects. Thus, science does not provide us with reality, but offers us a model of reality, which explains what we know of reality.

The theory of science of the last century is one of the most fascinating periods of the history of human thought. During this century, there are two different stages which occur consecutively and in which philosophers did a complete different intellectual job. The first stage corresponds to the formal philosophers, whose main concern was the logical structure of propositions in science. It has two perfectly distinguishable sub-stages:

1) the neo-positivist philosophers, also called logical positivists, which campaigned for a systematic reduction of human knowledge to logical and scientific foundations. Because the resulting logical positivism (or "logical empiricism") allowed only for the use of logical tautologies and first-person observations from experience, it dismissed as nonsense the metaphysical and normative pretensions of the philosophical tradition. Although participants sometimes found it difficult to defend the strict principles on which their programme depended, this movement offered a powerful vision of the possibilities for modern knowledge, introducing the symbolic logic in the theory of science. One of its main proponents is Rudolf Carnal, who outlined the logic of confirmation, an account of the methods and procedures by means of which we employ sensory observations to verify (or at least to confirm) the truth of scientific hypotheses about the operation of the physical universe. Using the formal methods of mathematical logic, then, the goal is to construct a strictly scientific language that perspicuously represents the structure of the world as a whole. Unfortunately, this could not be done. There is no inductive method: the association of the characteristics in space and time are not enough to postulate necessary and constant relations among them. In other words, the fundamental problem is that empirical generalizations are themselves incapable of direct support within such a system. The basis of the scientific method must be searched somewhere else.

2) In the 1940s a new formalist philosophy is developed, which moves away from the positivism but which still highlights the logical structure of the propositions and its rigorous treatment. It is represented by Karl Popper and his followers. This philosophy rejects the possibility of the logic of confirmation (or inductive logia) and argues for a logic of falsification or deductive logic. Popper's account of the logical asymmetry between verification and falsifiability lies at the heart of his philosophy of science. It also inspired him to take falsifiability as his criterion of demarcation between what is and is not genuinely scientific: a theory should be considered scientific if and only if it is falsifiable. In su esfuerzo por fundamentar las reglas del método científico, Popper explota, diestramente pero tal vez de manera extrema, la idea de que si bien muchos casos favorables no pueden confirmar una hipótesis, basta un solo caso contrario para refutarla. Lógicamente, el análisis es correcto, pero no resulta respaldado por las lecciones de la historia.

In the 60s a new trend is introduced by Thomas Kuhn, who shifts the main philosophical interest from the logical structure of utterances to the notion of scientific paradigm, changing tus the way of doing science. The notion of paradigm is the best contribution of the history of science to the development of epistemology. In recent years, the task of epistemologists has focused on evaluating the work carried out by the philosophy of science throughout the 20th century.
2.2. Types of thinking

Interpreting is an innate process in human beings. We all interpret the reality around us continuously without having learnt to do it. However, often times our interpretations are incoherent and/or slanted, and thus are primary. What is taught is, first, to have ideas and, secondly, how to relate them to get more coherence. Ideas arise from three main sources: from our perception, from what we are communicated and from the inferential operations that we carry out on them. On the other hand, the ideas communicated may come from any referente frame, even from pure scientific analyses, being thus scientific. From a cognitive point of view, the content of an idea is not what characterizes it as scientific, but the way in which it is processed and the constraints on them. Once we integrate ideas into interpretation processes, they are not properly scientific but work as premises in the interpretation process. Therefore, they can be processed and valued as poetry.

People doing science are able to interpret the world in which they live and their heads are full of ideas; some are coming from their innate processes and some are coming from an explanation. Scientific explanations do not arise from a mental innate process but from something that has to be learnt with effort. Both types of processes are different and make our minds richer in representations and its manipulation. Interpretation is the process that leads to culture, but science is another process with other means.

Human beings are animals that process the information surrounding them in a specific way. In other words, our devices for processing information have evolved and separated us from our neighbours. Thus, it must be possible to follow, in the evolution of human beings, similar (or maybe identical) mental stages to those of animals which are not close to us and to test how other processes have only appeared in our species. Therefore, we are going to try to describe them.

It seems probable that the information processing, which from now onwards we will call primary, it is the most similar one to that of animals close to our species. The characteristic of that type of processing are the following:

1. The data the system uses as input are the objects/events/states of the world potentially encompassed by the senses.
2. The computations that the mental device does with these data are inferential (f. ins. computations such as if X, then Y (which can be entailed in another one of the same kind). Thus,
   a. If there is a lion on my way [if I get close to the lion, then it will try to eat me] then I shouldn’t get close to it.
3. The limits of this device are given by the species. For instance, although possibly the elephant and the human being may both have this kind of device, the data perceived by a human being are different from those perceived by an elephant. On the other hand, in an animal without a mind such as the mosquito, the operations allowing its behaviour are also different.

But our mind evolved little by little separating itself from that of our neighbours. This new kind of processing information was called secondary thinking. Its characteristics are the following:

1 Sometimes it has been called common sense, but I don’t consider this convenient as animals lack common sense in the way human beings understand it.
2 The proverb, The human being is the only animal that stumbles twice on the same stone, shows intuitively that we perceive that we still possess an irrational quality, just as the mosquito which, regardless of the slaps, comes back over and over to sting us. (Vide infra).
The data in this evolved device are not only the ones encompassed by the senses but also those resulting from the inferential processing. That is, the data perceived by our senses are processed together with data invented by our processing system.

The computations are still inferential, so the possible metarepresentations can lead us to represent the world in a quite different way. Thus,

2. If something happens [if there is a tempest then my boat may sink [if my boat sinks then I’ll drown myself]] then [if everything comes from a creator then the tempest has one that I’ll call Neptune] begging him earnestly [if someone begs then sometimes one gets what one wants] I might save myself. Etc.

Here, there are some beliefs that we do not necessarily share; for instance, why are we going to call the creator of tempests Neptune? ie, it seems that

3. The limits in this case are not those from the species, as different human beings may drop different conclusions such as calling Poseidon to the creator of tempests or not believing in god. Thus, limits need to be established; otherwise, any explanation would be valid and nothing would be explained and so, this type of information processing would not do the task that evolution assigned to it (see Damasio, 1994 further below). In this case, the limits are given by the cultural group that imposes them on the others, mainly if that representation of the world benefits them.

This type of processing constitutes the most important feature of our mental activity, because it allows us to explain things (ie., tempests) in relation with others (ie., the notion of a creator of all). Above, we have called this typical information processing interpretation and it is, as aforementioned, the fundamental basis of culture. The more ideas we hold in our head to relate and interpret the world, the more learned we will be. Nearly everything we think about and we share with other beings comes from this way of thinking about the world surrounding us.

However, among certain cultural groups there exists the widespread idea that limit assigning to this kind of processing, instead of being used for the content of the representations to interpret the world (ie., it is not to forbidding that Poseidon is not the creator of tempests, because we believe it is Neptune), should be used for how to realize these interpretations. Then, it arises a possible third way of information processing called tertiary thinking with the following characteristics:

1. The data are those perceivable and non-perceivable by the senses (no one has perceived an subatomic particle or a black hole).

2. The computations, nevertheless, must have serious limitations: they cannot freely relate mental representations to offer interpretations. The only valid computations are those of a cause-and-consequence effect among representations and those acting upon them. Besides, the actual scientific trend requires that those causal relations are exclusively material or mechanical. The results of those computations will acquire the interpretation level of explanations. This is the only thing that scientific thinking tries to achieve: something that requires an enormous control effort on human interpretive abilities.

3. The limits are neither specific. They are still cultural, but as it is enormously difficult to adopt them without being incoherent, the final decision is individual. That is to say, each person will have to individually face the challenge of solving a specific problem, using the scientific explanation or tertiary thinking.

---

3 Some call this process traditional thinking. However, this noun seems to only focus on the limits (of tradition, social group) rather than on the process itself.
Now we are going to study the two strategies proposed for analysing tertiary or scientific thinking. The first one was proposed by Noam Chomsky (1957) for linguistic studies and the second one was proposed by David Marr (1982) for cognition. Both of them can be applied to other scientific fields.

According to Chomsky, for a thought to be scientific it has to be about an object (or event or state) we can clearly point out. In other words, to think in a tertiary way we should achieve the level of observational adequacy. It is not only valid to name something – i.e. Neptune – if we don’t know which reality we are referring to with it. In this case, if and only if we are aware that with that term we point out a mental representation – instead of a supernatural being ruling the sea and tempests -, we will have begun to think scientifically.

A second level that should be reached in this type of thinking is the descriptive adequacy, i.e. the ability to describe an object (or event or state) in a precise way. This level is intuitively necessary to clearly circumscribe our analysis to the reality chosen in the first level. This is what Chomsky has tried to do in his work as a linguist: describing the set of mental representations called language and which are universal to the human species. Nevertheless, descriptions can be carried out in a natural language, being thus subjected to multiple interpretations as all linguistic expressions used to communicate our ideas, or algorithmically, where expressions are exempted from interpretations for having a pre-established meaning and for logically following one another.

The last level is called the explanatory adequacy, which tries to explain how it is possible the appearance of the studied object in the world and its role in it. It is clear that in human sciences the only possible explanation is to find out the material and/or mechanical causes which have produced the phenomenon. As Chomsky himself recognizes, this only can be done from speculative models trying to achieve similar effects to those obtained by the studied object. A good way to start speculating in human sciences could be by asking the following questions:

1. Which are the proximal causes of a mental object, i.e. its neuronal substrates, the cognitons (or cognitive units) constituting it, the environment in which it works, its social influence, etc.?

2. Which is the ontogeny (the individual history) of the mental object?

3. Which is the adaptive function (the original and the actual one in case they differ) of the mental object?

4. Which is the evolving history of the object?

If we want to answer these questions following the tertiary thinking, the effort of explanation is considerable, mainly if we want to follow the material or mechanical causes and effect.

On the other hand, for David Marr, if we try to study a cognitive process, we will have, first, to determine which operations are necessary for that process to reach its goals. The clearest example is that from Hutchins (1996), which deals with the cognitive problems involved in sailing. In sailing, we will have to carry out some cognitive operations that allow us to (1) know where we are, (2) know where we want to go to and (3) perform the necessary actions to achieve that change of place. A set of additional operations would be (4) to know how to determine the speed of movement from the origin to the target. As what is tried is to find out the necessary operations or mental computations, Marr calls this first study stage, computational level.

---

4 An algorithm is explicitly determined in each of its stages so that it is always possible to know which computation is being done. Arithmetic operations are algorithms.
The second level of analysis, the *representational*, would imply how such computations are represented in the minds of people who practise sailing. In western cultures, these computations are represented as a quiet world and a sailing activity which moves across it. However, for Micronesian sailors, for instance, the representation is just the opposite: the world moves and sailing consists in maintaining the ship quiet so that it finally reaches its target. So different representations will need different analyses, although the computations are the same.

The third level is the *implementational* one, in which Marr proposes to analyze any culturally inherited knowledge left in the different objects that human beings have created. That is, in any culture there are physical objects (in sailing, they would be compasses, gears, etc.), which somehow have incorporated mental representations socially shared and which help us to sail more easily.

We will end this section with a thought to avoid misunderstandings. Firstly, the division in the three different types of processing information does not imply any evaluation on the obtained results. As stated above, primary thinking allows us to live and develop in the world in a similar way to animals. Secondary thinking, however, distinguishes us from other living beings. But it must be based on primary thinking (stating or even rejecting it) to have some kind of adaptive value. Believing in Neptune as the creator of tempests is based on the primary thinking that there are things created by a creator; just the same type of belief that assures that every effect has a cause. However, believing that the earth moves around the sun contradicts our primary thinking. From a cognitive point of view, all this knowledge is identical, none is better than the other. The evaluation that each one does of them is again a result of the secondary thinking, typically human. One of the possible evaluations is, as we all know, religions. For them, both Neptune and the Virgin are unanswerable truths and so the limits of their processing are within their traditions.

Another evaluation is the one we have called tertiary thinking. It is thought that by means of constraints and strategies we will achieve truer knowledge than the one based on traditional representations. But it has to be taken into account that this belief is, still, a product of secondary thinking. Priests have just been substituted by scientists, but the type of thinking is the same. However, the findings that each one achieves in their domain using these constraints and limitations of imagination produce a new type of knowledge able to, at least, simulate in (real or mental) models the reality sought to be analyzed. The effort it requires avoids its use in all life domains, but in a limited field of research, so it cannot be considered an evolving step yet. However, it has been observed that societies encouraging its use have a higher state of well-being than those using the secondary thinking. If this is true, then the tertiary thinking could become in the future a further step in our evolution. Regardless of these considerations, the knowledge obtained with this type of thinking is very interesting, as I tried to prove in what follows.

### 2.3. Deductive thinking

Thinking is an activity we are constantly doing. When we are in a traffic jam, we begin to decide which alternative route to take; when we have two job offers, we choose the best one; when we are about to travel, we decide which clothes to include in our suitcase depending on where we are going. Thus, thinking is an activity greatly influenced by the external processing of information, mainly by perception or by the internal recuperation of mnestic systems. Fortunately, most of the times our thinking processes are right and allow us to solve problems. Other times, they offered us partial solutions. Finally, some situations amaze us with irrational answers, away from any logical principle. For instance, when playing the lotto, why most of us

---

5 Some people often imagine that their minds are scientific because they relieve or have faith in the representations that true scientific thinkers have achieved in some fields. Astronomy was not scientific, but just a set of Olympic Gods. It is when the way of thinking about it changes that it becomes a scientific product; thus, any subject can be thought about in a tertiary way, by making the controlling effort required to achieve it.
bet on a combination as 3, 15, 18, 22, 41 and 49, instead of a continuous combination such as 1, 2, 3, 4, 5 and 6? Has it ever occurred to you that both of them are equally possible?

Our thinking acts occupy time and sometimes are slow when compared with those of a computer. As a proof, I suggest the reader to make the following multiplication (or a rough estimation) in less than 10 seconds:

\[ 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = ? \]

Surely, the computer’s computational ability will be superior to yours, both in speed and precision. Within cognitive sciences, a great effort has been laid on whether human beings follow the logical principles when solving tasks, different strategies used in task-solving and decision-making. Therefore, we will begin analyzing human thinking from the perspective of formal logic.

In deductive reasoning, the conclusion follows from the content of the premises; i.e., the conclusion is included within the premises and, as a matter of fact, the conclusion is just an explanation of the content of the premises. So knowing that John Smith is millionaire, deductive reasoning allows us to make the following inferences:

- *John is millionaire*
- *Somebody is millionaire*

Deductive reasoning follows the criterion for deductive validity, according to which an argument is formally valid if and only if the truth of its premises entails the truth of its conclusion (Rips, 1990). It would be self-contradictory to affirm the premises and deny the conclusion. The conclusion is a logical consequence of its premises. An argument that is not valid is said to be invalid. An example of a valid argument is given by the following well-known syllogism:

- *All men are mortal.*
- *Socrates is a man.*
- *Therefore, Socrates is mortal.*

What makes this a valid argument is not that it has true premises and a true conclusion, but the logical necessity of the conclusion. Given the two premises, the argument would just be valid were the premises and the conclusion false. The following argument is of the same logical form but with false premises and a false conclusion, and it is equally valid:

- *All cups are green*
- *Socrates is a cup.*
- *Therefore, Socrates is green.*

In this case, the conclusion does not follow from the premises: a universe is easily imaginable in which *Socrates* is not a man but a woman, so that in fact the above premises would be true but the conclusion false. This possibility makes this argument invalid.

The validity of an argument is a matter of the argument’s logical form:

- All P are Q.
- S is a P.
- Therefore, S is a Q.

These abbreviations make plain the logical form of each argument. At this level, notice that we can talk about any arguments that may take on one or the other of the above two configurations by replacing the letters *P*, *Q* and *S* by appropriate expressions. Of particular
interest is the fact that we may exploit an argument’s form to help discover whether or not the argument from which it has been obtained is or is not valid. To do this, we define an “interpretation” of the argument as an assignment of sets of objects to the upper-case letters in the argument form, and the assignment of a single individual member of a set to the lower-case letters of the argument form. Thus, letting P stand for the set of men, Q stand for the set of mortals, and S stand for Socrates is an interpretation of each of the above arguments. Using this terminology, we may give a formal analogue of the definition of deductive validity:

An argument is **formally valid** if its form is one such that for each interpretation under which the premises are all true also the conclusion is true.

As already seen, the interpretation given above does cause the second argument form to have true premises and false conclusion, hence demonstrating its invalidity.

### 2.3.1. Categorical syllogism

**Formal properties.** A categorical syllogism is the most representative example of deductive reasoning. A **categorical syllogism** is an argument consisting of exactly three categorical propositions (two premises and a conclusion) in which there appear a total of exactly three categorical terms, each of which is used exactly twice. One of those terms must be used as the subject term of the conclusion of the syllogism, and we call it the **minor term** of the syllogism as a whole. The **major term** of the syllogism is whatever is employed as the predicate term of its conclusion. The third term in the syllogism doesn't occur in the conclusion at all, but must be employed in somewhere in each of its premises; hence, we call it the **middle term**.

Since one of the premises of the syllogism must be a categorical proposition that affirms some relation between its middle and major terms, we call that the **major premise** of the syllogism. It contains the predicate of the conclusion. The other premise, which links the middle and minor terms, we call the **minor premise**. It contains the subject of the conclusion.

Consider, for example, the categorical syllogism:

*No geese are felines* (major premise).
*Some birds are geese* (minor premise).

*Therefore, some birds are not felines* (Conclusion).

Clearly, "Some birds are not felines" is the conclusion of this syllogism. The major term of the syllogism is "felines" (the predicate term of its conclusion), so "No geese are felines" (the premise in which “felines” appears) is its major premise. Similarly, the minor term of the syllogism is "birds,” and "Some birds are geese" is its minor premise. "geese" is the middle term of the syllogism.

**Diagramming syllogisms.** Categorical statements present two properties: quantity and polarity. By means of quantity, they can be universal or particular, while polarity will determine whether they are affirmative or negative. Combining those two properties, we will obtain the four types of categorical statements:

<table>
<thead>
<tr>
<th>Formal symbol</th>
<th>Statement</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Universal-affirmative</td>
<td>All A are B.</td>
</tr>
<tr>
<td>E</td>
<td>Universal-negative</td>
<td>No A is B.</td>
</tr>
<tr>
<td>I</td>
<td>Particular-affirmative</td>
<td>Some A are B.</td>
</tr>
<tr>
<td>O</td>
<td>Particular-negative</td>
<td>Some A are not B.</td>
</tr>
</tbody>
</table>
Combining the four types of statements with the premises and the conclusion, we get 64 possible distinct syllogistic forms. A way to visually formalize these syllogisms is by means of the Euler-Venn diagrams:

1. First draw three overlapping circles and label them to represent the major, minor, and middle terms of the syllogism.

2. Next, on this framework, draw the diagrams of both of the syllogism's premises.
   - Always begin with a universal proposition, no matter whether it is the major or the minor premise.
   - Remember that in each case you will be using only two of the circles in each case; ignore the third circle by making sure that your drawing (shading or ×) straddles it.

3. Finally, without drawing anything else, look for the drawing of the conclusion. If the syllogism is valid, then that drawing will already be done.

Since it perfectly models the relationships between classes that are at work in categorical logic, this procedure always provides a demonstration of the validity or invalidity of any categorical syllogism. Consider the following example,

No M is P.
Some Ms are S.
Therefore, Some Ss are not P.

First, we draw and label the three overlapping circles needed to represent all three terms included in the categorical syllogism:

Second, we diagram each of the premises:

Since the major premise is a universal proposition, we may begin with it. The diagram for "No M are P" must shade in the entire area in which the M and P circles overlap. (Notice that we ignore the S circle by shading on both sides of it.)

Now, we add the minor premise to our drawing. The diagram for "Some M are S" puts an × inside the area where the M and S circles overlap. But part of that area (the portion also inside the P circle) has already been shaded, so our × must be placed in the remaining portion.
Third, we stop drawing and merely look at our result. Ignoring the M circle entirely, we need only ask whether the drawing of the conclusion "Some S are not P" has already been drawn.

Remember, that drawing would be like the one at left, in which there is an × in the area inside the S circle but outside the P circle. Does that already appear in the diagram on the right above? Yes, if the premises have been drawn, then the conclusion is already drawn. But this models a significant logical feature of the syllogism itself: if its premises are true, then its conclusion must also be true. Any categorical syllogism of this form is valid.

Here are the diagrams of several other syllogistic forms. In each case, both of the premises have already been drawn in the appropriate way, so if the drawing of the conclusion is already drawn, the syllogism must be valid, and if it is not, the syllogism must be invalid.

**AAA-1** (valid)
All Ms are P.
All Ss are M.
Therefore, All Ss are P.

**AAA-3** (invalid)
All Ms are P.
All Ms are S.
Therefore, All Ss are P.

**OAO-3** (valid)
Some Ms are not P.
All Ms are S.
Therefore, Some Ss are not P.

**EOO-2** (invalid)
No P is M.
Some Ss are not M.
Therefore, Some Ss are not P.
Some Ms are P.
Some Ss are not M.
Therefore, Some Ss are not P.

Human performance in syllogism-solving. Humans seem to make many mistakes when solving these syllogisms due to the atmosphere effect, the semantic content of the premises and the partial or mistaken interpretation of those.

1) Mistakes due to the atmosphere effect: Check whether the following syllogism is deductively valid, i.e. that the conclusion follows from the content of its premises:

- All Cs are B
- All As are B
- All A are C

Some people seem to pay attention exclusively to the kind of utterances, stating that they way in which the premises are formulated creates an atmosphere or context leading their attention. Thus, if one premise is particular, the conclusion will frequently be particular; otherwise it will be universal. If, at least, one premise is negative, the conclusion will be negative; otherwise, it will be affirmative. The aforementioned example is deductively invalid, as if all A are B and all C are B, there should be an identical or similar relationship between A and C. Let us provide it with meaning:

- All cows are mammals.
- All men are mammals.
- All men are cows.

2) Sometimes semantic meaning helps to understand abstract statements, but sometimes it can create some confusion, as it should be distinguished between what is empirically certain and what is formally valid. For instance:

- Some mammals are terrestrial.
- All lions are mammals.
- All lions are terrestrial.

This reasoning isn’t valid because the conclusion is not necessarily deduced from the content of the premises. But it is considered valid by many speakers, as the conclusion is empirically certain.

3) Mistakes and difficulties in premise interpretation or illicit conversion of the premises: The statement All A are B is interpreted by some subjects as All B are A; so that all men are mammals does not necessarily imply that all mammals are men.
2.3.2. Transitive inference or linear syllogistic thinking.

**Formal properties.** Another type of thinking are linear syllogisms in which the subject is asked to establish an order relationship among different objects from the information available in two or more premises. For instance:

- Paul is taller than Mary (A>B)
- Mary is taller than Javier (B>C)

**Who is taller?**

The conclusion is obtained from the content of the premises, i.e. applying the formal principles and given that A<B and B>C, we conclude that A>C (that is, Paul is the tallest). The order relationship involves a quantitative comparison in which each term displays either, more or less, a particular attribute or quality, and the reasoned must draw conclusions based on quantification. The quantitative element is usually a comparative adjective (**taller than**). Premises have also a middle term (in this example, Mary), which allows to establish the nexus or transitive relation between the non-related terms (Paul and Javier). The number of premises in this syllogism is unlimited. The only requirement is that the number of terms must be one more than the number of premises.

**Human performance in syllogism-solving.** These syllogisms have been used to explain visual-spatial models, as subjects solving them must entail an imagery representation, in which terms are hierarchically ordered. Manipulating the difficulty of these syllogisms, some principles have been proposed. But before describing them, try to solve the following problem:

- Phillip is smaller than John (A>B)
- John is smaller than Frank (B>C)

**Who's the tallest? (<?)**

Solving this syllogism, takes longer than the previous one, as the question is formulated inversely and so Frank being the tallest (C<A). Thus, these models propose two mechanisms for solving linear tasks: the flow-chart form of the terms and the extreme items at the marked end of the continuum. The first mechanism establishes a hierarchical representation of the elements, either vertical or horizontally. The second mechanism orders adjectives into bipolar relations (good-bad, tall-small), in which one adjective is the marked one (the contrastive one) and the other one unmarked (the neuter one).

2.3.3. Conditional thinking

**Formal properties.** This is a propositional thinking. A proposition is an utterance that affirms or denies something. We can thus define the truth-value of a claim by saying that the truth-value of a true proposition is "true" and that of a false proposition is "false". **Thus the truth-value of a proposition is its truth or falsity.** We can display the truth-value of propositions by using a technique called truth-tables. A truth-table displays the conditions under which a claim is true. For any given proposition, P:

<table>
<thead>
<tr>
<th>P</th>
<th>T</th>
<th>F</th>
</tr>
</thead>
</table>

This says that P is either true or false. Given that fact, we can rigorously define the relationship between any two propositions as follows:

<table>
<thead>
<tr>
<th>P</th>
<th>Q</th>
<th>T</th>
<th>T</th>
<th>T</th>
<th>F</th>
</tr>
</thead>
</table>
The relations between propositions can be as follows: conjunction, inclusive disjunction, exclusive disjunction, conditional and bi-conditional:

1) **CONJUNCTION**: a compound proposition made up of two simpler claims. These simpler propositions are called conjuncts. A conjunction is true if and only if both of its conjuncts are true. Thus, if either part of a conjunct is false, then the conjunction itself is false. In English, conjunctions are indicated by "and", "but" or "even though". Thus, **John is in class even though Jane is outside** (P and Q).

2) **DISJUNCTION**: a compound made of two simpler propositions, called disjuncts. There are two types of disjunctions, or two ways to interpret a disjunction. An exclusive disjunction is one which says either P or Q, but not both. An inclusive or weak disjunction is one which says that either P or Q, possibly both (and/or is often used here). Logicians are interested in inclusive disjunctions. Thus, for our purposes, a disjunction is true if and only if at least one of its disjuncts is true.

3) **CONDITIONALS** are compound propositions which take the form of "if/then" statements. An example is: **If the train is late, then we will miss our flight.** We all the claim before the "then" but after the "if" the antecedent of the conditional. The claim after the "then" is called the consequent. A conditional does not state either that its antecedent or that its consequent is true; rather, it states that if the antecedent is true, then the consequent is true. Therefore, logicians define conditionals this way: A conditional is false if and only if its antecedent is true and its consequent is false; otherwise, a conditional is true (if P, then Q).

**Deductive inferences in conditional thinking.** One type of mistaken inferences people usually apply during their processing are the following: imagine you are in London and your friend in Glasgow tells you on the phone: **If it rains, I'll go to the cinema.** Next day, you phone him/her and they tell you the film was gorgeous, etc. Thus, you suppose it was raining. Surprisingly, the weather forecast says it was a shinny and warm day. So, was your friend lying you? No, probably it is you acting irrationally.

The table below shows the deductive inferences that can be done with the conditional. **Modus ponens** is usually used in our daily life and it guarantees a truth-value conclusion if the antecedent is true. Something similar happens with **Modus Tollens**, which guarantees the unoccurrence of the antecedent, as the consequence is not occurring. Most people the modus ponens but very few use the modus tollens. Both the affirmative fallacy of the consequent and the negative fallacy of the antecedent show that the conditional is symmetrical, in the sense that **if p then q** is interpreted as **if q then p** and if p then q as **if not p then not q**.

<table>
<thead>
<tr>
<th>Argument</th>
<th>Subject’s conclusion</th>
<th>Formal name</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>If p, then q</td>
<td>Then q</td>
<td>Modus Ponens</td>
<td>If it rains, then I won’t go to the cinema.</td>
</tr>
<tr>
<td>If p, then q not q</td>
<td>Then not p</td>
<td>Modus Tollens</td>
<td>If an intruder is detected by the alarm, the alarm goes off. The alarm does not go off. Therefore, no intruder is detected.</td>
</tr>
<tr>
<td>If p, then q q</td>
<td>Then p</td>
<td>Affirmative fallacy of the consequent (invalid inference)</td>
<td>I go to the cinema, then it rains (the rule doesn’t imply that I’m going to the cinema because it rains).</td>
</tr>
</tbody>
</table>
If $p$, then $q$

<table>
<thead>
<tr>
<th>not $p$</th>
<th>Then not $q$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negative fallacy of the antecedent (invalid inference)</td>
</tr>
<tr>
<td></td>
<td>It doesn't rain, then I'm not going to the cinema (the rule doesn't imply that I can't go to the cinema because it doesn't rain).</td>
</tr>
</tbody>
</table>

When you inferred that it was raining in Glasgow the previous day because your friend told you she/he was in the cinema, then you are committing the affirmative fallacy of the consequent. That is, *if it rains, I’ll go to the cinema* you interpreted as *if I go to the cinema is because it rains.*

**Watson selection task.** Imagine you have a deck of cards with two sides: a blue side and a green side. In the first game, called the Watson Selection Task, every card has a number on one side and a letter on the second side. Each represents a logical pair. We draw four cards, two are laying blue-side-up and two are laying green-side-up. We get “16,” “61,” “P,” and “E”.

```
16 | 61 | P | E
p | no p | q | no q
```

Your task is to indicate which card or cards you would turn up to find if the utterance is true. Surely, you would choose 16 and P (p and q), when the right answer is 16 and E (p and no-q). We would choose 16 because if on the green side there was a different letter from E, the utterance would be immediately falsified. Until now, we have acted according to Modus ponens.

Choosing the card with 61 would mean the negative fallacy of the antecedent, something which is not very common. It would be an invalid inference.

Choosing P is one which creates much confusion, it is chosen by the affirmative fallacy of the consequent: if 16, then P. It would confirm the rule but it wouldn't prove its truth, as if we would turn E and have 16, it would falsify the rule. On the other hand, if when turning P we would find 61, the information would be irrelevant. Then, 16 doesn't provide us with falsifying, so there is no point in turning it.

Finally, E could have 16 or 61. If it were 61, it would be irrelevant; if it were 16, it would be a falsifying example. However, most people turn P and not E in trying to prove the confirmation of the rule rather than its falsity. The results of the test are generalizable. We could ease the task by introducing semantic content in the task, as for instance:

```
I ALWAYS TRAVEL TO PARÍS BY PLANE
```

```
PARÍS    LONDON    PLANE    CAR
```

**Are we rational in our thinking?** People are capable of applying logic principles and make formal inferences correctly. The mistakes made are exclusively due to questions which require abstract thought take longer to complete and are more error-prone. According to Johnson-Laird, people don't think exclusively of pure logical inferences but they build representations of
the premises based on semantic or contextual aspects. The difficulty of an inference lies in both
the difficulty of the processing and understanding of the premises, and in the short-term
memory functioning to decide the validity of the inference. Three stages are implied in the
inference-solving:

1) Interpretation of the premises: subjects first build a model of each of the premises.
2) Integration of the premises: from the previous mental models, subjects create a new
representation of mental model, integrating the information from the second premise
with that from the first one.
3) Validation of the premises: Subject checks that the conclusion is applicable to all the
mental models generated in the previous stage.

2.4. Inductive thinking

Alternatively to deductive thinking, in deductive thinking the conclusion goes further beyond the
premises. The conclusion of an inductive argument is not a question of formal validity but of
probability. For instance, an inductive process takes us to make inferences as the following ones:

Peter is millionaire

Businessmen are millionaire.

Inductive thinking is strong when there is a strong probability of its conclusion being true if its
premises are true (Rips, 1990). In the example, we infer that all businessmen are millionaire,
although there are some in the most absolute misery. Inductive thinking is non-demonstrative,
being rule derivation and event prediction the two main inductive systems. Rule derivation tries
to obtain a rule that explains events. Event prediction tries to predict the outcome of an event
from previous evidence. Thus, inductive thinking goes from the particular to the general
knowledge. If all the students of a class get high marks, we might conclude that all the
students of that school also get high marks. But this conclusion is mainly probabilistic, not
infallible.

The main problem with inductive thinking has been the extrapolations to series and analogies.
In extrapolations, a set of letters or numbers is presented and subject is required to complete
the set from the rule inferred from previous examples:

A B A C A D A F A...

In analogy, a subject is given a relation of two terms A:B and his/her task is to complete
another pair C:D in which that relation can be applied:

Dogs... bark

Cats...

Apart from these classical tasks, we are going to study two research fields where inductive
thinking has been applied: event prediction and categorization.

Event prediction. Predicting an event is a probabilistic question: if a person knows that a coin
is under a bucket, he/she is capable of predicting that the probabilities of obtaining the coin
under that bucket are higher than those of another bucket. However, people predict events
contrary to their probabilities of occurrence, which seems to show that human behavior isn't
always rational and logical. Imagine that you go to a casino and at the roulette table you
observe that an even number has won for the last seven games. Which would be your election
if you were to bet? Even or odd? Most people bet on odd numbers, ignoring that the probabilities of occurrence of both events are identical as they don't depend on the occurrence of previous events. It’s the gambler’s fallacy!, not made by professional gamblers.

This shows that most people ignore the real probability of occurrence of a given event and choose to base on representativity (Tversky and Kahneman, 1983). In one of the many examples, subjects are asked to judge whether a described person is a pilot, seller, scientist or farmer. Subjects’ prediction is based on comparing to what extend that person has the features that they consider characteristic of that profession. Subjects often ignore the objective probability of finding a farmer or a scientist, although the experimenter has previously informed them that 20% are farmers and only 2% are scientists. So introverted, absent-minded, intellectually looking people with glasses are categorized as scientists and not as farmers. So representativity stands for the correspondence or similarity degree between a result and a model. It is true that judgments based on representativity can be right but they often lead to mistakes. Think about wrong predictions we all make: we thought an elegantly dressed person was respectable and he/she was a pickpocket, we thought our football team would win, even though it had lost the last seven matches, etc. So the question is why we use representativity instead of logic, risking making mistakes.

Representativity allows us to make inferences from a set of data. As a matter of fact, the underlying mechanism that accounts for the degree of representativity of an event is an inductive process by means of which we infer general conclusions from a set of particular events: usually all individuals dressed elegantly have proved not to be pickpockets. See that the prediction of events used by the individual is not that far from the logic principles. Some predicting behavior obeys to cognitive acts of inductive thinking, because the knowledge of some particular events allows us to create a general principle of anticipation.

**Categorization.** This is a basic process by means of which organisms try to group outer stimuli into significant classes searching shared features. It is based on two principles: one related to mechanisms of adaptation to the environment (adaptation principle) and the other to inductive thinking mechanisms (induction principle). The diverse stimuli from the environment must be overcome to ease up our adaptation. Thus, we categorize the different stimuli into classes, according to their common characteristics, to treat them in a similar way.

The induction principle arises from the class creation, which implies applying inductive thinking mechanisms. Categorizing is grouping and grouping can only be done going from the particular elements to the general properties, finding out the features or properties shared by exemplars of a given category. Categorization has been studied from a deterministic point of view to then evolve to probabilistic models inspired in natural categories.
TEST 1: ARE WE RATIONAL SOLVING TASKS?

The aim of the following practice is to replicate Wason selection task to prove subjects’ inconsistency when applying logical rules to abstract reasoning. In order to demonstrate subjects’ logical reasoning accurately, two different tasks will be conducted. In the first task, you will apply the original version of the Wason task (or four-card problem) and contrast your results with those obtained by Wason in 1966 (adapted from Cabestrero et al., 2002). In the second task, you will analyze the semantic effect on subjects’ performance, comparing the results with the ones obtained in the first task (adapted from Bye, 2012). It is essential to choose different subjects for each task so as not to contaminate the experiment.

First task

In the first part of this practice, you are required to carry out an experiment and collect a set of data. You will need to test this experiment on 15 subjects, having equal numbers of both males and females, and of foreigners and Spaniards, and following the Wason selection task. Write down their answers. The procedure of this part is as follows:

1. Cut out the four cards in Appendix I. Two of these four cards have letters and the other two have numbers. There is another card with the rule written on it for the subject to think about.
2. Place the subject somewhere where he/she can do the task quietly and without interruptions; then, run the experiment.
3. Place the cards before the subject in the following order: E, D, 4, 7
4. Give the subjects the following instructions:

   We are going to carry out a short task using some cards. Please don’t turn them over. Imagine that each of them has a number on one side and a letter on the other. Now pay attention to the following rule:

   **If a card has a vowel on one side, then it has an even number on the other side.**

   The question is the following: *Which card(s) do you need to turn over in order to determine if the rule is true or false?* Don’t turn over any more cards than are absolutely necessary.

---

6 Both an English and a Spanish version of the instructions are provided for foreign or Spanish students, respectively.
Vamos a realizar una tarea en la que va a trabajar con una serie de tarjetas. No les de la vuelta. Imagine que cada una de las tarjetas tiene un número en una cara y una letra en la otra. Ahora preste atención a la siguiente regla:

**Cada carta que tiene una E en una cara tiene un 4 en la otra.**

Su tarea consiste en decir qué tarjeta o tarjetas giraría para averiguar si esta frase es verdadera o falsa.

Make sure your subjects fully understand the instructions, clarifying any possible doubts they may have without giving them any hints of what the aim or purpose of the experiment is about otherwise you could spoil it/condition/undermine the results. Once your subjects have answered the test, do not tell them whether their answer is right or wrong.

Subjects can take as much time as they want to answer. They must simply say aloud which card(s) they would turn over and you should write their answers down on the record sheet provided in Appendix II, marking with an X the answer each subject provides out of the five possible options (E/ E & 4/E, 4 & 7/ E & 7/ Others).

Once you have noted all the answers, calculate the percentage of subjects giving each type of answer and write them down on the record sheet in Appendix II. Compare the possible similarity of your results with those obtained by Wason (shown in the following table):

<table>
<thead>
<tr>
<th>Type of answer</th>
<th>E</th>
<th>E,4</th>
<th>E,4,7</th>
<th>E,7</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage</td>
<td>42%</td>
<td>59%</td>
<td>9%</td>
<td>5%</td>
<td>13%</td>
</tr>
</tbody>
</table>

**Second task**

The second task is almost identical to the first one, except for the fact that it uses semantic content and thus is a concrete Wason selection task. The cards to be used this time are the ones in Appendix III and IV (English and Spanish version, respectively). The procedure will be exactly the same as in the first task and the instructions subjects should be provided with are the following:
We are going to carry out a task using some cards. Please
don't turn them over. I want you to imagine that you work as
a bouncer at the door of a bar and your job is to make sure
that the law governing alcohol consumption is strictly
enforced. This law states:

**If a person drinks an alcoholic drink, then they must be
over the age of 18 years old.**

The cards below represent the preferred drink and age of
four customers at your bar. Each card represents one
person: one side shows what they're drinking – note the
card on the left shows beer, the alcoholic drink! – the other
side their age.

The question is the following: *which card(s) do you need to
turn over in order to determine whether the rule is being
followed?* Don’t turn over any more cards than are absolutely
necessary.

---

Vamos a realizar una tarea en la que va a trabajar con una
serie de tarjetas. No les de la vuelta. Imagine que trabaja
como guardia de seguridad en un pub. Su labor es hacer que
se cumpla la siguiente ley relacionada con el consumo de
alcohol:

**Si una persona consume una bebida alcohólica debe ser
mayor de 18 años.**

Las tarjetas representan la bebida elegida y la edad de
cuatro clientes del bar en el que trabaja. Cada tarjeta
representa a una persona: una cara muestra lo que están
bebiendo – la tarjeta de la izquierda es cerveza, la bebida
alcohólica – y la otra cara, su edad. Su tarea consiste en decir
*qué tarjeta o tarjetas tendría que girar para determinar si la
regla se está cumpliendo.* No gire más tarjetas de las que sean
precisas.
Make sure subjects understand the instructions, clarifying any possible doubts they may have. Once again it is important not to give them any clues about what the experiment is about; otherwise, you could spoil it. Once subjects have answered the test, do not tell them whether their answer is correct or wrong.

Subjects can take as much time as they want to answer. They must simply say aloud which card(s) they would turn over and you should note their answers down on the record sheet provided in Appendix V, marking with an X the answer each subject provides out of the five possible options (Beer/ Beer & 35/Beer, 35 & 17/ Beer & 17/ Others).

Once you have noted down all the answers, counting each type of answer separately, calculate the percentage of subjects answering each type of answer and write them down on the record sheet in Appendix V. Compare the results of this second task with those of the first task to see if there is an improved performance in the selection of the correct answer in this second task \((p, \neg q; \text{ that is, beer & 17})\) and thus if there is a semantic effect.
APPENDIX I

Cards to present to the subjects in the first task (English and Spanish version):

“E”

“If a card has a vowel on one side, then it has an even number on the other side.”

“D”

“Cada carta que tiene una E en una cara tiene un 4 en la otra.”
APPENDIX II

Register sheet for subjects’ answers in the Wason selection task (classical version, first task):

<table>
<thead>
<tr>
<th>Number of subjects</th>
<th>Types of answer</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E</td>
<td>E, 4</td>
<td>E, 4, 7</td>
<td>E, 7</td>
<td>Others</td>
</tr>
<tr>
<td>Subject 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>Σ</td>
<td>Σ</td>
<td>Σ</td>
<td>Σ</td>
<td>Σ</td>
</tr>
<tr>
<td>Percentage</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
</tbody>
</table>

There are five possible types of answer. The first four [E], [E, 4], [E, 4, 7] and [E, 7] are the most frequent answers you will obtain. Use the category [Others] when the subject provides an answer which does not exactly correspond to any of the other four, for instance [E, D, 4]. In order to register the answer, mark the corresponding answer with an X. Then, add the answers obtained in each category in the frequency row. And finally, calculate the percentages of each type of answer dividing the number of answers in each category or column by 15 and multiplying it by 100, as in the following example:

$$\frac{(8 \div 15) \times 100}{100} = 53.33$$
APPENDIX III

Cards to present to the subjects in the second task (English version):

Beer

Coke

“If a person drinks an alcoholic drink, then they must be over the age of 18 years old.”

35

17
APPENDIX IV

Cards to present to the subjects in the second task (Spanish version):

Cerveza

Coca-cola

“Si una persona consume una bebida alcohólica debe ser mayor de 18 años.”

35

17
APPENDIX V

Register sheet for subjects’ answers in the Wason selection task (semantic content version, second task):

<table>
<thead>
<tr>
<th>Number of subjects</th>
<th>Types of answer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beer</td>
</tr>
<tr>
<td>Subject 1</td>
<td></td>
</tr>
<tr>
<td>Subject 2</td>
<td></td>
</tr>
<tr>
<td>Subject 3</td>
<td></td>
</tr>
<tr>
<td>Subject 4</td>
<td></td>
</tr>
<tr>
<td>Subject 5</td>
<td></td>
</tr>
<tr>
<td>Subject 6</td>
<td></td>
</tr>
<tr>
<td>Subject 7</td>
<td></td>
</tr>
<tr>
<td>Subject 8</td>
<td></td>
</tr>
<tr>
<td>Subject 9</td>
<td></td>
</tr>
<tr>
<td>Subject 10</td>
<td></td>
</tr>
<tr>
<td>Subject 11</td>
<td></td>
</tr>
<tr>
<td>Subject 12</td>
<td></td>
</tr>
<tr>
<td>Subject 13</td>
<td></td>
</tr>
<tr>
<td>Subject 14</td>
<td></td>
</tr>
<tr>
<td>Subject 15</td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>Σ</td>
</tr>
<tr>
<td>Percentage</td>
<td>%</td>
</tr>
</tbody>
</table>

There are five possible types of answer. The first four [E], [E, 4], [E, 4, 7] and [E, 7] are the most frequent answers you will obtain. Use the category [Others] when the subject provides an answer which does not exactly correspond to any of the other four, for instance [E, D, 4]. In order to register the answer, mark the corresponding answer with an X. Then, add the answers obtained in each category in the frequency row. And finally, calculate the percentages of each type of answer dividing the number of answers in each category or column by 15 and multiplying it by 100, as in the following example:

\[(8 \div 15) \times 100 = 53.33\]
3. Cognitivism

3.1. Introduction: Cognition and mind

By the term cognition, we refer to some phenomenon derived from the functioning of that unobservable entity called mind. We are going to study with the mind and its functioning, although there may be discrepancy on what topics to deal with when talking about cognition. Some people will argue that several arithmetic tasks, such as communicating with other people by means of language or remembering a phone number, require given cognitive abilities. However, other people may consider cognition as experiences related to the mind, such as predicting an event beforehand, telepathy or similar phenomena. Apparently, we only use minimal part of our mind capacity and thus, we cannot predict the lotto numbers or communicate with long distance friends without a phone or the internet. Unfortunately, we are not going to be able to satisfy the curiosity of this last group of people; we are not going to study such phenomena in this course, as they are not part of the scientific thinking.

Human cognition means studying mental activities from a systematic point of view, analyzing how science has understood the mind, how knowledge is manipulated in memory and which cognitive aspects are involved in, for instance, reading and understanding a text, as you are doing now, or in the multiplication $445 \times 7$. Juan Luis Arsuaga, a paleontologist, has insisted in some of his works that human beings were not the elected species by nature to impose upon the rest of the species. The development of their cognitive capacities simply caused that a group of primates evolved to a human society thanks to the fact that in a given moment environmental conditions favored the symbolic capacity, which progressively developed ended in the ability to communicate with other beings by means of an articulated language (Arsuaga y Martínez, 2001). Thus, our cognitive system has some specific properties that distinguish it from other species.

What does the study of human cognition comprise? According to Brook, Andrew & Robert. J. Stainton (2000), Knowledge and Mind. A Philosophical Introduction, Bradford Books, MIT Press, basic topics are the following:

1. What is knowledge?
2. What can we know?
3. How do we acquire knowledge?

The answers, in general, are the following:

1. Knowledge is something that follows, at least, the two (or maybe three) following conditions:
   a) It is true (if my knowledge is about ghosts, as they do not exist, I do not have any knowledge about them)
   b) It is believed (if I do not believe that Paris is the capital of France, although it is, I do not have any knowledge about which is the capital of France)
   c) Some philosophical theories think that knowledge is acquired by means of a competent authority.

2. There are two kinds of skepticism (based on the two first aforementioned conditions):
   a) Absolute: human beings cannot know anything about the world.
   b) Relative: certain aspects of the world are impossible to know (for instance, morality or what we attribute to other minds, etc.)

3. In this case, we have a continuum between the empiricists, who consider knowledge as derived from sensory experience, and rationalists, who think knowledge is not sensory but intellectual and deductive.

According to Brook & Stainton, another part of the philosophy, the philosophy of mind, deals with:
a) What is the mind? (what one thinks, according to the tradition).
b) How does something incorporeal like the mind relate to something material like our body? There is an apparent relationship: when we are drunk or drugged, our body sometimes gets sick and our mind works in an abnormal way. This basic point has given birth to two different theories: the dualist and monist, as previously mentioned.

### 3.2. Summary of the ontological theories:

1) Two basic questions:

   A) What ENTITIES does a scientific theory explain?
   
   B) How do those entities RELATE to the ones other scientific theories study?

2) Analysis of ideas:

   A) In natural sciences, strongly influenced by MATERIALISM, there is the possibility of collaboration among sciences.

   1) **Materialism:** Everything that exists is material and follows the rules of physicalism: all things are composed of *material* and all phenomena (including consciousness) are the result of material interactions. In other words, matter is the only substance.

      a) Materialism belongs to **MONIST ONTOLOGY:** the universe is really just one thing, despite its many appearances and diversities.

      b) Materialism **opposes to**

      1. **MONISTIC Idealism**

         a. For **Johann Gottlieb Fichte** (1762-1814), the I posits (i.e.\(I = I\)). The 'I' must set itself in order to be an 'I' at all. The expansion of the individual \(I\) into a universal \(I\) made it difficult to distinguish between the \(I\) and the world.

         b. **Georg W.F. Hegel** (1770-1831) was a monist, like Parmenides and, above all, Spinoza.

            1) Hegel believed that everything is one. All concepts belong to an Absolute Idea, all spirit is part of the Absolute Knowledge.

               a) He considered monism a search.

               b) He understood monism as preserving the difference in the identity, multiplicity in the unity.

            2) He was also an idealist:

               a) Thinking is not a mental game but the world. The world is the development and manifestation of thinking itself.
b) "Spirit" became a matter of global culture, or "World-Spirit" — i.e.,
groups of individuals and events interacting with themselves and
others at a world-historical level.
c) But the most relevant thing is to think reality: think the thoughts
which constitute the Ultimate Reality. That is to say, his idealism
is the result of a process.

2. ONTOLOGICAL PLURALISM (vitalism, emergentism, Cartesian dualism, etc.).

a. Ever since biology made vitalism obsolete, pluralism is only valid in human
   and social sciences.

b. Its best argument is our ignorance: psychological and social ignorance.

c. All true discoveries are framed within the ontological materialism.

B) The ontological border between natural sciences, and human and social sciences:

1) Defended by humanists and sociologists.

2) But, naturalist scientists only believe in ontological materialism.

   a) Astronomy, chemistry, physics, geology and biology have acquired an important
      combination of logical coherence, causal description, explicative power and
      falsifiability, which has transformed them in satisfactory and accurate examples of
      what knowledge can be.

   b) Several disciplines have integrated themselves into this unlimited system
      of interconnected knowledge, as a way to acquire knowledge and progress.

   c) The birth of computers and of modern cognitive science completed the unification
      of the material and mental world, showing how physical systems can
      incorporate information and meaning.

   d) Artificial networks emerged a few decades ago, but these systems can simulate
      many cognitive processes such as reason, memory, knowledge, choice, problem
      solving, prediction and language. Until then the mind had been considered as a
      metaphysical object, completely different form the physical world, and thus human
      beings were animals completely disconnected from the causal network connecting
      the rest of the universe.

   e) These developments have taken beings, mental phenomena and human
      beings -three disconnected domains from the scientific world – to the scientific
      world of causality.

C) Materialistic monism in human sciences:

1) The theoretical distinction between maximalist and minimalist materialism is classic.

2) Both types of materialism AGREE in that every particular mental state or process is a
   particular physical process. That is, for a mental token (exemplar) there is always a physical
   token.
a) For **maximalists**, every mental state or process is a kind of physical state or process. For instance, face recognition is a well-defined mental process. A maximalist will argue that, if this is the case, there is at least a well-defined type of neurological process and that both types are just one.

b) The **minimalist**, on the other hand, will consider face recognition as a function to be conducted by a material device, though it can be carried out by very different material devices (such as time can be indicated by sun watches, quartz watches, that is, by heterogeneous physical mechanisms).

An excellent analogy of the level Independence is the relation of the *software* with the *hardware* of the computers. Both are independent but for a computer to work we have to establish materialistic links between the physical implementation of the machine and the program. That is to say, a physical computer is needed to implement programs in it. A WORD program does not work in a microwave or in a fridge.

**FINAL OUTLINE:**

![Ontology](image.png)

**Vitalism** (Pluralism: several substances)

**Emergentism** (Vitalism: (...) Cartesian Dualism)

**Dualism** (Idealism: Fichte, Hegel)

**Materialism** (Physicalism)

**Monism** (one substance)

**Maximalist** (everything is reducible to physical matter: individual processes and analysis categories: connectionism)

**Minimalist** (All individual processes are possible because there is a material basis necessary to be explicit at some point, but the analysis categories are different and irreducible to physical matters).

### 3.3. Historical introduction to cognitivism

#### 3.3.1. First revolution: behaviorism and mind

*Behaviorism* was born in 1913 with a paper from J. Watson “Psychology as a behaviorist views it”. In this
paper, Watson claimed for a psychology focusing on control and behavior prediction; by behavior he understood the observable and manifest events, i.e. what the organism says or does.

Behaviorism was a reaction against mental structuralism, because it substituted mental analysis for behavior analysis. It has four basic rules:

1. Influenced by Darwin, it assumes an evolving continuum among species.
2. It is positivist, that is, understands behavior as an observable, quantifiable and falsified fact.
3. It is mechanist because it defends that explanatory principles for simple and complex behaviors are similar.
4. It is associationist because behaviors must be explained in terms of associations between elements. Such elements are stimuli (physical events from the environment) and (motor) answers, being the S-A (stimulus-answer) the only valid method of study to explain behavior. S is what is perceived and after some specific neuronal connections, it would elicit a specific A.

From the 1930s on, Watson’s behaviorism gave rise to Hull’s neo-behaviorism, which has other motivational aspects, apart from the learning mechanisms, to explain behavior. Hull distinguished between learning and performance. Thus, neo-behaviorism proposes the model S-O(organism)-A, as between the stimulus and the answer it is necessary to consider internal mechanisms, whose existence can be inferred from behavior observation and experimental manipulation. This model is hypothetical deductive, because of its enormous explanatory power with learning processes and behavior performance.

Tolman was the neo-behaviorist who emphasized most on the cognitive aspects without leaving associationism. He claimed that cognition was part of behavior, as organism behavior is goal-driven, interacts with the environment and chooses the best alternatives for problem solving. Thus, he talked about purposive behavior, goals, intentions and cognitive maps.

Nevertheless, the most famous behaviorist was B. F. Skinner. His idea of behavior was controversial, especially in the cognition field. According to Skinner, the organism should be understood as rejecting all mentalism and describes behavior in terms of environment. He rejects the hypothetical deductive method and proposes instead an inductive methodology based on the experimental analysis of behavior; thus, changes in behavior are the result of an individual's response to events (stimuli) that occur in the environment.

Behaviorism’s most important contribution has been the introduction of a scientific methodology.

### 3.3.2. Second revolution: cognition and mind

Behaviorism is considered like the prehistory of cognitivism. In the 1930s and 40s, there was a famous symposium celebrated in 1948 by the California Institute of Technology, in which the psychologist Karl Lashley presented a paper against behaviorist theory. Lashley criticized the online S-A model, arguing that, if human behavior is considered complex, then it should be explained by hierarchical principles. Lashley critique showed the difficulty of behaviorism to explain human cognition.

In the 1950s, the development experienced by disciplines such as mathematics (with the amazing works of Norbert Wiener and John von Neumann in the USA and Alan Turing in the UK), or neurobiology (with research on mental incapability such as aphasia, agnosia, etc.) constituted a source of knowledge which provided many models. Psychologists such Bruner and Millar and linguists like Chomsky began to ask whether it was possible to apply those models to the study of human cognition. In this context, the Second World War began and it required an enormous technological effort and development with the aim of deciphering secret messages or calculating ballistic tables, but as a consequence it brought about the design and construction of the first computers performing intelligent tasks. This was the beginning of a new way pointing at the creation of thinking machines that could perform similar tasks to those of the
The birth of cognitive science has an accurate date: the 11th of September 1956 in a symposium on information theory, organized by the Massachusetts Institute of Technology. In it, Noam Chomsky surprised the Academy with his communication “Three Models of Language” in which he proved that the mathematical theory of Shannon and Weaver could not be applied to the description of the natural language; however, his cognitive model of language, universally known transformational generative grammar, could solve the challenge of achieving a serious scientific approach. Let us analyze the most important ideas of Chomsky’s theory for cognitive science:

1. Green ideas sleep furiously.
2. Green sleep colorless furiously ideas.

According to Chomsky, both (1) and (2) contain identical words and have no apparent meaning; thus they are not relevant. However, our intuition suggests that (1) is more right than (2) because (1) has a correct syntactical structure. Thus, Chomsky assumes the existence of innate, formal, abstract knowledge in the mind, independent from the semantic meaning. Such knowledge is a set of internal rules that the speaker applies intuitively to distinguish between grammatically correct and agrammatically sentences. Chomsky’s grammar is generative because human cognitive system can produce grammatically correct sentences, from some internalized, innate rules, that can be decomposed in elements such as subject, verb, etc.; it is transformational because it defends the existence of a set of rules that allow the abstract representation of sentences that can relate to each other or transform themselves into different ones:

3. John kicked Mike.
4. Mike is kicked by John.

In a transformational grammar, there are a deep and a surface structure. Let us suppose the following example, in which both sentences have the same surface structure but different deep structure:

5. John is easy to please.
6. John is eager to please.

In (5) the subject receives the action, whereas in (6) the subject performs the action. Sometimes the transformation of a deep structure to a surface one can generate ambiguities:

7. The marks of the teacher are high

The study of the formal aspects in language analysis was opposed to Skinner’s strict behaviorist approach. Chomsky criticizes Skinner’s book Verbal Behavior because of its environmentalism and associationism and highlights as a distinctive feature of speakers creativity, as the speaker is not just a passive receptor of stimuli that provides answers but their cognitive capacity allows them to generate sentences they have never heard before. This capacity argues in favor of the internalized rules, sustained in the LAD (Language Acquisition Device).

Chomsky’s assumptions emphasized the possibility of analyzing cognition from a formal point of view, in line with AI (Artificial Intelligence) purposes of creating computer programs that simulated human cognitive activity. Software programs were and are formal languages with accurate instructions to perform a task. Besides, Chomsky’s assumptions admit a new mentalism: the mind has abstract knowledge, rules, that can be studied without empirical data. This favors the idea of a computer simulation of cognitive processes and of a computer with mental states. J. Fodor, a philosopher, applies Chomsky’s epistemological ideas to the human cognitive system and argues that abstract innate knowledge must be represented in a mentalésé or language of thought, as a set of rules that determine knowledge manipulation and generate cognitive processes such as language, perception, memory or reasoning in problem solving. This approach is rationalism as the rules of the mind (res cogitans) are
different from the rules of the res extensa (physical substrate of the mind).

The development of the first computers, together with the rationalist view insisting in theory formalization, gave rise to the idea of the mind as an intelligent system performing computational tasks similar to those of a computer. This was the beginning of the mind as a computational system manipulating symbols, like software.

3.4. The study of human cognition: representation and computation

Cognition is the general name of the operations performed by mental mechanisms when processing information. However, many researchers distinguish two types of mental operations: the cognitive and the precognitive. In the former, processes are (semi-) conscious; whereas the latter would be automatic operations with no possibilities of becoming conscious. Cognitive science tries to describe and explain both types of operations by models that show the development of the unconscious process. For instance, Noam Chomsky has proposed theoretical models that explain scientifically the generation of encoded messages in a human language.

The basic elements of cognitive models are of two kinds: representations and computations. The first ones have to do with the forms and content of the information processed. Computations refer to the operations performed with the representations to get a result. There are two problems: it is not clear what a representation is neither the part of the brain where it is located. But the idea in all models is that it is something structured with some content. There is also the problem that there is not any clear and explicit theory able to cope with the unclear aspects of the linguistic content. However, formal computational systems are very developed and its functioning is clearly understood. So computations are in most models formal: they work according to the structures, not to the contents.

Cognition can be modeled according to three description levels: the computational level, the representation level and the implementation level. In the first one, the description just establishes which is the output information, which the result and which operations to apply to the output to transform it into the result. It is a general level of analysis. The second level describes the representations each individual belonging to a group uses to algorithmically perform these operations. It is the cultural level of analysis. The third level tries to describe how the computations of the first level and the representation of the second are physically performed in the world.

There are nowadays two big cognitive approaches: the ones studying the mental macro-processes of information (language use, reasoning, etc.), phenomenologically conceptualized; and the ones describing the micro-processes which work in parallel and which allow simulating complex operations. The first approach is the classic cognitivism while the second is the connectionism.

TEST 2: THE GAME OF LIFE
The **GAME OF LIFE**, invented by John Horton Conway, is an example of a cellular automaton. The universe of Game of Life is an infinite two-dimensional orthogonal grid of square cells, each of which is in one of two possible states, **live** [+] or **dead** [-]. Every cell interacts with it eight neighbors: cells directly horizontally, vertically (N-S y E-W) or diagonally adjacent (NE-SW y SE-NW):

<table>
<thead>
<tr>
<th></th>
<th>NE</th>
<th>N</th>
<th>NW</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td></td>
<td>cell</td>
<td>W</td>
</tr>
<tr>
<td>SE</td>
<td>S</td>
<td>SW</td>
<td></td>
</tr>
</tbody>
</table>

Time in the world is discrete, not continuous. Births and deaths happen simultaneously and the discrete moment at which this happens is called a **tick**. In other words, each generation is a pure function of the one before. At each step in time (ie. tick), the following transitions occur:

1) Any live cell with fewer than two live neighbors dies, as if caused by underpopulation.

2) Any live cell with more than three live neighbors dies, as if by overcrowding.

3) Any live cell with two or three live neighbors lives onto the next generation (survival).

4) Any dead cell with exactly three live neighbors becomes a live cell (birth).

(Both overcrowding (more than 3 live neighbors) and underpopulation (less than 2 live neighbors) lead to extinction).

Let’s study some cases: (1) In the following configuration, only cells **D** and **F** have each three live neighbors [+] . Thus, they will be the only ones living onto the next generation. Cells **B** and **H** have just one neighbor [+], so they will die. Cell **E** has two neighbors, so it survives [+]:

A) PRESENT MOMENT:

<table>
<thead>
<tr>
<th></th>
<th>A-</th>
<th>B+</th>
<th>C-</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-</td>
<td>E+</td>
<td>F-</td>
<td></td>
</tr>
<tr>
<td>G-</td>
<td>H+</td>
<td>I-</td>
<td></td>
</tr>
</tbody>
</table>

B) NEXT TICK, IE. NEXT GENERATION:
Obviously, this configuration will go backwards in the next tick and so successively and indefinitely until new cells are born. This configuration is called blinker and belongs to the oscillator class of objects. They usually change from generation to generation, but eventually repeat themselves after two generations.

2) What happens in this other configuration?

Nothing. Each cell [+] has three live neighbors [+] and no dead cell [- or in white] has three neighbors [+], so it remains as it is: no live cells die and no new cells are born. This configuration is called block and belongs to still life objects.

The most important thing is that applying the aforementioned rules, the configuration of the following generation can be exactly predicted. In other words, if we apply the rules to the Game of Life, we have the possibility of predicting without error the products. Everything is clear and visible in the world of the Game of Life. The following configurations are more interesting than the previous ones. Let us consider the following diagonal pattern:
FIRST EXERCISE:

What makes these configurations interesting is relations among them. Thus, we have gliders, eaters, the queen bee shuttles, etc. Let’s see the simplest: the glider:
SECOND EXERCISE: Try to make the configurations of four generations to test how it glides or moves. It is exciting!

THIRD EXERCISE: What happens when a glider and an eater meet in the world? Try four generations:
4. Knowledge representation in the mind

It is fine to define what it means to have a mind. It is fine to assert that representations can be manipulated in a process called thought, but the necessary condition for its functioning is the amazing capacity that humans have to meta-represent representations in deeper levels. Thus, it is not the same to represent ourselves that the floor is hard and not liquid; but we can also think that it is so and manipulate that information result as we want, if we re-introduce it in another representation and so ad infinitum. From now on, when we imagine that the human being is able to make things impossible for other species, we should know that we are referring to the meta-representation ability. Hence, ART, HUMOUR, RELIGION, etc. that distinguish us from other species are, in the first place, attitudes (which is how some meta-representations are called, such as the ironic or artistic attitude).

4.1. Processable elements and their functioning

4.1.1. A possible causal history of representations

We all have an intuitive idea of what a representation is; morphologically re seems to indicate that something is present again. It is much more difficult to explicit, in causal terms, the relation between what is represented and its representation.

It seems logic to suppose that, as the result of evolution, the representation ability has to have its basis on elements shared with other animals. We could say that the simplest organisms (i.e., the bacteria) react against the environment: their sensors are joined to their organic structure and react directly without intermediate representations. In human beings, this kind of automatic reactions (for instance, close the eyes when something unknown comes to our eyes) also exist. Those are not real representations, although some scholars consider them representations preprogramed in the genotype of the species by the evolving history. More complex organisms have perceptive mechanisms on which external stimuli produce some effects, once they have been transformed into internal stimuli around a detector which joins them into a representation according to certain guidelines which must be scientifically described. Representations are not copies of the thing represented, so the set of internal stimuli necessary and sufficient to create a given representation should also be studied. For instance, if our external stimuli come from a dangerous predator of the organism, then there is no direct relation between the external and the internal stimuli; if there were, it would take much time to process information and the result would be fatal for the individual. This is an outline of the food representation [FOOD]:

\[
\begin{align*}
A & \rightarrow a \\
B & \rightarrow b \\
C & \rightarrow c \\
D & \rightarrow d \\
E & \rightarrow e \\
F & \rightarrow f \\
N & \rightarrow n
\end{align*}
\]

Detector

[FOOD]

There are two types of relation: the one among all external and internal stimuli, and the one of some internal stimuli (in this case, \(a\), \(d\), \(e\) and \(n\)) with the detector. The detector is responsible for the adequate representation.

As we analyze more complex organisms, the organic relation is more sophisticated with the following elements:

1. Specific movement perceptors for food.

\(7\) Whenever I use capital letters and brackets, I will refer to a representation or concept.
2. The detector **represents** [FLY] (the aforementioned mechanism is now just one element in the more complex organism)

3. Motivator, similar to detector, but **representing** [HUNGER ] + [FLY]

4. Motor program that **represents** the action to perform for [HUNT] + [FLY]

5. Motor possibility to act in the world according to that program.

In this later case, there are two types of semantic relations: the causal interaction between the individual environment and their mental representations by means of their perceptors, having as a result the representation [FLY]. This relation constituted by steps (1) and (2) is an exosemantic relation. Secondly, we have the inferential interaction between the representations stored in memory and the exosemantic relations that the perceptor mechanism obtains. As those are internal relations of the organism, they are called endosemantic.

In the case of the human beings, everything is more complex, as we will see in the next paragraphs.

### 4.1.2. Levels of representational accessibility

Our causal description of the origin of some representations refers to the idea of the relation between them and the individuals. That is to say, with a representation like [FLY]) I can adopt a positive or negative attitude: {HUNGER FOR [FLY]} o {NO HUNGER FOR [FLY]}. This relation between the subject and the representation is called attitude. Stephen Schiffer uses the metaphor of mental boxes to describe the levels of accessibility.

Let us suppose that there are some storing boxes for representations. The most important ones are the wish and belief boxes, like in this case {HERE THERE IS ALWAYS [FLY]}8. Then there is belief when there is a representation in the mental box of the beliefs and desire. We will accept this metaphor right now to focus, on the one hand, on the way representations are stored and recover and, on the other, to distinguish between two kinds of representations.

Sperber (1997) relieves that there are two kinds of representations in human minds, each with a different cognitive both in mental inference operations and in physical behaviour operations:

1. **Intuitive representations**: This type of representations are directly stored in the desire or belief boxes, as we saw in the fly example. In human beings (and maybe in some of the nearest species), it does not need to be always directly represented in the box, but it is an spontaneous process effect of dealing with beliefs which actually are represented. For instance, the impossibility of Cervantes driving is one of these intuitive beliefs. Intuitive beliefs can be used as premises in inferential operations carried out by the mind, or even in inferences conducting to action. Sperber calls them intuitive because we do not need to think about them, we think with them. It is not species-specifically human.

2. **Reflective representations**: It is formed by those representations meta-represented in other intuitive representations. In human beings, it acquires such an important development that it changes the way in which information is processed. These representations give rise to secondary and even tertiary thinking. This processing type is species-specifically human. For instance, we can hold the following representations unthinkable for any other being.

   1. {THE POPE SAYS THAT [THE MOTHER OF JESUS WAS VIRGEN MARIA]}

---

8 Maybe there is the need to postulate some more boxes such as hope or frighten.
2. {CATHOLICS BELIEF THAT [THE MOTHER OF JESUS WAS VIRGEN MARIA]}

3. {THE IDEA THAT [THE MOTHER OF JESUS WAS VIRGEN MARIA] IS ABSURD}

The representations in brackets are meta-represented in other representations directly stored in the belief box, which makes it somehow valid. Thus, (1) and (2) are valid for people who consider certain what the Pope says or for Catholics. (3) is certain for those people who focus on their experiences and observation. Notice that this last representation could have been directly stored in the belief box, working as intuitive. Only when we think about it, it becomes reflexive. This is one of the most important effects of the meta-representational ability: that we can think about our thoughts. Imagine how boring your life would be if you could not think about cognition!

The origin of intuitive beliefs is spontaneous. It arises from sensory information and from automatic inferences about that sensory data, although it can also be spontaneously formed from some type of communication instead of experience, i.e. when all concepts in that communicative exchange are intuitive.

Reflexive beliefs are not spontaneous. They are created either by reasoning or by deriving them from an authority. Therefore, they acquire the strength of the meta-representational proposition. This strength can be almost absolute, when talking about scientific data; very strong if based on a recognized authority or well-built hypotheses; very possible or doubtful. Reflexive beliefs contain intuitive concepts, because they are the ones we think with, but also reflexive concepts which have been consciously introduced, specifying their inferential role and meaning, although it is not always so.

One of the advantages of this meta-representational ability is to allow us to have beliefs about concepts we do not understand. Millions of people relieve in a divine trinity that nobody has been able to fully understand. The advantage of this human ability is that it allows us to understand certain realities, as we are able to believe many things we do not fully understand but that will be gradually understood. So human learning is partly possible thanks to this type of mental operations. When we come across (half-) understood concepts, we either convince ourselves that their thoughts are very deep or we admire them as opere aperte.

For instance, a Zen maxim is not fully understandable and is open to several (maybe infinite) interpretations:

4. [satori] is contemplation without contemplating anything, wisdom without wisdom, which is practiced without practicing anything. Therefore, it is not that either.

Finally, the poem by Ezra Pound is considered, above all, artistic, at least by his editor and I suppose that by many other people:

5. In a Station of the Metro
The apparition of these faces in the crowd;
Petals on a wet, black bough.

This ability allows us to meta-represent abstract or incomprehensible things which become comprehensible as we are able to relate them with other representations of the same abstract or specific kind. This is what will be studied in the following paragraphs:

4.1.3. Types of human representations

There are two kinds of basic representations, as Rivière (1986) mentions: implicit representations (usually called images, not only the visual ones) and explicit representations (representations in which their structural relations have to be explicitly mentioned, i.e. propositions).
To see this distinction, let us consider an example: the following figure (visual image) is a public representation of a donkey, whose constituent parts (ears, eyes, paws, etc.) are related in a similar way to how they should be in reality. Therefore, it is not necessary to express them explicitly. They are directly represented and implicitly perceived as part of the drawing.

However, if I want to explicitly represent how this figure is, I would have to explicit how its parts relate to one another: the donkey is sitting, its ears are open forming, its eye is too big compared to the rest of the face, etc. Thus, these propositions are explicit and are the first step before verbal communication. The explicit representation would not be possible if we did not have our metarepresentational ability, as we could not think that pensar que {THE WORD DONKEY CORRESPONDS TO [THE OBJECT DONKEY]}. However, this metarepresentational ability is not always easy. Think, for instance, of the smell of your boyfriend/ girlfriend/ husband/ wife, etc. We remember it but it is very difficult to transform that smell into a proposition. Only a creative poet with a linguistic imagery ability could create a public representation that vaguely suggests that smell.

Often the processor device of our mind works with different types of representations, creating systems that require some limitations to avoid pseudo-solutions in subsistence problems. When mentioning the three types of human thinking, we talked about the limits imposed by the group and the individual ones. Let us elaborate on this point.

Within propositional representations, we could distinguish between interpretative, descriptive and defining representations. The interpretive representations are those representing nothing, they are not justified by an object or event. They have sense because they relate to others semantically, not causally. For instance, the [GNOME] representation is semantically similar to the [ELF] and [DWARF] one, although they are different types of meta-representations: {I believe there are [magic beings (like the gnomes)]}. Evidently, there is no real object that has to do with this representation. Cognitive psychology tries to study the mental representation of [GNOME], not as a referential content, but as something causal, analyzing why it arises and how it is maintained in the human minds of a specific culture. In those cases, the important thing is to find out what kind of interpretations is allowed with such representations. For instance, if the members of an X culture believe in djins, we can interpret this belief by relating our interpretive representations about magic beings with those new characters. It seems clear that part of our interpretation of others’ beliefs is by means of these representations, without it meaning an ontological entity to our representations. That is, interpreting that [DJINS] are spirits does not imply that the djins really exist.

Another kind of representations has a referent to which they are alike. For instance, our concept of [GAME] refers to, at least, a reality: activities of human beings; i.e. the representation [GAME] refers, in non-figurative cases, to a kind of activity. But in this case, the characteristics that describe the activity are not enough nor necessary. A descriptive characteristic of that activity would be that is a competitive activity, although lonely people are

---

9 This distinction belongs to Sperber (1996: 27-37).
not. So we have representations referring to something and so describing an object or event, but not expliciting it fully. If the definition of interpretive was sense similarity, the descriptive are similar to something real, at least.

Some representations, a few compared with the two previous, not only describe a reality by means of its similarity with it, but they are that reality and, as such, we can access the necessary and sufficient conditions defining them. They are the defining representations: the true representations in the etymological sense of the word and we can causally track them from their origin, describe them and define them. The concept [MOTHER], for instance, is defined by [+FEMENINE] and [+PROGENITOR] in a sufficient and necessary way.

Tertiary thinking is the process by which we, first, try to transform interpretive concepts into descriptive ones and then, to determine as accurately as possible which conditions enable their existence in the world to represent ontologically material existing entities.

4.2. Evolving problems: intentionality

The American philosopher, Daniel Dennett, in his work *Consciousness Explained*, reminds us that Hellen Keller, the deaf and blind child, who, thanks to her teacher, learnt to talk and read, wrote the following autobiographical datum:

> Before my teacher came to me, I did not know that I am. I lived in a world that was a no-world. I cannot hope to describe adequately that unconscious, yet conscious time of nothingness.... Since I had no power of thought, I did not compare one mental state with another. (Keller (1908), en Dennett, 1991: 227)

Keller’s description of its not knowing that it was her, of that world/ no-world, of that unconsciousness yet consciousness, etc. is an exercise of heterophenomenology 10, which could be the world without consciousness of the animals. The question is: what is the radical different between animals and human beings? The answer to this question is surely our linguistic ability (without taking into account the metaphors about animal languages11, i.e. bee language, dolphin one, etc.)

When human language became perceptible and representations were observable, the meta-representation arose, as our meta-representational ability is closely linked to the talking ability in a human language. Daniel Dennett argues in favor of this conception. The linguistic actualization of a language in a communicative act directs the concatenation of representations and its structure. In his own words,

> It is not that first goes into a higher-order state of self-observation, creating a higher-order thought, so that one can then report the higher-order thought by expressing the higher-order thought. It is rather that the second-order state (the better-informed state) comes to be created by the very process of framing the report. [...] The emergence of the expression is precisely what creates or fixes the content of the higher-order thought expressed. [...] The higher-order state literally depends on – causally depends on – the expression of the speech act. (Dennett, 1991: 315)

That is, the meta-representational possibility is, first, a linguistic possibility, which consequently is reflected on the structures of the mind.

However, there is a basic philosophical problem with this conception, as Sperber (2000) points out. We could think that

---

10 According to Dennett (1991: 442), this word would indicate our imaginary projection of the consciousness of animals when talking about how Bambi enjoyed the colors of the sunset, although deers are blind to chromatic changes.

11 As stated before, it is possible that some animals do have a mental structure of representations of the world, so they would have *language*. However, the metaphors I mention refer to the human linguistic tool that we use to help us with communication: language.
[...] Either ancestral linguistic communication, though presumably simpler in many respects, was based on the same kind of communicative mechanism as modern linguistic communication. If so, it presupposed meta-representational ability, and therefore could not precede it. Or else, ancestral linguistic communication was strictly a coding-decoding affair, like other forms of non-human animal communication, and then there is no reason to assume that our ancestors had the resources to become aware of the representational character of their signals, anymore than bees or vervet monkeys do. (Sperber, 2000: 4)

Thus, although the communicative behaviors of those animals are undoubtedly public, perceptible and potentially observable representations, there is no possibility of updating that potentiality in the minds of such animals. Besides, it does not seem very likely that the ability to attribute mental states to other animals and the decoding capacity are inter-related in the rest of the species as they are in the human beings. The most appropriate is to relieve that stimuli decoding set animals in a cognitive, emotional or motivational state appropriate for the situation (Sperber, ibid.).

Contrary to what Dennett thinks, the mutation allowing meta-representation has to take place before the configuration of the mentalese and then it reflected on language. This possibility is more in line with the modular approach of the human mind, which, according to Cosmides, is formed by the following aspects:

[..] Did the human mind evolve to resemble a single general-purpose computer with a few or no intrinsic content-dependent programs (Gould, 1979)? Or does its evolved architecture more closely resemble an intricate network of functionally dedicated computers, each activated by different classes of content or problem, with some more general-purpose computers embedded in the architecture as well (e.g. Chomsky, 1975, 1980; Cosmides & Tooby, 1987; Gallistel, 1990; Gazzaniga, 1985; Rozin, 1976; Symons, 1987)? (Tooby & Cosmides, 1992: 94).

That is, the meta-representational possibilities which, in a first moment, arose in the mental devices of representation creation by an evident mutation, which separated us from animals settled down in the module associated to those devices that allow us to transform them into public representations. A cause for this mutation was surely survival. Individuals capable of predicting the intentions of the surrounding human beings could escape before a rival could show an openly aggressive behavior. Hence, attributing (or meta-representing) mental states has an obvious advantage, compared to those that only reacted to danger by processing general or encoded stimuli. Its origin had, thus, a purely biological function.

(1º) What is the explicative power of a psychological/social phenomenon? and (2º) what is the extent of interpretive representations?\(^{12}\)

Firstly, the description of the effects of a psychological or social phenomenon can only be explained if we are conscious that mentioning those effects will never be able to explain the phenomenon existence. Besides, we need to show its development or continuity.

Secondly, we must be aware that functionalist theories do not always allow us to find principles to identify the social or psychological phenomena. In most of the cases, this phenomenon identification is realized in an interpretive way, as the subject’s judgments are part of the previously established or discovered function.

Ruth Millikan (1984) establishes for the evolving history some general adaptations. The basic idea deals with the evolution of behavioral or corporal elements in the selective mechanical process. This author sees distinguishes between the different types of FUNCTIONS in mutations:

All adaptations have a function. However, sometimes these functions are collateral and can only be considered non-fundamental accidents in the evolution of adaptation; that is, they do not directly relate to survival or reproduction. So, for instance, trying to attribute mental states to other people can lead us to earn money, as it happens with psychologists. This function—earning money—is not basic to explain the adaptation development. This is a collateral function.

Regarding basic functions, they deal with survival and reproduction. Within basic functions, Millikan distinguishes between direct basic function, which are the direct product of mutation and obtain a vital advantage (for instance, chameleons’ change of color) and derived basic function, derived from the basic functions for a specific goal (for instance, if we put the chameleon on a blue surface, it turns blue). The basic biological function of this meta-representational seems to be to predict danger before there is a real threatening. The derived basic function would be the one each subject does by means of that adaptation. Imagine that Peter runs away when he sees Charles going to his house with an enormous stick. However, John asks his cousin to protect him when Paul tries to kick him for flirting with his girlfriend. In each situation, danger and the way we deal with it are different. But let us suppose that one of these derived functions allows Margaret not only to run away but to find out something Dave wants to show her by doing weird movements with his index finger and grumbling. Margaret performs the following mental operation:

1. Margaret sees Dave doing finger signs and grumbling.
2. This [Dave doing finger signs and grumbling] is not a normal behavior of Dave’s.
3. This [Dave doing finger signs and grumbling] is not a normal behavior of Dave’s because he wants something.
4. Margaret looks around to see if she finds it out.
5. In the distance, she sees Dirk, her enemy.
6. Dave by doing finger signs and grumbling wanted to show me the danger.

In this narration, the derived basic function of Margaret’s meta-representational device is not to escape from danger, but to find out what Dave wants with his unusual behavior. With time, this derived basic function will become the direct basic function of another problem: communication among human beings. By finding out speakers’ intentions, hearers are able to interpret the signals that speakers produced in a given situation. In this way, a chain of possible mutations going from direct basic to derived basic are established. The distinction between direct-basic and derived-basic explains respectively (1) how a given characteristic stabilizes by ancestors’ results and (2) how a given characteristic (not derived from any ancestral model) realizes an indirect basic function.

As Sperber y Wilson (1986/1995) shown, two aspects should be taken into account in human communication: (1) the basic, which is the attempt to inform somebody of something, and (2) the behavior with which we try to inform somebody that we try to communicate that information. Sperber and Wilson call (1) informative intention and (2) communicative intention. Let us represent this relation visually:
COMMUNICATIVE INTENTION: Somebody has the intention to inform about sth

INFORMATIVE INTENTION: The intention to inform about something

How many meta-representations can humans do? Is it possible to have different meta-representational levels in non-human animal communication? According to Noh,

(...) se puede alcanzar un cierto nivel comunicativo en cuanto la intención informativa sea reconocida, incluso si la capacidad de meta-representar la intención comunicativa de orden superior no se da (Noh, 2000:3)

For instance, bees encode certain genetically pre-established movements, the information about the situation of some food resources, or the tail movement of a dog means that it is happy. But it seems that they do not do it intentionally, but in an automatic way. It is a very basic and perfectly encoded information processing.

Human beings show these characteristics in several aspects when interchanging information, but their processes are much more complex than encoding and decoding. The human being uses its code to form an expression, with which it can send different, even contradictory, messages, for instance:

1. James has always being very careful with his health.

If we meta-represent (1) in the following proposition:

2. Mary says ironically that it will communicate the message that James does not take care of himself.

In some cases, depending on the speaker’s ATTITUDE, messages can acquire an ironic, sarcastic tone and express just the opposite of what they should mean according to the sentence’s linguistic decoding. In general, human attitudes (such as irony) are the ability to meta-represent representations (linguistically decoded meanings) to provide them with a sense appropriate for that attitude.

To understand a message, according to Sperber & Wilson (1995):

1. The speaker has to show their COMMUNICATIVE INTENTION by means of any behavior (talking is just one of them).

2. Once the hearer perceives that act, they prepare to interpret the message.

3. Once the speaker knows the hearer is prepared (i.e., they pay attention), they communicate their INFORMATIVE INTENTION by means of some act (pronouncing the words in an understandable language, etc.)

4. Then the hearer interprets the message (1º) according to the speaker’s attitude of making manifest an encoded expression, and (2º) according to the decoding that the hearer does about that expression.

Decoding is not the most important part of communication among human beings. It can be omitted but be communication (for instance, if I show the door with my index finger and put an annoyed expression, it can be easily interpreted that I am asking someone to leave the room). The most important part is interpreting speaker’s informative intention, what Grice called mind reading. It is amazing the ability we have to interpret this informative intention from gestures,
face expressions, voice tones, etc.

In a simple linguistic exchange there are two meta-representations:

{WHEN BEGINNING TO TALK MARK TRIES TO COMMUNICATE THAT [in this moment Ann tries to inform by saying that (James has always been very careful with his health)]}.

**TEST Nº 3: INTENTION**

**A.** Imagine the following situation:

Ann is finishing an economic report that she has to hand in the following day. As it usually happens, the phone, which is in the lounge, rings at that very moment. Ann knows that Miguel is in the lounge, watching a football match. Ann shouts desperately:

\[ \text{ANA: Phone!!} \]
\[ \text{MIGUEL: I'm in the toilette!} \]

What is happening in this communication exchange? Which are the informative and the communicative intentions of each of the speakers?

**B.** Can you explain what happens in this communicative exchange, in terms of the informative and the communicative intention, between a saleswoman, called Margot, and Vincent, a client:

\[ \text{MARGARITA: Fruit!} \]
\[ \text{VICENTE: It was called Trinity.} \]
5. Intelligent systems: computers and mind

5.1. Introduction

Up to now we have dealt with interpretation problems, but it is time to point at the way in which minds work. Nowadays, the two systems producing results or intelligent behaviors are the computer and the brain. These two big approaches deal with what is known as artificial intelligence (AI). The distinction between computer and neural network is very important because the contemporary study of the mind is very much influenced by both.

The computer, as an intelligent system, receives an input, encodes it in its own language, manipulates it according to some programmed instructions and provides an output in a comprehensible language. A computer is based on a binary encoding (1 and 0 which physically correspond to the presence or absence of electricity). Between the input and output the computer only performs a task following the instructions of the program or software. So the computer just processes the information we provide it to.

The brain, on the contrary, is an analogical system, constituted by billions of cells, the neurons. Each of them receives inputs from other neurons, integrates them and generates an output which sends to other neurons or effector organs such as the muscles. Information Exchange among neurons is done by liberating neurotransmitters. Neurons send their signals through the axon. Neurotransmitters affecting the neuron’s dendrite can be excitatory or inhibitory: the first ones increase the probability of the neuron synapses and the second ones diminish that possibility. So for a neuron to synapse the inputs must reach the unit’s threshold value.

5.2. Mind architectures

Both computers and the brain are used as inspiration sources to elaborate theoretical models to explain the cognitive functioning. These models are represented in formal abstract languages, like an architect presents its building. By its similarity, in cognitive science we talk about mental architectures to refer to the design specifications of cognitive models.

Contemporary cognitivism has used two conceptual analogies as explicative metaphors to elaborate models and theories:

1) Some researchers have inspired in online computers to explain their models. So just as a computer distinguishes among a logical level, the program or software and the physical substrate or hardware, the mind is understood as a software which manipulates symbolic information and that is implemented in brain acting as a hardware. This is the computer metaphor or analogy and the elaborated models are symbolic or classic architectures of the mind (they are called classic because they were chronologically the first ones).

2) Other researchers focus on the brain to develop explicative models of the mind. According to this perspective, the mind is understood as a computational system described by models of artificial neuronal networks (ANNs), in which information is distributively processed among many units analog to the neurons. This is the brain metaphor or analogy of the mind and constitutes the basis of the connectionist architectures of the mind.

When using metaphors as conceptual analogs, it is important to discriminate between the physical and the functional level. The physical or implementation level is constituted by the specific components of a system: chips and circuits in the computer and neurons in the brain. The most important part is the functional or logical level, the operative way that an intelligent system adopts, independently from the physical substrate. In other words, neither classic cognitivism nor connectionism aimed at proposing explanatory theories of the physical functioning of a computer or a brain, but to elaborate models of the mind functioning.
5.3. Symbolic architectures

5.3.1. The origins of the symbolic cognitivism: Alan Turing

Two theoretical lines in 1930s and 1940s gave rise to the symbolic cognitivism: the theory of computation and cybernetics.

The theory of computation

Biography:

He was born in London in 1912. He was an exceptional mathematician. When he was 20, he went to Cambridge University. The relativity theory was out of fashion and what apparently was in vogue was the quantum mechanics and the mathematics logic.

For two millennia, Maths has evolved in an uncontrolled way. Although it was the most objective and evolved scientific object of human mankind, nobody knew if it was possible that its results were wrong. The Greek discovered some paradoxes such as the Epimenides, according to which a proposition was denied and asserted at the same time. Euclides tried to formalize its rules in a work called *Elements*. Later on, René Descartes, Immanuel Kant, Frank Boole, Gottlob Frege and Giuseppe Peano tried to do the same with statistics and infinitesimal calculus but without success. At the beginning of the 20th century, Bertrand Russell and Alfred North Whitehead tried to re-elaborate *all* mathematics in some principles called *theory of types*. But the work was so complex that nobody was convinced.

David Hilbert elaborated the *Hilbert Program*, which was a list of the big problems without solving\(^{13}\), among which there was *completion*. The question was whether this discipline was coherent and complete, and whether any proposition could be derived from its postulates. Hilbert thought so. In 1931, Kurt Gödel solved the problem by showing that Mathematics were incomplete and rejecting the Hilbert Program. His theorem stated that *Any axiomatic formulation of theory of the numbers includes nondecidable propositions*. In sum, Gödel proved that in any system, science, language and mind there are assertions that are true but that cannot be falsified.

Turing was fascinated by all those questions and, above all, with the idea that a machine obeying rules could write poetry, produce mathematical truths or play chess. The most important thing for Turing was that Gödel did not say clearly, if there is an axiomatic system and we have any of its proposition, whether it could be mechanically determined if that proposition was *non-decidable* in that system. That is, he did not answer Hilbert’s third question.

Turing’s machine

Turing worked on that problem and one day while running along the Cam, he solved the problem. Turing thought of machine and considered very plausible Brewster paradox of *the body being a machine*. Turing was interested in a machine’s lack of decision but he realized that due to Gödel no machine could be built that infallible recognized the non-decidable propositions of a system.

\(^{13}\) The three key questions were the following:

1) Is Mathematics *complete* (in the technical sense that any mathematical expression can be falsified)?
2) Is Mathematics *consistent*?
3) Is Mathematics *decidable* (is there a definite method that can be applied to any mathematical expression and guarantees the right decision about the truth of that expression)?
He first specified the general notion of what a machine was, taking the typing machines as an example. He asked himself what it meant to call *mechanical* to that machine. It means to any answer to an action of the operator is right. One can predict the machine’s behavior but the answer would always depend on the real **configuration** of the machine. So he decided to have internal configurations and a variable position in a typing line, although the action of the machine would not depend on its position.

The typing machine was just too limited to be used as an example. It manipulates symbols but it can only write them and needs an operator to choose the symbols and to change them configurations and positions at each moment. Turing asked himself: *how would a more general machine manipulating symbols be?* It would have to have a finite number of configurations and an exactly determined behavior in each of them. But he imagined much more and described the following machine:

a) It a machine that writes in one line.

b) There is an unlimited paper role, divided in several discrete units so that only one symbol can be written in each of them. Thus, the functioning of the machine is defined in a finite way but with an unlimited space to act.

c) The writing point of the machine can be to the right or to the left at any moment.

d) The machine could scan the symbol on the box.

e) It would be able to write and erase symbols.

**Description of the Turing test**

Turing wrote a classic paper called: "Computing machinery and intelligence", in which he proposed the **simulation game** the **Turing test**.

![Figure 1](image)

As shown in the figure, in two different rooms there are a man and a woman; in another separate room is the person who is going to play and connected to the two other rooms by some kind of device. The game consists in that the person has to guess in which room the man is by the answers provided by both the man and the woman. The woman will answer as a woman and the man will have to try to cheat ting and answer as a woman. Eventually, the person will find out where the man is. But let us suppose that a man knows feminine psychology so well that he cheats the person; in this case we could say that he *knows* women’s mental functioning in a way that he simulates their reactions perfectly.
Turing’s idea was to substitute the man with a machine, in the simulation game. In this case, the player would have to find out where the machine. That person who could achieve that the machine worked out perfectly, simulating the mind of a human being, would have understood the functioning of this kind of mind.

The **cognitivism change of paradigm** claims that created models will help to see whether the invented mechanism has some kind of reality in the brain and vice verse: whether models are functionally valid for human nature. In this way, we will be able to know how the brain works.

**Cybernetics**

There are two basic ideas that cybernetics contributes to cognitive systems: 1) The *feedback*, for an organism to be able to auto-regulate itself it has to receive information about its behavior. 2) *Computation*, which is the calculus each system does to modify its states.

**5.3.2. Fundamental concepts of the computational theory**

Turing’s idea helped to develop computer sciences, although it was John von Neumann with his work, *The Computer and the Brain* (1958), the father of computers. He created an architecture with a CPU operating according to a binary logic (1, 0) and with storing possibilities.

Classic cognitivism understands the mind, just like a computer, which manipulates information by means of symbols. A *symbol* is a representation of a given object or (abstract/specific) knowledge. One of the biggest developments of the scientific world has been the discovery that by *manipulating forms*, contents can be related by means of the appropriate operations. By reading these lines, you are manipulating the 28 symbols of the alphabet. In sum, symbol is the way in which knowledge is represented in the mind.

Besides the symbol, classic cognitivism understands that the mind operates on symbolic information by means of *computations*. Any algorithm has three features:

1. **Neutrum substratum**: for instance, a division (a type of algothythm) can be done on a paper or a blackboard, with a pen or pencil. Its power is due to its *logical structure* rather than its causal qualities.
2. **Unconscious mechanization**: although the result may be brilliant, each of the constituent steps is absolutely simple.
3. **Guaranteed results**: whatever the algorithm does, it does it always. Thus, in each step we know exactly the result we have obtained.
The software of a computer needs a mechanism that applies the instructions of the program. This mechanism is the CPU (Central Processing Unit). In conclusion, classic cognitivism considers that the mind works as the software of a computer, because it is system of manipulating symbols, which applies a set of cognitive operations on mental representations containing information upon a set of rules.

A symbolic model can be analyzed from three different levels: the semantic, the symbolic and the physical one. The semantic or knowledge level tries to explain why intelligent systems show given performances. Human beings have a set of beliefs and goals that drive their performance in a given moment. In the symbolic level the knowledge from the previous level is formally represented by means of an algorhythm that implies symbol manipulation. In the human being, this level formally explains why some performances require more time or errors than others. Symbolic architectures assume that knowledge is stored as abstract mental representations, which can be described by symbols:

A: John is tall
B: Mary is tall
C: A – [taller than] – B

The physical level is the material support with its own functioning on which the intelligent system resides. In the case of a calculator, it is the memory, its circuits and the display panel; in the human case, it is the brain.

5.3.3. Online information processing

The term information processing (IP) began to be used by the end of the 1950s and comes from AI, where it was usually used to refer to the computer simulation of cognitive processes (Newell y Simon, 1972). IP theories are cognitive theories, but the other way around is not try. From AI, the work by Newell, Shaw and Simon (1958) constitutes the precedent of IP systems. In a classical work of 1972, Newell and Simon described the general structure of an IP:

![Diagram of IP system]

As shown in this figure, the environmental inputs are caught by the receptors which send information to a central processor. This has an interpreter to identify each signal and to associate it with the corresponding letter. The interpreter recognizes HELP using the symbolic information stored in the memory. Besides, the system can elicit an answer to the environment.

Basic idea of an IP

An IP system requires two elements: knowledge (information) and manipulation (processing). The basic idea is that the external informative input is received by an intelligent system (human or machine). Some rules are applied on that information, which is manipulated. Once the operations have been performed, an output is provided.

Theoretical assumptions of online IP models
Palmer and Kimchi (1986) consider that a cognitive model which wants to become an IP should follow these four postulates:

1. **Information description**: mental events can be described as *information events* with 3 basic elements: an input, an operation to be applied to that input and an output. A mental event can include any type of internal (conscious or unconscious) experience.

2. **Recursive decomposition**: Recursive decomposition implies mapping the input and the output. Any information event can be decomposed in other low-level events, until a final level formed by the *primitive elements*. Processing activity linked to human memory can be decomposed into some components: iconic memory, echoic memory, short-term memory and long-term memory.

3. **Flow continuity**: it implies that the informative input of an operation is the output of the precedent one.

4. **Flow dynamics**: it has to do with the temporality; i.e. an operation cannot elicit an output until it has received the corresponding input and processed it.

5. **Physical implementation**: when a physical system (brain or machine) pretends to be described in terms of information processing, such system will adopt a set of representations upon which some operations will be applied to transform them.

### 5.3.4. AI and computer simulation

AI, as a discipline, was born in 1956 in a summer seminar in Darmouth College with researchers such as Newell, McCarthy, Simon or Minsky. They showed interest in developing computers that could simulate human behaviors, usually classified as *intelligent*.

The AI idea is that a machine carries out an intelligent task as fast and best as possible. In the cognition field, the computer program serves as explicative model of human performance and thus, serves to support or reject a cognitive theory. Simulation is obliged to help in the elaboration of a cognitive theory on human intelligence. To contrast a given theory by simulation, there are several steps: 1) the theory is deduced from a model which is a formal interpretation of the theory; 2) to transcribe that model into a computer program; 3) to check theory predictions in both the simulation and experiments; 4) to contrast the intermediate proceedings as the obtained results in the simulation experiment. If a simulation is valid, it offers two criteria:

1. **Identity in the final result**: the program must produce the same result as human beings.

2. **Proceeding equivalence**: the program must perform the task in the same number of steps and time as the subject, and reproduce them.

### 5.3.5. Can machines think?

Can machine experience mental states similar to those of humans – consciousness, desires, emotions, intentionality? For many scientists, this idea does not longer belong to science fiction, although this question has given rise to two different positions in AI.

**Weak AI**

Most researchers understand that the Turing Test does not presuppose that the machine has mental states in the human sense, when it simulates. Simulation programs are simple formal analogies of cognitive functioning, but do not assume that the computer can have mental status in the sense of having internal experiences, self-consciousness or intentionality. According to this approach, our identity is preserved from the machines.

**Strong AI**
Other group of researchers understands mind and its states as a consequence of the particular brain functional organization. The dualism body/mind is a unimportant discussion, as mental states arise because the brain has a particular functional organization. Therefore, any system with the same functional organization as the brain can have identical mental states. The brain is a hardware in which there is software, the mind, which can be understood as a functional system. Thus, the mind, like software, can be implemented in any physical system. A conveniently programmed computer reproducing the mind’s organization functioning can have mental states in the most strict sense of the word, i.e., conscious experiences, intentionality, etc.

5.3.6. The computational paradox and the intentionality problem

Authors like Gardner (1985) consider that the computer analogy has given place to a computational paradox, as it has shown two limiting aspects in which human beings are different from computers: 1) it is not possible to implement in machines principles driving thought and social behavior, which determine many subjects’ behaviors when discriminating between right and wrong, as machines never make mistakes in executing the program; those mistakes are attributable to human beings; 2) there is a difference in the artificial processor and the human hardware, because the circuits and the chips cannot reflect the evolving complex processes present in the neurons.

Another difficulty is the intentionality problem. Intentionality refers to mental states (beliefs, expectations, desires, ...). Mental states are intentional in the sense of aboutness, they are about something, contrary to the physical effects referring to nothing. The intentionality problem is that, when operating on mental symbolic representations, those must be interpreted and referred to something to provide at each moment the most appropriate answer, i.e. provide them with meaning. The big difference between machines and minds is that the computer lacks intentionality; that is, it does not understand nor operate according to beliefs, expectations, desires, etc. but just according to the executed program.

Throughout these three decades, it has been observed that human performances are not adequately explained by means of these online models, inspired in a step by step logic. According to many researchers, the relevant thing is to search for analogies between machines in parallel and the mind, without ignoring the way the brain computes information. This has been precisely the goal of connectionist architectures, which will be studied in the next unit.
TEST 4: THE TURING TEST

A Coffeehouse Conversation

Douglas Hofstadter

PARTICIPANTS

Chris, a physics student;
Pat, a biology student;
and
Sandy, a philosophy student.

Chris: Sandy, I want to thank you for suggesting that I read Alan Turing's article "Computing Machinery and Intelligence." It's a wonderful piece and it certainly made me think -and think about my thinking.

Sandy: Glad to hear it. Are you still as much of a skeptic about artificial intelligence as you used to be?

Chris: You've got me wrong. I'm not against artificial intelligence; I think it's wonderful stuff -perhaps a little crazy, but why not? I simply am convinced that you AI advocates have far underestimated the human mind, and that there are things a computer will never, ever be able to do. For instance, can you imagine a computer writing a Proust novel? The richness of imagination, the complexity of the characters...

Sandy: Rome wasn't built in a day!

Chris: In the article Turing comes through as an interesting person. Is he still alive?

Sandy: No, he died back in 1954, at just forty-one. He'd only be sixty-seven this year, although he is now such a legendary figure it seems strange to imagine him still alive today.

Chris: How did he die?

Sandy: Almost certainly suicide. He was homosexual and had to deal with a lot of harsh treatment and stupidity from the outside world. In the end, it apparently got to be too much, and he killed himself.

Chris: That's a sad story.

Sandy: Yes, it certainly is. What saddens me is that he never got to see the amazing progress in computing machinery and theory that has taken place.

Pat: Hey, are you going to clue me in as to what this Turing article is about?

Sandy: It is really about two things. One is the question "Can a machine think?" -or rather, "Will a machine ever think?" The way Turing answers this question (he thinks the answer is "yes," by the way) is by batting down a series of objections to the idea, one after another. The other point he tries to make is that the question is not meaningful as it stands. It's too full of emotional connotations. Many people are upset by the suggestion that people are machines, or that machines might think. Turing tries to defuse the question by casting it in less emotional terms. For instance, what do you think, Pat of the idea of "thinking machines"?

PAT: Frankly, I find the term confusing. You know what confuses me? It's those ads in the newspapers and on TV that talk about "products that think" or "intelligent ovens" or whatever. I just don't know how seriously to take them.

SANDY: I know the kind of ads you mean, and I think they confuse a lot of people. On the one hand we're given the refrain "Computers are really dumb, you have to spell everything out for them in complete detail," and on the other hand we're bombarded with advertising hype about "smart products."

CHRIS: That's certainly true. Did you know that one computer terminal manufacturer has even taken to calling its products "dumb terminals" in order to stand out from the crowd?

SANDY: That's cute, but it just plays along with the trend toward obfuscation. The term "electronic brain" always comes to my mind when I'm thinking about this. Many people swallow it completely, while others reject it out of hand. Few have the patience to sort out the issues and decide how much of it makes sense.

PAT: Does Turing suggest some way of resolving it, some sort of IQ test for machines?

SANDY: That would be interesting, but no machine could yet come close to taking an IQ test. Instead, Turing proposes a test that theoretically could be applied to any machine to determine whether it can think or not.

PAT: Does the test give a clear-cut yes or no answer? I'd be skeptical if it claimed to.

SANDY: No, it doesn't. In a way, that's one of its advantages. It shows how the borderline is quite fuzzy and how subtle the whole question is.

PAT: So, as is usual in philosophy, it's all just a question of words.

SANDY: Maybe, but they're emotionally charged words, and so it's important, it seems to me, to explore the issues and try to map out the meanings of the crucial words. The issues are fundamental to our concept of ourselves, so we shouldn't just sweep them under the rug.

PAT: So tell me how Turing's test works.

SANDY: The idea is based on what he calls the Imitation Game. In this game a man and a woman go into separate rooms and can be interrogated by a third very sex-oriented questions? I can imagine the man, overeager to act convincing, giving away the game by answering some very blunt questions that most women would find too personal to answer, even through an anonymous computer connection.

SANDY: It sounds plausible.

CHRIS: Another possibility would be to probe for knowledge of minute aspects of traditional sex-role differences, by asking about such things as dress sizes and so on. The psychology of the Imitation Game could get pretty subtle. I suppose it would make a difference if the interrogator were a woman or a man. Don't you think that a woman could spot some telltale differences more quickly than a man could?

PAT: If so, maybe that's how to tell a man from a woman.

SANDY: Hmm ... that's a new twist! In any case, I don't know if this original version of the Imitation Game has ever been seriously tried out, despite the fact that it would be
relatively easy to do with modern computer terminals. I have to admit, though, that I'm not sure what it would prove, whichever way it turned out.

**PAT:** I was wondering about that. What would it prove if the interrogator (say, a woman) couldn't tell correctly which person was the woman? It certainly wouldn't prove that the man was a woman!

**SANDY:** Exactly! What I find funny is that although I fundamentally believe in the Turing test, I'm not sure what the point is of the Imitation Game, on which it's founded!

**CHRIS:** I'm not any happier with the Turing test as a test for "thinking machines" than I am with the Imitation Game as a test for femininity.

**PAT:** From your statements I gather that the Turing test is a kind of extension of the Imitation Game, only involving a machine and a person in separate rooms.

**SANDY:** That's the idea. The machine tries its hardest to convince the interrogator that it is the human being, while the human tries to make it clear that he or she is not a computer.

**PAT:** Except for your loaded phrase "the machine tries," this sounds very interesting. But how do you know that this test will get at the essence of thinking? Maybe it's testing for the wrong things. Maybe, just to take a random illustration, someone would feel that a machine was able to think only if it could dance so well that you couldn't tell it was a machine. Or someone else could suggest some other characteristic. What's so sacred about being able to fool people by typing at them?

**SANDY:** I don't see how you can say such a thing. I've heard that objection before, but frankly it baffles me. So what if the machine can't tap-dance or drop a rock on your toe? If it can discourse intelligently on any subject you want, then it has shown it can think -to me, at least! As I see it, Turing has drawn, in one clean stroke, a clear division between thinking and other aspects of being human.

**PAT:** Now you're the baffling one. If one couldn't conclude anything from a man's ability to win at the Imitation Game, how could one conclude anything from a machine's ability to win at the Turing game?

**CHRIS:** Good question.

**SANDY:** It seems to me that you could conclude something from a man's win in the Imitation Game. You wouldn't conclude he was a woman, but you could certainly say he had good insights into the feminine mentality (if there is such a thing). Now, if a computer could fool someone into thinking it was a person, I guess you'd have to say something similar about it -that it had good insights into what it's like to be human, into "the human condition" (whatever that is).

**PAT:** Maybe, but that isn't necessarily equivalent to thinking, is it? It seems to me that passing the Turing test would merely prove that some machine or other could do a very good job of simulating thought.

**CHRIS:** I couldn't agree more with Pat. We all know that fancy computer programs exist today for simulating all sorts of complex phenomena. In physics, for instance, we simulate the behaviour of particles, atoms, solids, liquids, gases, galaxies, and so on. But nobody confuses any of those simulations with the real thing!

**SANDY:** In his book *Brainstorms*, the philosopher Daniel Dennett makes a similar point about simulated hurricanes.
**CHRIS**: That's a nice example too. Obviously, what goes on inside a computer when it's simulating a hurricane is not a hurricane, for the machine's memory doesn't get torn to bits by 200-mile-an-hour winds, the floor of the machine room doesn't get flooded with rainwater, and so on.

**SANDY**: Oh, come on! That's not a fair argument. In the first place, the programmers don't claim the simulation really is a hurricane. It's merely a simulation of certain aspects of a hurricane. But in the second place, you're pulling a fast one when you imply that there are no downpours or 200-mile-an-hour winds in a simulated hurricane. To us there aren't any—but if the program were incredibly detailed, it could include simulated people on the ground who would experience the wind and the rain just as we do when a hurricane hits. In their minds (or, if you prefer, in their simulated minds) the hurricane would not be a simulation but a genuine phenomenon complete with drenching and devastation.

**CHRIS**: Oh, boy—what a science-fiction scenario! Now we're talking about simulating whole populations, not just a single mind!

**SANDY**: Well, look! I'm simply trying to show you why your argument that a simulated McCoy isn't the real McCoy is fallacious. It depends on the tacit assumption that any old observer of the simulated phenomenon is equally able to assess what's going on. But, in fact, it may take an observer with a special vantage point to recognize what is going on. In this case, it takes special "computational glasses" to see the rain and the winds and so on.

**PAT**: "Computational glasses"? I don't know what you're talking about!

**SANDY**: I mean that to see the winds and the wetness of the hurricane, you have to be able to look at it in the proper way.

**CHRIS**: No, no, no! A simulated hurricane isn't wet! no matter how much it might seem wet to simulated people, it won't ever be genuinely wet! And no computer will ever get torn apart in the process of simulating winds!

**SANDY**: Certainly not, but you're confusing levels. The laws of physics don't get torn apart by real hurricanes either. In the case of the simulated hurricane, if you go peering at the computer's memory expecting to find broken wires and so forth, you'll be disappointed. But look at the proper level. Look into the structures that are coded for in the memory. You'll see that some abstract links have been broken, some values of variables radically changed, and so forth. There's your flood, your devastation—real, only a little concealed, a little hard to detect.

**CHRIS**: I'm sorry, I just can't buy that. You're insisting that I look for a new kind of devastation, a kind never before associated with hurricanes. Using this idea, you could call anything a hurricane as long as its effects, seen through your special "glasses," could be called "floods and devastation."

**SANDY**: Right—you've got it exactly! You recognize a hurricane by its effects. You have no way of going in and finding some ethereal "essence of hurricane," some "hurricane soul," located right in the middle of the eye! It's the existence of a certain kind of pattern—a spiral storm with an eye and so forth that makes you say it's a hurricane. Of course there are a lot of things that you'll insist on before you call something a hurricane.

**PAT**: Well, wouldn't you say that being an atmospheric phenomenon is one vital prerequisite? How can anything inside a computer be a storm? To me, a simulation is a simulation!
Sandy: Then I suppose you would say that even the calculations that computers do are simulated— that they are fake calculations. Only people can do genuine calculations, right?

Pat: Well, computers get the right answers, so their calculations are not exactly fake—but they're still just patterns. There's no understanding going on in there. Take a cash register. Can you honestly say that you feel it is calculating something when its gears turn on each other? And a computer is just a fancy cash register, as I understand it.

Sandy: If you mean that a cash register doesn't feel like a schoolkid doing arithmetic problems, I'll agree. But is that what "calculation" means? Is that an integral part of it? If so, then contrary to what everybody has thought till now, we'll have to write a very complicated program to perform genuine calculations. Of course, this program will sometimes get careless and make mistakes and it will sometimes scrawl its answers illegibly, and it will occasionally doodle on its paper... It won't be more reliable than the post office clerk who adds up your total by hand. Now, I happen to believe eventually such a program could be written. Then we'd know something about how post office clerks and schoolkids work.

Pat: I can't believe you could ever do that!

Sandy: Maybe, maybe not, but that's not my point. You say a cash register can't calculate. It reminds me of another favorite passage of mine from Dennett's Brainstorms—a rather ironic one, which is why I like it. The passage goes something like this: "Cash registers can't really calculate; they can only spin their gears. But cash registers can't really spin their gears either; they can only follow the laws of physics." Dennett said it originally about computers; I modified it to talk about cash registers. And you could use the same line of reasoning in talking about people: "People can't really calculate; all they can do is manipulate mental symbols. But they aren't really manipulating symbols; all they are doing is firing various neurons in various patterns. But they can't really make their neurons fire; they simply have to let the laws of physics make them fire for them." Et cetera. Don't you see how this Dennett-inspired reductio ad absurdum would lead you to conclude that calculation doesn't exist, hurricanes don't exist, nothing at a higher level than particles and the laws of physics exists? What do you gain by saying a computer only pushes symbols around and doesn't, truly calculate?

Pat: The example may be extreme, but it makes my point that there is a vast difference between a real phenomenon and any simulation of it. This is so for hurricanes, and even more so for human thought.

Sandy: Look, I don't want to get too tangled up in this line of argument, but let me try out one more example. If you were a radio ham listening to another ham broadcasting in Morse code and you were responding in Morse code, would it sound funny to you to refer to "the person at the other end"?

Pat: No, that would sound okay, although the existence of a person at the other end would be an assumption.

Sandy: Yes, but you wouldn't be likely to go and check it out. You're prepared to recognize personhood through those rather unusual channels. You don't have to see a human body or hear a voice (all you need is a rather abstract manifestation—a code, as it were). What I'm getting at is this. To "see" the person behind the dits and dats, you have to be willing to do some decoding, some interpretation. It's not direct perception; it's indirect. You have to peel off a layer or two, to find the reality hidden in there. You put on your "radio-ham's glasses" to "see" the person behind the buzzes. Just the same
with the simulated hurricane! You don't see it darkening the machine room - you have to decode the machine's memory. You have to put on special "memory-decoding glasses." Then what you see is a hurricane!

PAT: Oh, ho ho! Talk about fast ones - wait a minute! In the case of the shortwave radio, there's a real person out there, somewhere in the Fiji Islands or wherever. My decoding act as I sit by my radio simply reveals that that person exists. It's like seeing a shadow and concluding there's an object out there, casting it. One doesn't confuse the shadow with the object, however! And with the hurricane there's no real hurricane behind the scenes, making the computer follow its patterns. No, what you have is just a shadow hurricane without any genuine hurricane. I just refuse to confuse shadows with reality.

SANDY: All right. I don't want to drive this point into the ground. I even admit it is pretty silly to say that a simulated hurricane is a hurricane. But I wanted to point out that it's not as silly as you might think at first blush. And when you turn to simulated thought, you've got a very different matter on your hands from simulated hurricanes.

PAT: I don't see why. A brainstorm sounds to me like a mental hurricane. But seriously, you'll have to convince me.

SANDY: Well, to do so I'll have to make a couple of extra points about hurricanes first.

PAT: Oh, no! Well, all right, all right.

SANDY: Nobody can say just exactly what a hurricane is - that is, in totally precise terms. There's an abstract pattern that many storms share, and it's for that reason that we call those storms hurricanes. But it's not possible to make a sharp distinction between hurricanes and nonhurricanes. There are tornados, cyclones, typhoons, dustdevils... Is the Great Red Spot on Jupiter a hurricane? Are sunspots hurricanes? Could there be a hurricane in a wind tunnel? In a test tube? In your imagination you can extend the concept of "hurricane" to include a microscopic storm on the surface of a neutron star.

CHRIS: That's not so far-fetched, you know. The concept of "earthquake" has actually been extended to neutron stars. The astrophysicists say that the tiny changes in rate that once in a while are observed in the pulsing of a pulsar are caused by "glitches" (starquakes) that have just occurred on the neutron star's surface.

SANDY: Yes, I remember that now. The idea of a "glitch" strikes me as wonderfully eerie - a surrealistic kind of quivering on a surrealistic kind of surface.

CHRIS: Can you imagine - plate tectonics on a giant rotating sphere of pure nuclear matter?

SANDY: That's a wild thought. So starquakes and earthquakes can both be subsumed into a new, more abstract category. And that's how science constantly extends familiar concepts, taking them further and further from familiar experience and yet keeping some essence constant. The number system is the classic example - from positive numbers to negative numbers, then rationals, reals, complex numbers, and "on beyond zebra," as Dr. Seuss says.

PAT: I think I can see your point here, Sandy. We have many examples in biology of close relationships that are established in rather abstract ways. Often the decision about what family some species belongs to comes down to an abstract pattern shared at some level. When you base your system of classification on very abstract patterns, I suppose that a broad variety of phenomena can fall into "the same class," even if in many superficial ways the class members are utterly unlike each other. So perhaps I
can glimpse, at least a little, how to you a simulated hurricane could, in some funny sense, be a hurricane.

**Chris**: Perhaps the word that's being extended is not "hurricane" but "be"!

**Pat**: How so?

**Chris**: If Turing can extend the verb "think," can't I extend the verb "be"? All I mean is that when simulated things are deliberately confused with the genuine article, somebody's doing a lot of philosophical wool-pulling. It's a lot more serious than just extending a few nouns such as "hurricane."

**Sandy**: I like your idea that "be" is being extended, but I think your slur about "wool-pulling" goes too far. Anyway, if you don't object, let me just say one more thing about simulated hurricanes and then I'll get to simulated minds. Suppose you consider a really deep simulation of a hurricane -I mean a simulation of every atom, which I admit is impossibly deep. I hope you would agree that it would then share all that abstract structure that defines the "essence of hurricanehood." So what's to hold you back from calling it a hurricane?

**Pat**: I thought you were backing off from that claim of equality!

**Sandy**: So did I, but then these examples came up, and I was forced back to my claim. But let me back off, as I said I would do, and get back to thought, which is the real issue here. Thought, even more than hurricanes, is an abstract structure, a way of describing some complex events that happen in a medium called a brain. But actually thought can take place in any of several billion brains. There are all these physically very different brains, and yet they all support "the same thing" thinking. What's important, then, is the abstract pattern, not the medium. The same kind of swirling can happen inside any of them, so no person can claim to think more "genuinely" than any other. Now, if we come up with some new kind of medium in which the same style of swirling takes place, could you deny that thinking is taking place in it?

**Pat**: Probably not, but you have just shifted the question. The question now is, how can you determine whether "the same style" of swirling is really happening?

**Sandy**: The beauty of the Turing test is that it tells you, when!

**Chris**: I don't see that at all. How would you know that the same style of activity was occurring inside a computer as inside my mind, simply because it answered questions as I do? All you're looking at is its outside.

**Sandy**: But how do you know that when I speak to you, anything similar to what you call "thinking" is going on inside me? The Turing test is a fantastic probe, something like a particle accelerator in physics. Chris, I think you'll like this analogy. Just as in physics, when you want to understand what is going on at an atomic or subatomic level, since you can't see it directly, you scatter accelerated particles off the target in question and observe their behavior. From this you infer the internal nature of the target. The Turing test extends this idea to the mind. It treats the mind as a "target" that is not directly visible but whose structure can be deduced more abstractly. By "scattering" questions off a target mind, you learn about its internal workings, just as in physics.

**Chris**: More exactly put, you can hypothesize about what kinds of internal structures might account for the behavior observed -but they may or may not in fact exist.
SANDY: Hold on, now! Are you saying that atomic nuclei are merely hypothetical entities? After all, their existence (or should I say "hypothetical existence") was proven (or should I say "suggested") by the behavior of particles scattered off of atoms.

CHRIS: Physical systems seem to me to be much simpler than the mind, and the certainty of the inferences made is correspondingly greater.

SANDY: The experiments are also correspondingly harder to perform and to interpret. In the Turing test, you could perform many highly delicate experiments in the course of an hour. I maintain that people give other people credit for being conscious simply because of their continual external monitoring - which is itself something like a Turing test.

PAT: That may be roughly true, but it involves more than just conversing with people through a teletype. We see that other people have bodies, we watch their faces and expressions - we see they are fellow human beings and so we think they think.

SANDY: To me, that seems a highly anthropocentric view of what thought is. Does that mean you would sooner say a mannikin in a store thinks than a wonderfully programmed computer, simply because the mannikin looks more human?

PAT: Obviously I would need more than just vague physical resemblance to the human form to be willing to attribute the power of thought to an entity. But that organic quality, the sameness of origin, undeniably lends a degree of credibility that is very important.

SANDY: Here we disagree. I find this simply too chauvinistic. I feel that the key thing is a similarity of internal structure (not bodily, organic, chemical structure, but organizational structure) software. Whether an entity can think seems to me a question of whether its organization can be described in a certain way, and I'm perfectly willing to believe that the Turing test detects the presence or absence of that mode of organization. I would say that your depending on my physical body as evidence that I am a thinking being is rather shallow. The way I see it, the Turing test looks far deeper than at mere external form.

PAT: Hey now! you're not giving me much credit. It's not just the shape of a body that lends weight to the idea there's real thinking going on inside - it's also, as I said, the idea of common origin. It's the idea that you and I both sprang from DNA molecules, an idea to which I attribute much depth. Put it this way: The external form of human bodies reveals that they share a deep biological history, and it's that depth that lends a lot of credibility to the notion that the owner of such a body can think.

SANDY: But that is all indirect evidence. Surely you want some direct evidence. That is what the Turing test is for. And I think it is the only way to test for "thinkinghood."

CHRIS: But you could be fooled by the Turing test, just as an interrogator could think a man was a woman.

SANDY: I admit, I could be fooled if I carried out the test in too quick or too shallow a way. But I would go for the deepest things I could think of.

CHRIS: I would want to see if the program could understand jokes. That would be a real test of intelligence.

SANDY: I agree that humor probably is an acid test for a supposedly intelligent program, but equally important to me perhaps more so would be to test its
emotional responses. So I would ask it about its reactions to certain pieces of music or works of literature—especially my favorite ones.

**CHRIS:** What if it said, "I don't know that piece," or even "I have no interest in music"? What if it avoided all emotional references?

**SANDY:** That would make me suspicious. Any consistent pattern of avoiding certain issues would raise serious doubts in me as to whether I was dealing with a thinking being.

**CHRIS:** Why do you say that? Why not say that you're dealing with a thinking but unemotional being?

**SANDY:** You've hit upon a sensitive point. I simply can't believe that emotions and thought can be divorced. Put another way, I think that emotions are an automatic by-product of the ability to think. They are implied by the very nature of thought.

**CHRIS:** Well, what if you're wrong? What if I produced a machine that could think but not emote? Then its intelligence might go unrecognized because it failed to pass your kind of test.

**SANDY:** I'd like you to point out to me where the boundary line between emotional questions and nonemotional ones lies. You might want to ask about the meaning of a great novel. This requires understanding of human emotions! Is that thinking or merely cool calculation? You might want to ask about a subtle choice of words. For that you need an understanding of their connotations. Turing uses examples like this in his article. You might want to ask it for advice about a complex romantic situation. It would need to know a lot about human motivations and their roots. Now if it failed at this kind of task, I would not be much inclined to say that it could think. As far as I am concerned, the ability to think, the ability to feel, and consciousness are just different facets of one phenomenon, and no one of them can be present without the others.

**CHRIS:** Why couldn't you build a machine that could feel nothing, but that could think and make complex decisions anyway? I don't see any contradiction there.

**SANDY:** Well, I do. I think that when you say that, you are visualizing a metallic, rectangular machine, probably in an air-conditioned room—a hard, angular, cold object with a million colored wires inside it, a machine that sits stock still on a tiled floor, humming or buzzing or whatever, and spinning its tapes. Such a machine can play a good game of chess, which, I freely admit, involves a lot of decision making. And yet I would never call such a machine conscious.

**CHRIS:** How come? To mechanists, isn't a chess-playing machine rudimentarily conscious?

**SANDY:** Not to this mechanist. The way I see it, consciousness has got to come from a precise pattern of organization—one that we haven't yet figured out how to describe in any detailed way. But I believe we will gradually come to understand it. In my view consciousness requires a certain way of mirroring the external universe internally, and the ability to respond to that external reality on the basis of the internally represented model. And then in addition, what's really crucial for a conscious machine is that it should incorporate a well developed and flexible self-model. And it's there that all existent programs, including the best chess-playing ones, fall down.

**CHRIS:** Don't chess programs look ahead and say to themselves as they're figuring out their next move, "If you move here, then I'll go there, and then if you go this way, I could go that way..."? Isn't that a sort of self-model?
Sandy: Not really. Or, if you want, it's an extremely limited one. It's an understanding of self only in the narrowest sense. For instance, a chess-playing program has no concept of why it is playing chess, or the fact that it is a program, or is in a computer, or has a human opponent. It has no ideas about what winning and losing are, or...

Pat: How do you know it has no such sense? How can you presume to say what a chess program feels or knows?

Sandy: Oh, come on! We all know that certain things don't feel anything or know anything. A thrown stone doesn't know anything about parabolas, and a whirling fan doesn't know anything about air. It's true I can't prove those statements, but here we are verging on questions of faith.

Pat: This reminds me of a Taoist story I read. It goes something like this. Two sages were standing on a bridge over a stream. One said to the other, "I wish I were a fish. They are so happy!" The second replied, "How do you know whether fish are happy or not? You're not a fish." The first said, "But you're not me, so how do you know whether I know how fish feel?"

Sandy: Beautiful! Talking about consciousness really does call for a certain amount of restraint. Otherwise you might as well just jump on either the solipsism bandwagon ("I am the only conscious being in the universe") or the panpsychism bandwagon ("Everything in the universe is conscious")!

Pat: Well, how do you know? Maybe everything is conscious.

Sandy: If you're going to join those who claim that stones and even particles like electrons have some sort of consciousness, then I guess we part company here. That's a kind of mysticism I can't fathom. As for chess programs, I happen to know how they work, and I can tell you for sure that they aren't conscious! No way!

Pat: Why not?

Sandy: They incorporate only the barest knowledge about the goals of chess. The notion of "playing" is turned into the mechanical act of comparing a lot of numbers and choosing the biggest one over and over again. A chess program has no sense of shame about losing or pride in winning. Its self-model is very crude. It gets away with doing the least it can, just enough to play a game of chess and do nothing more. Yet, interestingly enough, we still tend to talk about the "desires" of a chess-playing computer. We say, "It wants to keep its king behind a row of pawns," or "It likes to get its rooks out early," or "It thinks I don't see that hidden fork."

Pat: Well, we do the same thing with insects. We spot a lonely ant somewhere and say, "It's trying to get back home" or "It wants to drag that dead bee back to the colony." In fact, with any animal we use terms that indicate emotions, but we don't know for sure how much the animal feels. I have no trouble talking about dogs and cats being happy or sad, having desires and beliefs and so on, but of course I don't think their sadness is as deep or complex as human sadness is.

Sandy: But you wouldn't call it "simulated sadness," would you?

Pat: No, of course not. I think it's real.

Sandy: It's hard to avoid use of such teleological or mentalistic terms. I believe they're quite justified, although they shouldn't be carried too far. They simply don't have the same richness of meaning when applied to present-day chess programs as when applied to people.
**CHRIS:** I still can't see that intelligence has to involve emotions. Why couldn't you imagine an intelligence that simply calculates and has no feelings?

**SANDY:** A couple of answers here! Number one, any intelligence has to have motivations. It's simply not the case, whatever many people may think, that machines could think any more "objectively" than people do. Machines, when they look at a scene, will have to focus and filter that scene down into some preconceived categories, just as a person does. And that means seeing some things and missing others. It means giving more weight to some things than to others. This happens on every level of processing.

**PAT:** What do you mean?

**SANDY:** Take me right now, for instance. You might think that I'm just making some intellectual points, and I wouldn't need emotions to do that. But what makes me care about these points? Why did I stress the word "care" so heavily? Because I'm emotionally involved in this conversation! People talk to each other out of conviction, not out of hollow, mechanical reflexes. Even the most intellectual conversation is driven by underlying passions. There's an emotional undercurrent to every conversation—it's the fact that the speakers want to be listened to, understood, and respected for what they are saying.

**PAT:** It sounds to me as if all you're saying is that people need to be interested in what they're saying, otherwise a conversation dies.

**SANDY:** Right! I wouldn't bother to talk to anyone if I weren't motivated by interest. And interest is just another name for a whole constellation of subconscious biases. When I talk, all my biases work together and what you perceive on the surface level is my style, my personality. But that style arises from an immense number of tiny priorities, biases, leanings. When you add up a million of these interacting together, you get something that amounts to a lot of desires. It just all adds up! And that brings me to the other point, about feelingless calculation. Sure, that exists—in a cash register, a pocket calculator. I'd say it's even true of all today's computer programs. But eventually, when you put enough feelingless calculations together in a huge coordinated organization, you'll get something that has properties on another level. You can see it (in fact, you have to see it) not as a bunch of little calculations, but as a system of tendencies and desires and beliefs and so on. When things get complicated enough, you're forced to change your level of description. To some extent that's already happening, which is why we use words such as "want," "think," "try," and "hope," to describe chess programs and other attempts at mechanical thought. Dennett calls that kind of level switch by the observer "adopting the intentional stance." The really interesting things in AI will only begin to happen, I'd guess, when the program itself adopts the intentional stance toward itself!

**CHRIS:** That would be a very strange sort of level-crossing feedback loop.

**SANDY:** It certainly would. Of course, in my opinion, it's highly premature for anyone to adopt the intentional stance, in the full force of the term, toward today's programs. At least that's my opinion.

**CHRIS:** For me an important related question is: To what extent is it valid to adopt the intentional stance toward beings other than humans?

**PAT:** I would certainly adopt the intentional stance toward mammals.

**SANDY:** I vote for that.
**CHRIS:** That's interesting! How can that be, Sandy? Surely you wouldn't claim that a dog or cat can pass the Turing test? Yet don't you think that the Turing test is the only way to test for the presence of thought? How can you have these beliefs at once?

**SANDY:** Hmm... All right. I guess I'm forced to admit that the Turing test works only above a certain level of consciousness. There can be thinking beings that could fall the test -but on the other hand, anything that passes it, in my opinion, would be a genuinely conscious, thinking being.

**PAT:** How can you think of a computer as a conscious being? I apologize if this sounds like a stereotype, but when I think of conscious beings, I just can't connect that thought with machines. To me consciousness is connected with soft, warm bodies, silly though that may sound.

**CHRIS:** That does sound odd, coming from a biologist. Don't you deal with life in terms of chemistry and physics enough for all magic to seem to vanish?

**PAT:** Not really. Sometimes the chemistry and physics just increase the feeling that there's something magical going on down there! Anyway, I can't always integrate my scientific knowledge with my gutlevel feelings.

**CHRIS:** I guess I share that trait.

**PAT:** So how do you deal with rigid preconceptions like mine?

**SANDY:** I'd try to dig down under the surface of your concept of "machines" and get at the intuitive connotations that lurk there, out of sight but deeply influencing your opinions. I think that we all have a holdover image from the Industrial Revolution that sees machines as clunky iron contraptions gawkily moving under the power of some loudly chugging engine. Possibly that's even how the computer inventor Charles Babbage viewed people! After all, he called his magnificent many-geared computer the Analytical Engine.

**PAT:** Well, I certainly don't think people are just fancy steam shovels or even electric can openers. There's something about people, something that... that... they've got a sort of flame inside them, something alive, something that flickers unpredictably, wavering, uncertain but something creative!

**SANDY:** Great! That's just the sort of thing I wanted to hear. It's very human to think that way. Your flame image makes me think of candles, of fires, of thunderstorms with lightning dancing all over the sky in crazy patterns. But do you realize that just that kind of pattern is visible on a computer's console? The flickering lights form amazing chaotic sparkling patterns. It's such a far cry from heaps of lifeless clanking metal! It is flamelike, by God! Why don't you let the word "machine" conjure up images of dancing patterns of light rather than of giant steam shovels?

**CHRIS:** That's a beautiful image, Sandy. It changes my sense of mechanism from being matter-oriented to being pattern-oriented. It makes me try to visualize the thoughts in my mind (these thoughts right now, even) as a huge spray of tiny pulses flickering in my brain.

**SANDY:** That's quite a poetic self-portrait for a spray of flickers to have come up with!

**CHRIS:** Thank you. But still, I'm not totally convinced that a machine is all that I am. I admit, my concept of machines probably does suffer from anachronistic subconscious flavors, but I'm afraid I can't change such a deeply rooted sense in a flash.
SANDY: At least you do sound open-minded. And to tell the truth, part of me does sympathize with the way you and Pat view machines. Part of me balks at calling myself a machine. It is a bizarre thought that a feeling being like you or me might emerge from mere circuitry. Do I surprise you?

CHRIS: You certainly surprise me. So tell us -do you believe in the idea of an intelligent computer, or don't you?

SANDY: It all depends on what you mean. We have all heard the question "Can computers think?" There are several possible interpretations of this (aside from the many interpretations of the word "think"). They revolve around different meanings of the words "can" and "computer."

PAT: Back to word games again....

SANDY: That's right. First of all, the question might mean "Does some present-day computer think, right now?" To this I would immediately answer with a loud "no." Then it could be taken to mean, "Could some present-day computer, if suitably programmed, potentially think?" This is more like it, but I would still answer, "Probably not." The real difficulty hinges on the word "computer." The way I see it, "computer" calls up an image of just what I described earlier: an air-conditioned room with cold rectangular metallic boxes in it. But I suspect that with increasing public familiarity with computers and continued progress in computer architecture, that vision will eventually become outmoded.

PAT: Don't you think computers, as we know them, will be around for a while?

SANDY: Sure, there will have to be computers in today's image around for a long time, but advanced computers (maybe no longer called computers) will evolve and become quite different. Probably, as in the case of living organisms, there will be many branchings in the evolutionary tree. There will be computers for business, computers for schoolkids, computers for scientific calculations, computers for systems research, computers for simulation, computers for rockets going into space, and so on. Finally, there will be computers for the study of intelligence. It's really only these last that I'm thinking of -the ones with the maximum flexibility, the ones that people are deliberately attempting to make smart. I see no reason that these will stay fixed in the traditional image. Probably they will soon acquire as standard features some rudimentary sensory systems -mostly for vision and hearing, at first. They will need to be able to move around, to explore. They will have to be physically flexible. In short, they will have to become more animal-like, more self-reliant.

CHRIS: It makes me think of the robots R2D2 and C3PO in *Star Wars.*

SANDY: As a matter of fact I don't think of anything like them when I visualize intelligent machines. They're too silly, too much the product of a film designer's imagination. Not that I have a clear vision of my own. But I think it is necessary, if people are going to try realistically to imagine an artificial intelligence, to go beyond the limited, hard-edged image of computers that comes from exposure to what we have today. The only thing that all machines will always have in common is their underlying mechanicalness. That may sound cold and inflexible, but what could be more mechanical (in a wonderful way) than the operations of the DNA and proteins and organelles in our cells?

PAT: To me what goes on inside cells has a "wet," "slippery" feel to it, and what goes on inside machines is dry and rigid. It's connected with the fact that computers don't make mistakes, that computers do only what you tell them to do. Or at least that's my image of computers.
SANDY: Funny -a minute ago your image was of a flame, and now it's of something "wet" and "slippery." Isn't it marvelous how contradictory we can be?

PAT: I don't need your sarcasm.

SANDY: I'm not being sarcastic -I really do think it is marvelous.

PAT: It's just an example of the human mind's slippery nature -mine, in this case.

SANDY: True. But your image of computers is stuck in a rut. Computers certainly can make mistakes -and I don't mean on the hardware level. Think of any present-day computer predicting the weather. It can make wrong predictions, even though its program runs flawlessly.

PAT: But that's only because you've fed it the wrong data.

SANDY: Not so. It's because weather prediction is too complex. Any such program has to make do with a limited amount of data (entirely correct data) and extrapolate from there. Sometimes it will make wrong predictions. It's no different from the farmer in the field gazing at the clouds who says, "I reckon we'll get a little snow tonight." We make models of things in our heads and use them to guess how the world will behave. We have to make do with our models, however inaccurate they may be. And if they're too inaccurate, evolution will prune us out -we'll fall over a cliff or something. And computers are the same. It's just that human designers will speed up the evolutionary process by aiming explicitly at the goal of creating intelligence, which is something nature just stumbled on.

PAT: So you think computers will make fewer mistakes as they get smarter?

SANDY: Actually, just the other way around. The smarter they get, the more they'll be in a position to tackle messy real-life domains, so they'll be more and more likely to have inaccurate models. To me, mistake making is a sign of high intelligence!

PAT: Boy! You throw me sometimes!

SANDY: I guess I'm a strange sort of advocate for machine intelligence. To some degree I straddle the fence. I think that machines won't really be intelligent in a humanlike way until they have something like that biological wetness or slipperiness to them. I don't mean literally wet -the slipperiness could be in the software. But biological-seeming or not, intelligent machines will in any case be machines. We will have designed them, built them -or grown them! We will understand how they work -at least in some sense. Possibly no one person will really understand them, but collectively we will know how they work.

PAT: It sounds like you want to have your cake and eat it too.

SANDY: You're probably right. What I'm getting at is that when artificial intelligence comes, it will be mechanical and yet at the same time organic. It will have that same astonishing flexibility that we see in life's mechanisms. And when I say "mechanisms," I mean "mechanisms." DNA and enzymes and so on really are mechanical and rigid and reliable. Wouldn't you agree, Pat?

PAT: That's true. But when they work together, a lot of unexpected things happen. There are so many complexities and rich modes of behavior that all that mechanicalness adds up to something very fluid.

SANDY: For me it's an almost unimaginable transition from the mechanical level of molecules to the living level of cells. But it's what convinces me that people are
machines. That thought makes me uncomfortable in some ways, but in other ways it is an exhilarating thought.

**CHRIS:** If people are machines, how come it's so hard to convince them of the fact? Surely if we are machines, we ought to be able to recognize our own machinehood.

**SANDY:** You have to allow for emotional factors here. To be told you're a machine is, in a way, to be told that you're nothing more than your physical parts, and it brings you face to face with your own mortality. That's something nobody finds easy to face. But beyond the emotional objection, to see yourself as a machine you have to jump all the way from the bottommost mechanical level to the level where the complex lifelike activities take place. If there are many intermediate layers, they act as a shield, and the mechanical quality becomes almost invisible. I think that's how intelligent machines will seem to us (and to themselves!) when they come around.

**PAT:** I once heard a funny idea about what will happen when we eventually have intelligent machines. When we try to implant that intelligence into devices we'd like to control, their behavior won't be so predictable.

**SANDY:** They'll have a quirky little "flame" inside, maybe?

**PAT:** Maybe.

**CHRIS:** So what's so funny about that?

**PAT:** Well, think of military missiles. The more sophisticated their target-tracking computers get, according to this idea, the less predictably they will function. Eventually you'll have missiles that will decide they are pacifists and will turn around and go home and land quietly without blowing up. We could even have "smart bullets" that turn around in midflight because they don't want to commit suicide!

**SANDY:** That's a lovely thought.

**CHRIS:** I'm very skeptical about these ideas. Still, Sandy, I'd like to hear your predictions about when intelligent machines will come to be.

**SANDY:** It won't be for a long time, probably, that we'll see anything remotely resembling the level of human intelligence. It just rests on too awesomely complicated a substrate (the brain) for us to be able to duplicate it in the foreseeable future. Anyway, that's my opinion.

**PAT:** Do you think a program will ever pass the Turing test?

**SANDY:** That's a pretty hard question. I guess there are various degrees of passing such a test, when you come down to it. It's not black and white. First of all, it depends on who the interrogator is. A simpleton might be totally taken in by some programs today. But secondly, it depends on how deeply you are allowed to probe.

**PAT:** Then you could have a scale of Turing tests - one-minute versions, five-minute versions, hour-long versions, and so forth. Wouldn't it be interesting if some official organization sponsored a periodic competition, like the annual computer-chess champion-ships, for programs to try to pass the Turing test?

**CHRIS:** The program that lasted the longest against some panel of distinguished judges would be the winner. Perhaps there could be a high prize for the first program that fools a famous judge for, say, ten minutes.

**PAT:** What would a program do with a prize?
**CHRIS**: Come now, Pat. If a program’s good enough to fool the judges, don’t you think it’s good enough to enjoy the prize?

**PAT**: Sure, especially if the prize is an evening out on the town, dancing with all the interrogators!

**SANDY**: I’d certainly like to see something like that established. I think it could be hilarious to watch the first programs flop pathetically!

**PAT**: You’re pretty skeptical, aren’t you? Well, do you think any computer program today could pass a five-minute Turing test, given a sophisticated interrogator?

**SANDY**: I seriously doubt it. It’s partly because no one is really working at it explicitly. However, there is one program called “Parry” which its inventors claim has already passed a rudimentary version of the Turing test. In a series of remotely conducted interviews, Parry fooled several psychiatrists who were told they were talking to either a computer or a paranoid patient. This was an improvement over an earlier version, in which psy psychiatrists were simply handed transcripts of short interviews and asked to determine which ones were with a genuine paranoid and which ones with a computer simulation.

**PAT**: You mean they didn’t have the chance to ask any questions? That’s a severe handicap—and it doesn’t seem in the spirit of the Turing test. Imagine someone trying to tell which sex I belong to just by reading a transcript of a few remarks by me. It might be very hard! So I’m glad the procedure has been improved.

**CHRIS**: How do you get a computer to act like a paranoid?

**SANDY**: I’m not saying it does act like a paranoid, only that some psychiatrists, under unusual circumstances, thought so. One of the things that bothered me about this pseudo-Turing test is the way Parry works. “He” (as they call him) acts like a paranoid in that he gets abruptly defensive, veers away from undesirable topics in the conversation, and, in essence, maintains control so that no one can truly probe “him.” In this way, a simulation of a paranoid is a lot easier than a simulation of a normal person.

**PAT**: No kidding! It reminds me of the joke about the easiest kind of human for a computer program to simulate.

**CHRIS**: What is that?

**PAT**: A catatonic patient—they just sit and do nothing at all for days on end. Even I could write a computer program to do that!

**SANDY**: An interesting thing about Parry is that it creates no sentences on its own—it merely selects from a huge repertoire of canned sentences the one that best responds to the input sentence.

**PAT**: Amazing! But that would probably be impossible on a larger scale, wouldn’t it?

**SANDY**: Yes. The number of sentences you’d need to store to be able to respond in a normal way to all possible sentences in a conversation is astronomical, really unimaginable. And they would have to be so intricately indexed for retrieval.... Anybody who thinks that somehow a program could be rigged up just to pull sentences out of storage like records in a jukebox, and that this program could pass the Turing test, has not thought very hard about it. The funny part about it is that it is just this kind of unrealizable program that some enemies of artificial intelligence cite when arguing
against the concept of the Turing test. Instead of a truly intelligent machine, they want you to imagine a gigantic, lumbering robot that intones canned sentences in a dull monotone. It’s assumed that you could see through to its mechanical level with ease, even if it were simultaneously performing tasks that we think of as fluid, intelligent processes. Then the critics say, “You see! It would still be just a machine—a mechanical device, not intelligent at all!” I see things almost the opposite way. If I were shown a machine that can do things that I can do (I mean pass the Turing test) then, instead of feeling insulted or threatened, I’d chime in with the philosopher Raymond Smullyan and say, “How wonderful machines are!”

**Chris**: If you could ask a computer just one question in the Turing test, what would it be?

**Sandy**: Uhmm...

**Pat**: How about “If you could ask a computer just one question in the Turing test, what would it be?”?
6. Connectionist architecture

6.1. The architecture of the connectionist mind

The academic year 81 was almost over. The subject *Cognition* that had generated much expectation in Bill had been upsetting. The contents were fine but the hypotheses and the conclusions were unsatisfactory. It was boring to constantly listen that his mind was similar to the computer he had in his room to write some course projects and to play computer games, mainly because the machine blocked itself in the most exciting part of the game. The analogy computer-mind didn't make any sense to him.

Moved by this skepticism, Bill went to class decided to check whether his ideas and conclusions were logic. He tried to find out if this understanding the mind as a computer manipulating symbols was the engineers’ fault. Thus, he went to the last class and saw that the results of his examinations had been excellent. However, he wasn’t convinced of what scientists were telling him. So he raised his hand and made a question. The lecturer was a David Rumelhart, who immediately paid attention to Bill:

- *Bill, what’s your question?*
- *Professor Rumelhart, what had happened if, in the past, the engineers who developed the first computers had chosen a parallel processing for their design, instead of the online processing?*

After a long silence, Rumelhart smiled and answered him that all the cognitive models would have been very different to how they had been explained throughout the academic year. The inexhaustible number of models and theories assuming that the human mind processes information online would not have existed. Bill was proud and satisfied. At last, someone recognized that his mind could work differently to the computer on his table (*Cognición Humana: Mente, ordenadores y neuronas, Antonio Crespo (2002)).*

If behaviorism arose against mentalism and classic cognitivism, based on the online computer manipulating symbols, against behaviorism, connectionism was born in the mid 80s as an alternative to the cognitive IP models. The connectionists proposed models that represented distributed information, i.e. a set of processing units operating simultaneously. The birth of connectionism is almost simultaneous to the one of classic cognitivism. In 1943, McCulloch and Pitt described how an architecture inspired in neural networks could calculate logical functions. Indeed, by interconnecting simple elementary processors (called *Perceptrons* by Frank Rosenblatt, see figure 6.1.) able to receive and transmit energy across their matrix, and to adapt their output value depending on the input value (as neurons do by means of their *axons* and *dendrites*), the truth table of the logical connectors (*y, o*, etc.) can be reproduced. In short, the perceptron is a neural network in which the output(s) of some neurons are connected by weights to the input(s) of other neurons. According to McCulloch and Pitt, there are no *mental* programs (i.e. the software), as the classic cognitivism claims, but a neuronal physical structure (i.e. the *hardware*) which carries out those logical operations, metaphorically called mental - which are simply *brain-like*. 
6.1. The perceptron has two layers: the input units (analogous to sensory neurons) and the output units (analogous to motor neurons).

Until the end of the 70s of the 20th century, this theory developed further together with classic cognitivism, but Marvin Minsky & Seymour Papert claimed in their book *Perceptrons: An Introduction to Perceptual Geometry* (1969) that those perceptrons had serious limitations when simulating brain functions. They proved that perceptrons were unable to calculate some simple mental functions (for instance, to distinguish objects from their parts) without performing a huge number of operations and thus, unable to clearly describe how the mind performed these operations effortlessly.

Consider the following example (due to Minsky and Papert and cited in Clark, 1990:26):

- Janet: "That isn’t a very good ball you have. Give it to me and I’ll give you my lollipop."

For the set of micro theories needed to understand Janet’s words, Minsky and Papert suggested a lengthy list of concepts including: time, space, things, people, words, thoughts, talking, social relations, playing, owning, eating, liking, living, intention, emotions, states, properties, stories, and places. Each of these requires filling out. This is a daunting list and one that shows no obvious signs of completion or even completability. Besides, there is the problem of accessing the right frame at the right time, as humans usually do.

At the end of the 80s, however, the connectionism reappeared with the publication of David Rumelhart and James McClelland, *Parallel Distributed Processing*, in which they described the multi-layer perceptron capable of fast learning.
6.1.1. The brain metaphor of the mind

According to Rumelhart, classic cognitivism focused on the logical or functional level – the software (in the computer analogy) – leaving aside mental models describing a more basic level: the implementation. In other words, classic cognitivism understood the mind as a computer program that could be executed in any machine (hardware). Connectionism focused on the implementation level, as it considered that some algorithms could be designed to explain human phenomena within this level. If we admit that the mind is implemented on a physical substrate –the brain–, a connectionist will elaborate cognitive models based on a neural architecture. Thus, connectionists substituted the computer metaphor for the brain metaphor of the mind and abandoned the symbolic computation for one inspired by neural functioning.

Connectionism pretended to design artificial intelligent systems inspired by the brain, not a physiological theory of the brain. Therefore, connectionist models considered the neuron as the basic processing unit. The neuron was understood as a computational abstract element, organized in artificial neural networks (ANN). Thus, just as there are electric synapses between neurons (see figure 6.2.),

![Biological neural network](image)

in the connectionist models the mapping between processing units or artificial neurons is shown by means of connecting lines which constitute complex processing networks; hence, it is said that these models propose a neuromorphic conception of the mind.
6.3. Basic architecture of a connectionist network

6.1.2. Structure of the connectionist models

An artificial neural network (ANN) tries to simulate the structure and/or functional aspects of biological neural networks. The basic elements of the connectionist architecture are the processing units or artificial neurons, which are organized in networks characterized by a connectivity pattern. These artificial networks are usually disposed in layers and relate to each other by a variable pattern of activation that represents the actions that the system can execute.

The connectionism proposes a formal system of computation similar to that of a neuron about to generate an action potential spike. An input unit sends its activation value to each of the hidden units to which it is connected. Each of these hidden units calculates its own activation value depending on the activation values it receives from the input units. This signal is then passed onto the output units or to another layer of hidden units. Those hidden units compute their activation values in the same way, and send them along to their neighbors. Eventually the signal at the input units propagates all the way through the net to determine the activation values at all the output units. Connectionists presume that cognitive functioning can be explained by collections of units that operate in this way. In essence, connectionist models involve two basic parameters for the processing units: their spreading activation and their transfer function.

- **Spreading activation**: At any time, a unit in the network has an activation, which is a numerical value intended to represent some aspect of the unit. The pattern of activation set up by a net is determined by the weights, or strength of connections between the units. Weights may be both positive and negative. A negative weight represents the inhibition of the receiving unit. The activation value for each receiving unit is calculated according to a simple activation function (activation value ϕ) and can be continuous or discrete. The function sums together the contributions of all sending units, where the contribution of a unit is defined as the weight of the connection between the sending and receiving units. This sum is usually modified further, for example, by adjusting the activation sum to a value between 0 (OFF) and 1 (ON) and/or by setting the activation to zero unless a threshold level for the sum is reached.
- **Transfer function**: This function determines both the activation value $t$ that the unit will adopt from the inputs it receives and the output of the unit. Thus, it is the result of combining the activation function and the output, which act sequentially.

- **Activation function**: It is a simple rule which decides the new activation state $t$ of a unit, if the input values from other units change.

**Multilayer organization**: ANNs are not isolated networks, but connect themselves with other processing units being thus complex systems. Units are disposed hierarchically in layers. Within this organization, the processing units can be of three different types:

- **Input units**: They send a specific value to other units of the system.
- **Hidden units**: They organize themselves in one or several intermediate layers. They receive the values from the input units and send them to the output units. Not all connectionists systems need hidden units.
- **Output units**: They receive values from other units of the system: the input units or the hidden ones, if they exist.

In table 6.4., the input units send a value to the output and hidden unit. The value inside the hidden and the output units is the threshold value.

In this network, the information is **feedforward propagated**, because the arrows are unidirectional. 6.5. shows the two kinds of feedforward propagated networks: *single layer networks*, those with just input and output units connected by a single layer, and *multi-layer networks*, those one or more layers of hidden units and thus with several layers of weights. The former is represented by the network on the left in 6.5., while the latter is the one on the right.
Connectivity pattern. This pattern is given by the weight, the value next to the arrows. In 6.4., the weight is +1 between the input units and the rest of the units, and –2 between the hidden unit and the output unit. The function of the weight is essential because:

a) It indicates the nature of the connection, which can be excitatory (red arrows) if the weight is positive or inhibitory (blue arrow) if the weight is negative. In our example, there is an inhibitory connection between the hidden unit and the output one (-2, blue arrow), being the rest excitatory (+1, red arrows). A value 0 between two units indicates that those are not related to each other.

b) It indicates the connection strength determined by the absolute value and establishes the connection degree of the units among themselves (independently of whether its connection is excitatory or inhibitory). In our network, the hidden unit and the output one are much more linked between themselves (absolute value 2) than the input units with the hidden and the output one (absolute value 1).

c) By means of its modification, the net is going to be able to learn a given task. The weight value is going to determine the connectivity pattern, which will determine what the system knows and how it will answer to the incoming signals. A simple connectivity pattern is shown in table 6.6., where 3 input units are linked to two output units (there are no hidden units in this net). Red lines indicate the excitatory connections whereas blue lines indicate the inhibitory ones.
6.2. Neuronal networks

6.2.1. Single layer networks: the perceptron

Once the basic structure of the ANNs has been described, it is necessary to study their functioning. Let us study a simple and historical network in the connectionist world: the perceptron. Its creator was Rosenblatt, who in 1958 proposed a pattern recognition system, in which the underlying idea is that of perceptive processing described in table 6.7.: a given object is first registered by the retina cells and later recognized by the corresponding brain cells; the retina cells are called input units. The signal goes through the respective projection and association area until an appropriate answer is provided.

The perceptron in this case is an ANN with two input units, one output unit and a layer of connection weights. Its task is to learn to recognize the input pattern in such a way that the output unit will be either on or off. The perceptron functioning is the binary proposal of McCulloch y Pitts. That is, all units answer in a discrete and binary way, with 0 when they are off (the inputs do not reach the threshold value) or with 1 when they are on (the inputs reach the threshold value).

Let us analyze an example. The formal representation of a perceptron is the one in table 6.8. The perceptron has an output unit with a threshold value 0 and two input units X1 and X2. It will have to learn to recognize connectivity patterns for the output units to be on or off.

Let us suppose that we want to check if with the threshold values our perceptron is able to
execute the inclusive disjunction OR (I'm going to the cinema with John or with Peter, or with both). The truth table for this logical task is the following:

<table>
<thead>
<tr>
<th>p</th>
<th>q</th>
<th>P OR q</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>

When an OR utterance is true if and only if p or q or both are true, being false in all other instances. If we transcribe this to a binary logic, we will have the following table (let us consider T = 1, F = 0):

<table>
<thead>
<tr>
<th>X1</th>
<th>X2</th>
<th>U. de salida</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

In our perceptron p and q are represented by input units X1 and X2, whereas p OR q is represented by the output unit. The task to be learnt by the perceptron is to activate the output unit when at least one or the two input units are on, being off in all other cases. The aim is to prove that the perceptron is able to carry out the logical OR task.

Thus, the weights are multiplied by the activation values of each input unit and the total input will determine the activation (1)/deactivation (0) of the output unit if the threshold level for the sum is reached or not. Let us study four simple examples:

1) Let us suppose that the two input units are ON (11). The activation value of each input unit is multiplied by its weight and the total input value.
   - X1 is ON (1) and its output is (1), so the output unit has the following input: 1(weight) x 1(activation value) = 1
   - X2 is ON (1) and its output is (1), so the output unit has the following input: 1(weight) x 1(activation value) = 1

   The sum of the total input is 1+1=2; and the value 2 reaches the threshold level 0, so the output unit will be ON (1). Thus the perceptron recognizes the first combination.

2) Let us suppose that X1 is ON and X2 OFF.
   - X1 is ON (1) and its output is (1), so the output unit has the following input: 1(weight) x 1(activation value) = 1
   - X2 is OFF (0) and its output is (0), so the output unit has the following input: 1(weight) x 0(activation value) = 0

   The sum of the total input is 1+0=1; and the value 1 reaches the threshold level 0, so the output unit will be ON (1).

3) Let us suppose that X1 is OFF and X2 ON.
   - X1 is OFF (0) and its output is (0), so the output unit has the following input: 1(weight) x 0(activation value) = 0
   - X2 is ON (1) and its output is (1), so the output unit has the following input: 1(weight) x 1(activation value) = 1

   The sum of the total input is 0+1=1; and the value 1 reaches the threshold level 0, so the output unit will be ON (1).

4) Let us suppose that both input units are OFF (00).

   In this situation, the output unit has the following input:
   - In A: 1(weight) x 0(activation value) = 0
In B: 1(weight) x 0(activation value) = 0

The sum of the total input is 0+0=0; and the value 0 doesn't reach the threshold level 0, so the output unit will be OFF (0). Thus, the perceptron also recognizes when both units are off.

However, this system that could solve the previous task could not solve tasks like the exclusive disjunction XOR (I'm going to the cinema with John or Peter, but not with both). An XOR utterance is true if and only if p or q are true, being false in all other cases.

<table>
<thead>
<tr>
<th>p</th>
<th>q</th>
<th>P XOR q</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>

If we transcribe this into a binary logic, we will obtain the following table:

<table>
<thead>
<tr>
<th>X1</th>
<th>X2</th>
<th>U. de salida</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

A perceptron is unable to solve this task because of linear geometrical problem. Thus, Minsky and Papert (1969) criticized the fact that perceptrons could only solve classificatory problems that were linearly separable, but many are not. So the perceptron is a very limited system for recognition and classification tasks.

### 6.2.2. Multilayer networks

The linear problem was solved with a multilayer network, as the one shown in table 6.4, formed by two input units, one hidden and one output unit. See how the network is able to solve the aforementioned XOR task with the given weights and threshold values. Let us take the exclusive disjunction XOR table; the output unit will only be on (1) if and only if one of the two input units receives a signal, but never if both of them do (0). Let us study the possibilities:

1) Both input units are on. The hidden unit receives the following input from X_1 and from X_2 respectively:

\[ [1 \text{ (weight) } \times 1 \text{ (activation value) }] + [1 \text{ (weight) } \times 1 \text{ (activation value) }] = 2 \]

This value reaches the threshold + 1.5, so the hidden unit will be on (1). The output unit will receive 3 inputs: from X_1, from the hidden unit and from X_2:

\[ [1 \text{ (weight) } \times 1 \text{ (activation value) }] + [-2 \text{ (weight) } \times 1 \text{ (activation value) }] + [1 \text{ (weight) } \times 1 \text{ (activation value) }] = 0 \]

This value does not reach the threshold value (+0.5), so the output unit will be off (0). So it receives two excitatory connections (X_1 and X_2) and one inhibitory of double value which impedes its activation.

2 and 3) Only one of the input units is on. Let us suppose that X_1 is on and X_2 is off. The hidden unit receives the following input:

\[ (1 \times 1) + (1 \times 0) = 1 \]

This value does not reach the threshold value (+1.5), so the hidden unit will be off (0). Hence, the output unit will receive the following inputs:

\[ (1 \times 1) + (-2 \times 0) + (1 \times 0) = 1 \]
This value reaches the threshold value (+0.5), so the output unit will be on (1). In this case the output unit receives only one excitatory connection which reaches the threshold value and so is on. If you do the calculus when $X_1$ is off and $X_2$ on, you will see that this connectivity pattern is also identified by the network.

4) Both input units are off. The hidden unit does not receive any input, as both input units are off:

$$ (1 \times 0) + (1 \times 0) = 0 $$

Thus, it does not reach the threshold value +1.5. The output unit will also be off:

$$ (1 \times 0) + (-2 \times 0) + (1 \times 0) = 0 $$

In conclusion, the output unit will neither be on when both input units are off, because all the input values are 0.

So a multilayer network does solve the exclusive disjunction XOR, which was one of the tasks the perceptron network could not do. This has been possible thanks to the hidden unit, which widen the computation possibilities by generating a new internal representation of the input data. In this example, it is the hidden unit the one that enables to solve this XOR task, because that unit is on if and only if both input units are on.

6.2.3. Feedback networks: Visual recognition

The multilayer ANNs perform feedforward propagation; that is, the signal is transmitted from the input units to the output ones. Besides, in the examples shown above, we have just considered one output unit, although depending on the task’s complexity we can introduce as many output units as necessary. The same goes for the input and hidden units.

However, there are other types of networks. In table 6.9 a net with two input, three hidden and two output units is presented. Notice that the output units resend a new signal to the input layer, creating thus some kind of feedback. This net is called recurrent neural network or feedback networks.

Feedback networks suppose a radical change to the feedforward propagation ones. In the latter the signal ends in the output unit, which constitutes the end of the chain, so they are stable networks and, as such, their behavior repertoire is limited: once the output is calculated, the process ends. However, feedback networks create loops which allow the signal feedback. Thus, they are dynamic networks, as the output is calculated according to the given input and feedback their signal to the input units; in this new iteration the input units modify their activation value and send a new signal causing a new output, which is again feedbacked to the input units. This is a dynamic process with two possible results: stability or instability. Feedback networks reach stability after several iterations, when the output units generate a stable output. But it may also happen that the output units continuously change their feedback signal; in this case, the net is in a permanent instability because the connectivity patterns are always changing.
Feedback networks offer the possibility of interaction. A feedback network is interactive when, at least, two units connect between themselves in a direct and bidirectional way. Interactive connections can be symmetric –if the weight value between two units is identical and thus, can be represented with one bidirectional arrow– or asymmetric –if the weight values are different, so two arrows are used for their representation– (see table 6.10.). However, not all feedback networks are necessarily interactive; feedback architecture without mutual interaction among its units is also possible, as shown in table 6.9.

An interesting feedback and interactive network is the one in table 6.11. Let us suppose that we create a network with a binary threshold value (1 or 0), but without the feedforward propagation. This type of network, in which units are completely interactive, and can interconnect and connect with themselves is called auto-associative networks. An example of an auto-associative ANN is the Hopfield network, in which units connect among themselves, except with themselves.
The most interesting aspect of this type of network is not only its highly simplified architecture (compared to the multilayer networks), but also its learning rules similar to the synapse mechanism.

**Visual recognition in auto-associative networks: identifying faces**

An example of an auto-associative network is presented in 6.12, in which ten different faces can be identified from four processing and interactive units detecting four different characteristics: Unit A distinguishes the type of hair; B, the eyes; C, the nose; and D, the mouth. The output values of the units are the following: +1 if a given characteristic is present, −1 if a specific characteristic is absent, 0 if the unit does not know. Thus, the first face means that A, B and C are on (+1) and D (−1). In the third face, the four units are (−1). And so on with the rest of the faces. The introduction of a third value is important because we can consider uncertainty in face recognition. Let us suppose for a minute that a unit in the net is uncertain and so is off (0). Hence, [0,-1,+1,+1] indicates that the network is executing a partial recognition, as it is not sure about the type of hair. Nevertheless, the network is able to provide an approximate answer. Imagine that you see a person on the street far away from you. You activate the information in your memory and immediately recognize her as your friend, Jane. She is far away but you have been able to recognize her, although your visual receptors in the retina have not sent so clear images as if you had come across her.
Have a look now at 6.13. The lion face in the central image is still recognizable regardless of the distortion.

6.13. Partial recognition, but ANNs can act effectively in these situations

Connectionists architectures can acquire such complexity levels as to store implicit information and recover it according to the input stimuli, even in partial or ambiguous informative situations. In order to understand this idea, the concepts of content-addressable memory
and adaptive resonance must be addressed.

**Content-addressable memory**

An auto-associative network acts as a content-addressable memory. In order to understand this concept, read the following sentence and try to contextualize it:

> To be or not to be: that is the question ...

If your literary knowledge is basic, you surely remember this sentence from *Hamlet* by Shakespeare. The surprising thing is asking oneself why a simple sentence of the book has activated in our memories the author’s name, the play, its reciting and even its continuation ...

> Whether 'tis nobler in the mind to suffer  
The slings and arrows of outrageous fortune,  
Or to take arms against a sea of troubles,  
And by opposing end them? To die: to sleep;  
No more; and by a sleep to say we end

An auto-associative network works in the same way as your memory: it is a content-addressable memory able to recover incomplete or ambiguous information.

**6.2.4. Winner-take-all (WTA) networks: Word recognition**

A WTA network is when all or part of the units try to get a higher activation degree than its partners. These are called *Winner-take-all networks*, because when one unit is fully activated (1), the others are completely off (0).

**Interactive activation and competition networks (IAC): word recognition**

It is common in WTA networks that competing units are *inhibitorily* connected among themselves, so they get a maximum activation value by inhibiting their adversaries. In this net, competition is combined with interaction among units. When this happens, the system is an interactive activation and competition network (IAC).

To illustrate the underlying idea in an IAC, McClelland and Rumelhart developed a classical connectionist work in 1981, as shown in 6.14.


The net’s goal is to recognize a four-letter word from a total of 1,179 English letters by means of excitatory and inhibitory connections among processing units. To understand this model, it is necessary to take into account that the ellipses have inside interactive processing units connecting with each other inhibitorily; whereas among layers there may be both inhibitory
(blue arrows) and excitatory (red arrows) connections. This IAC network is interactive because there is mutual connectivity (both excitatory and inhibitory) between the letter layer and the word layer and because within each ellipse all units are inhibitorily connected among themselves. The latter characteristic makes the network also competitive, because within a level and for each of the positions the units inside the ellipses compete among themselves to get a higher activation value than their partners, following the logic WTA.

Let us suppose that the model must recognize the word TAKE:

**Feature layer (input units):** In this layer the individual features of the individual letters are recognized: T, A, K and E. For each position and inside each ellipse, there are two units for each segment: a unit on for its presence and another for its absence. Thus, in this layer and inside the 4 ellipses there are 28 processing units. The units are inhibitorily connected with each other.

**Letter layer (hidden units):** Both the excitatory and inhibitory signals of the feature layer go into the letter level, where there are in each of the ellipses 26 processing units corresponding to each of the letters of the English alphabet, inhibitorily connected to get the maximum activation. The first position will detect letter T, the second A, the third K and the fourth E. The connections between the feature and letter layer are simple: for the first position, the feature layer detecting the horizontal line of letter T sends an excitatory signal to those letter units with this feature (B, Z, P...), and an inhibitory signal to those units without that feature (H, L, M...); the same would happen with A in the second position and E in the fourth one.

**Word layer (output units):** Word recognition is produced by the activation of one of the 1179 processing units of this third layer. The excitatory and inhibitory connections between the second (letter layer) and the third layer (word layer) will decide which word is recognized: for instance, letter unit T activated in the first position establishes excitatory connections with those word units which have a T at the beginning (TAKE, THIN ...) and inhibitory connections with those word units with any other letter (CAKE, LONG...). Besides, once the output unit is on (e.g. TAKE), this sends excitatory signals to the letter units detecting those in the respective positions and inhibitory signals to those not being detected: in our example, the activation of the word unit TAKE would send an excitatory signal to the letter unit T in the first position and an inhibitory signal to the rest of the units; it would excite the letter unit A in the second position and inhibit the rest of the letter units; and so on.

In conclusion, in such a multilayer connectionist system, word recognition is reduced to a PDP among the three layers, which establish excitatory and inhibitory connections and, within the same layer, units connect with negative weights to compete among themselves for the maximum activation value.

**ADAPTIVE RESONANCE**

The adaptive resonance supposes that, if there is noise during processing, the cognitive system will complete the information in the most possible coherent way to adapt itself to the environment (Grossberg, 1987). The auto-associative ANNs in uncertain or ambiguous situations are able to provide the most approximate answer.

Table 6.15. represents a classic (though adapted) example within connectionist models from McClelland, Rumelhart and Hinton (1986): it corresponds to the Jets and Sharks, two rival gangs of New York from the movie *West Side Story*. This is an IAC network that shows how the net stores information about some of its members. We will only consider 5 subjects: Rick, Art, Sam, Ralph and Lance. In the graphic, the irregular ellipses contain associated competitive units inhibitorily connected, just like blue arrows among all ellipses with negative weights: he is 20, 30 or 40 years; a thief, a bookie or a pusher; married, single or divorced; he belongs to JETS or to SHARKS; finally, he obtained Junior High School (JH), High School (HS) or College (COL). Units within the ellipses are inhibitorily linked, as they compete to get the highest activation value to classify a person as member of the Jets or Sharks. All arrows are bidirectional, so all linked units have identical weights.
### 6.15. Jets and Sharks IAC network

For instance, if the unit Ralph activated, it would inhibit the rest of the units of its ellipse (the other four members), but it would send excitatory signals to all the characteristics associated to this person: *he is 30, a pusher, single, belongs to the Jets and has gone to Junior High School.* Sam would activate other characteristics: *20 years, bookie, married, he belongs to Jets and has been to College.* Besides, if we activate the unit *married*, the net would tell us which members have that state. Furthermore, if we ask about any Sharks member around 30, the activation of *Sharks, 30* and all other excitatorily linked units would give us the output, *Rick.*

The adaptive resonance allows us to cope with uncertainty and with situations in which the information is partial. For instance, let us imagine that we want to get the name of a person who belongs to the Jets, is a bookie, married and has gone to Junior High School. As we can

<table>
<thead>
<tr>
<th>Nombre</th>
<th>Años</th>
<th>Estudios</th>
<th>Estado Civil</th>
<th>Ocupación</th>
</tr>
</thead>
<tbody>
<tr>
<td>Art</td>
<td>Jets</td>
<td>40's</td>
<td>J. H.</td>
<td>Sing.</td>
</tr>
<tr>
<td>Al</td>
<td>Jets</td>
<td>30's</td>
<td>J. H.</td>
<td>Mar.</td>
</tr>
<tr>
<td>Sam</td>
<td>Jets</td>
<td>20's</td>
<td>COL.</td>
<td>Sing.</td>
</tr>
<tr>
<td>Clyde</td>
<td>Jets</td>
<td>40's</td>
<td>J. H.</td>
<td>Sing.</td>
</tr>
<tr>
<td>Mike</td>
<td>Jets</td>
<td>30's</td>
<td>J. H.</td>
<td>Sing.</td>
</tr>
<tr>
<td>Jim</td>
<td>Jets</td>
<td>20's</td>
<td>J. H.</td>
<td>Div.</td>
</tr>
<tr>
<td>Greg</td>
<td>Jets</td>
<td>20's</td>
<td>H. S.</td>
<td>Mar.</td>
</tr>
<tr>
<td>John</td>
<td>Jets</td>
<td>20's</td>
<td>J. H.</td>
<td>Mar.</td>
</tr>
<tr>
<td>Doug</td>
<td>Jets</td>
<td>30's</td>
<td>H. S.</td>
<td>Sing.</td>
</tr>
<tr>
<td>Lance</td>
<td>Jets</td>
<td>20's</td>
<td>J. H.</td>
<td>Mar.</td>
</tr>
<tr>
<td>George</td>
<td>Jets</td>
<td>20's</td>
<td>J. H.</td>
<td>Div.</td>
</tr>
<tr>
<td>Pete</td>
<td>Jets</td>
<td>20's</td>
<td>H. S.</td>
<td>Sing.</td>
</tr>
<tr>
<td>Fred</td>
<td>Jets</td>
<td>20's</td>
<td>H. S.</td>
<td>Sing.</td>
</tr>
<tr>
<td>Gene</td>
<td>Jets</td>
<td>20's</td>
<td>COL.</td>
<td>Sing.</td>
</tr>
<tr>
<td>Ralph</td>
<td>Jets</td>
<td>30's</td>
<td>J. H.</td>
<td>Sing.</td>
</tr>
<tr>
<td>Phil</td>
<td>Shark</td>
<td>30's</td>
<td>COL.</td>
<td>Mar.</td>
</tr>
<tr>
<td>Ike</td>
<td>Shark</td>
<td>30's</td>
<td>J. H.</td>
<td>Mar.</td>
</tr>
<tr>
<td>Nick</td>
<td>Shark</td>
<td>30's</td>
<td>H. S.</td>
<td>Mar.</td>
</tr>
<tr>
<td>Don</td>
<td>Shark</td>
<td>30's</td>
<td>COL.</td>
<td>Sing.</td>
</tr>
<tr>
<td>Ned</td>
<td>Shark</td>
<td>30's</td>
<td>COL.</td>
<td>Mar.</td>
</tr>
<tr>
<td>Karl</td>
<td>Shark</td>
<td>40's</td>
<td>H. S.</td>
<td>Mar.</td>
</tr>
<tr>
<td>Ken</td>
<td>Shark</td>
<td>20's</td>
<td>H. S.</td>
<td>Sing.</td>
</tr>
<tr>
<td>Earl</td>
<td>Shark</td>
<td>40's</td>
<td>H. S.</td>
<td>Mar.</td>
</tr>
<tr>
<td>Rick</td>
<td>Shark</td>
<td>30's</td>
<td>H. S.</td>
<td>Div.</td>
</tr>
<tr>
<td>01</td>
<td>Shark</td>
<td>30's</td>
<td>COL.</td>
<td>Mar.</td>
</tr>
<tr>
<td>Neal</td>
<td>Shark</td>
<td>30's</td>
<td>H. S.</td>
<td>Sing.</td>
</tr>
<tr>
<td>Dave</td>
<td>Shark</td>
<td>30's</td>
<td>H. S.</td>
<td>Div.</td>
</tr>
</tbody>
</table>
see, there is no such person in our net. The most approximate person is Sam: a bookie, married person who belongs to JETS, but went to College, not to Junior High-School. In a computer simulation, the net would also answer Sam. The activation of bookie would also excite Sam, the activation of married would also excite Sam, the activation of JETS would excite Sam, Art, Ralph and Lance; finally, the activation of Junior High School would excite Lance, Ralph and Art. Thus, Sam receives a triple excitatory connection while Lance, Ralph and Art only receive a double connection; as there is competitive inhibitory connection between Sam and the rest of the units of the ellipse (Lance, Ralph y Art), the net would choose Sam as the most approximate answer. The net has been able to complete the provided partial information because it stores content-addressable connectivity patterns.

6.3. Connectionist models vs. symbolic models

Nowadays, when we talk about cognitivism we think of sub-symbolic or connectionist cognitivism, as many think that symbolic cognitivism should be called mentalism. What are the advantages of connectionist models vs. classic symbolic models? Connectionism shares with classic cognitivism the assumption that processing information and mind’s activity is their main object of study. Both pretend to create mind models, as both understand the mind as an information processing system. The differences lie in two points: connectionism describes the structure of the mind as interconnected processing units similar to neurons:

- **Classic (or mentalism):** algorithm and symbolic (or Online Processing)
- **Connectionism:** subsymbolic and connectionist (or Parallel Distributed Processing: PDP)

Connectionist models claim for a neural processing and have some features which distinguish them from classic cognitivism such as the following:

1) **PDP:** Online processing presents some difficulties when explaining the relative easiness and speed with which the mind performs several cognitive tasks. If we decompose a task step by step, we can perform it online or we can make all the elements of the system work in a simultaneous and cooperative way. This idea of a PDP is the main distinction between classic cognitive and connectionist models.

2) **Distributed representations:** Classic cognitivism represents knowledge in an analytical or analogical way. Connectionism does not use high-level symbolic formalisms such as propositions or images, but low-level subsymbolic mechanisms, which give birth to distributed knowledge representations. For instance, a distributed representation would be numbers in a chronometer or digital watch, formed onscreen by vertical or horizontal lines which can be on or off. When all are on, they represent number 8; i.e. the knowledge of a given number is a global distributed pattern among 7 units.

3) **No CPU needed:** While classic cognitive models need a CPU (central processing unit) to manipulate the stored symbolic information and apply a set of given rules, connectionist models do not a CPU. Information processing implies network activation and such activation is performed by many interconnected units.

4) **Superpositional memory:** Classic cognitivism understands memory as a set of stores with information mapping among them: sensory systems/registers, short-term memory and long-term memory. Connectionism considers memory as a superpositional system, because it is formed by many interconnected and superposed processing units called networks. The information is distributed among all units and thus, represented in all of them (adaptive resonance).
6.4. Advantages and disadvantages of connectionist models

1) Advantages:

- **Learning capacity:** Highly adaptive and flexible systems, able to learn by themselves with weight modification.

- **Neural inspiration:** Their implementation level is the brain, but they are mental theories, so their explanation power is more powerful than that of the symbolic systems.

- **Economy of representation:** A minimal number of processing units allows representing several knowledge types and relations among them. For instance, the perceptron allows representing several logic calculi.

- **Graceful degradation:** Information can be still present regardless of unit loss, as the rest of the units can establish new connectivity patterns.

- **Adaptive resonance and error tolerance:** ANNs can work in uncertain situation and recover incomplete information.

2) Drawbacks:

- **Low-spectrum models:** There is no general connectionist model that can comprise the research conducted in some given knowledge fields, as symbolic cognitive models have done (for instance, the modal model of memory).

- **Difficult to predict empirically:** Connectionist models are computer simulation models. Thus, it is difficult to establish hypothesis that can be empirically falsified, as the scientific method argues. Many connectionist models try to accomplish simulation with experimental data, but they have been criticized for carrying out tasks in online computers, when they are simulating parallel processes.

- **Learning paradox:** Learning capacity is paradoxically a great advantage but also a great limitation, as some learning methods are not real learning and performance mechanisms used by humans, i.e. they are far away from the physiological mechanisms that characterize the human brain when learning.
TEST 5: ARTIFICIAL NEURONAL NETWORKS: COMPUTATION AND LEARNING ALGORITHM

An artificial neural network (ANN) tries to simulate the structure and/or functional aspects of biological neural networks. The basic elements of the connectionist architecture are the processing units or artificial neurons, which are organized in networks characterized by a connectivity pattern. The computation that an ANN executes performs on the input values and applies a transformation. Understanding an ANN isn’t difficult. Have a look at the table 1.1. Two input units (A and B) are connected with an output unit. The value inside the unit is known as the threshold value. This value will determine whether this unit will be activated or not depending on the input it receives from the input units A and B. To determine its value, the weight is multiplied by the activation value of each input unit and the total input will determine the activation/deactivation of the output unit if the threshold level for the sum is reached or not. Let us suppose that the activation values of the units are binary: a value 1 means that the unit is activated and a value 0 represents that the unit is deactivated. The aim is to know the weight configuration among the different units, so if the input of the output unit reaches its threshold level, this will be activated (1) or will be deactivated (0) if it doesn’t reach the threshold level.

An example: two input units and one output unit
Let us have a look at the ANN in table 1.1. It is capable of executing the logical conjunction AND (p AND q) with the weights and the threshold level shown in the network. In other words, the output unit will be ON if and only if the two input units are ON, being OFF in any other case. The output unit uses a threshold activation type.

Table 1.1.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>Output unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

If all the inputs in this unit are higher than +1.5, it will be activated; if not, it will be deactivated. As shown in the table, the connecting weights between the input and output units are +1. In the table next to the network, the four possible input values are described with their corresponding outputs. There are four possible combinations for the input values: 11, 10, 01, 00. When both input units are activated (11), the output unit must be activated (1). However, when both input units are OFF (00) or just one of them is ON (10 ó 01), the output units must be OFF (0).

Two input units ON (11):
In this situation, the output unit has the following input:

- In A: 1(weight) x 1(activation value) = 1
In B: $1(\text{weight}) \times 1(\text{activation value}) = 1$

The sum of the total input is $1+1=2$; and the value 2 reaches the threshold level $+1.5$, so the output unit will be ON (1).

The unit A is ON and B is OFF (10):
- In A: $1(\text{weight}) \times 1(\text{activation value}) = 1$
- In B: $1(\text{weight}) \times 0(\text{activation value}) = 0$

The sum of the total input is $1+0=1$; and the value 1 doesn’t reach the threshold level $+1.5$, so the output unit will be OFF (0).

The unit A is OFF and B is ON (01):
- In A: $1(\text{weight}) \times 0(\text{activation value}) = 0$
- In B: $1(\text{weight}) \times 1(\text{activation value}) = 1$

The sum of the total input is $0+1=1$; and the value 1 doesn’t reach the threshold level $+1.5$, so the output unit will be OFF (0).

Both input units are OFF (00)
- In A: $1(\text{weight}) \times 0(\text{activation value}) = 0$
- In B: $1(\text{weight}) \times 0(\text{activation value}) = 0$

The sum of the total input is $0+0=0$; and the value 1 doesn’t reach the threshold level $+1.5$, so the output unit will be OFF (0).

**An example: two input units, two hidden units and one output unit**

Let us have a look at the ANN in table 1.2. It is a multilayer network with two input units (A and B), two hidden units (C and D) and one output unit. It is structurally a bit more complex than the previous one, but the logic of its calculation is identical to the previous one. Thus, for each unit, we will determine the inputs and outputs multiplying the weights by the activation values, to calculate the output.

**Two input units ON (11):**

Unit C has the following input:
- From A: $1(\text{weight}) \times 1(\text{activation value}) = 1$
- From B: $1(\text{weight}) \times 1(\text{activation value}) = 1$

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>Output unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
The sum of the total input is 1+1=2; and the value 2 reaches the threshold level +0.5, so the hidden unit C will be ON (1).

Unit D has the following input:

- From A: 1(weight) x 1(activation value) = 1
- From B: 1(weight) x 1(activation value) = 1

The sum of the total input is 1+1=2; and the value 2 reaches the threshold level +1.5, so the hidden unit D will also be ON (1).

Let us analyze what happens in the output unit, which receives two inputs, one from C and one from D. As both are ON, the output unit will receive the following input:

- From C: 1(weight) x 1(activation value) = 1
- From D: -1(weight) x 1(activation value) = -1

The sum of the total input is 1-1=0; and the value 0 doesn't reach the threshold +0.5, so the output unit will be OFF (0).

Unit A is ON and B is OFF (10):

Unit C has the following input:

- From A: 1(weight) x 1(activation value) = 1
- From B: 1(weight) x 0(activation value) = 0

The sum of the total input is 1+0=1; and the value 1 reaches the threshold level +0.5, so the hidden unit C will be ON (1).

Unit D has the following input:

- From A: 1(weight) x 1(activation value) = 1
- From B: 1(weight) x 0(activation value) = 0

The sum of the total input is 1+0=1; and the value 1 doesn't reach the threshold level +1.5, so the hidden unit D will be OFF (0).

Let us analyze what happens in the output unit, which receives two inputs, one from C and one from D:
From C: 1(weight) x 1(activation value) = 1
From D: -1(weight) x 0(activation value) = 0

The sum of the total input is 1+0=1; and the value 1 reaches the threshold +0.5, so the output unit will be ON (1).

Unit A is OFF and B is ON (01):
Unit C has the following input:
- From A: 1(weight) x 0(activation value) = 0
- From B: 1(weight) x 1(activation value) = 1

The sum of the total input is 0+1=1; and the value 1 reaches the threshold level +0.5, so the hidden unit C will be ON (1).
Unit D has the following input:
- From A: 1(weight) x 0(activation value) = 0
- From B: 1(weight) x 1(activation value) = 1

The sum of the total input is 0+1=1; and the value 1 doesn't reach the threshold level +1.5, so the hidden unit D will be OFF (0).

Let us analyze what happens in the output unit, which receives two inputs, one from C and one from D:
- From C: 1(weight) x 1(activation value) = 1
- From D: -1(weight) x 0(activation value) = 0

The sum of the total input is 1+0=1; and the value 1 reaches the threshold +0.5, so the output unit will be ON (1).

Both input units are OFF (00):
Unit C has the following input:
- From A: 1(weight) x 0(activation value) = 0
- From B: 1(weight) x 0(activation value) = 0

The sum of the total input is 0+0=0; and the value 0 reaches the threshold level +0.5, so the hidden unit C will be OFF (0).
Unit D has the following input:
- From A: 1(weight) x 0(activation value) = 0
- From B: 1(weight) x 0(activation value) = 0

The sum of the total input is 0+0=0; and the value 0 doesn’t reach the threshold level +1.5, so the hidden unit D will also be OFF (0).

Let us analyze what happens in the output unit, which receives two inputs, one from C and one from D:
- From C: 1(weight) x 0(activation value) = 0
- From D: -1(weight) x 0(activation value) = 0

The sum of the total input is 0+0=0; and the value 0 reaches the threshold +0.5, so the output unit will be OFF (0).

Learning processes
A network has learnt when it finds the appropriate connection weights. The output unit has the right answer according to the input data it has received. Thus, the network learns by itself performing a specific algorithm or rule. There are two basic learning processes: supervised and non-supervised processes. In this test, we will only deal with the former. Nevertheless, in both processes, the underlying idea is to modify the weights.
The delta or Widrow-Hoff rule consists in comparing the actual output to the desired results. The difference is called the error \( \delta \). Two kinds of errors are possible during learning: either the output units are off when they should be on or the output units are on when they should be off. So, according to the error made by the net, the delta rule offers the two following advices:

1) If your output unit is off when it should be on is because the net receives inputs with negative weights higher than the inputs with positive weights. Correction: increase the weights in a small quantity in a continuous way until the output unit is on.

2) If your output unit is on when it should be off is because if receives inputs with positive weights higher than the inputs with negative weights. Correction: remove a small quantity from the weights until the output unit is off (or add a small negative quantity).

These two advices are summed up in the following way:

When the net makes an error, pay less attention to the input unit making the error and more attention to the input units which haven't made the error.

How do we modify the weights? The difference between the desired value and the actual one \((t-^t)\) is the delta error \( \delta \). So we must increase or decrease the weights. Such a modification is shown by the following rule:

\[
\Delta w = n (t-^t) x
\]

where
- \( n \) = the number of examples
- \((t-^t)\) = the obtained delta error
- \( X \) = value of input units

When the net has learnt the rule, the values of the desired answer \((t)\) and the real answer \(^t\) are identical and their \( \delta \) will be zero. Thus, there will be no weight increase when the net provides the right answer. You might be thinking ... uff, it’s very difficult! But it’s not. Let’s see an example.

Let’s draw a net with three input units (A, B and C), connected to an output unit with a 0 threshold. Let’s suppose that we want the net to learn how to carry out the task in the following table, which is that the output unit is off when B and C are off, but is on in any other case:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>U. de salida</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

First point: If the weights are 0 and units A, B and C are off, will the output unit be on or off? Obviously, the output unit will be off, as all the products will be off. Then let’s begin the learning process.

1st task: if all the weights are 0, we take a sequence out of the four. Let’s presume that we take 111. With this sequence the output unit should be on (1). If we make the operation (weight x activation), we will have the following result:

\[(0x1) + (0x1) + (0x1) = 0\]
As 0 doesn’t reach the threshold value, the output unit is off (0), although it should be on (1). From this error, the net will begin to modify its weights according to the algorithm described in the delta rule:

\[ \Delta w = n (t^w - t) \times \]

The n value will be in this example 0,1. What is the error made by the net? The difference between the desired value (1) and the real on (0). What is the value of the input units? It is 1 for A, B and C. Let’s apply it for each of the inputs to obtain the increase for the weights of each of the connections:

\[ \Delta w_A = 0,1 \times (1-0) \times 1 = 0,1 \]
\[ \Delta w_B = 0,1 \times (1-0) \times 1 = 0,1 \]
\[ \Delta w_C = 0,1 \times (1-0) \times 1 = 0,1 \]

Put the new weights in the net that you drew. Now you will test that the net is able to carry out the task with the new weights.

**2nd task:** let’s suppose that we take the sequence 100. This means that A is on and B and C off, so the output unit should be off (0). With the weights 0,1, the net doesn’t provide the appropriate answer:

\[ (0,1 \times 1) + (0,1 \times 0) + (0,1 \times 0) = 0,1 \]

As this value is higher than the threshold 0, the output unit is on (1). A new error! So again it modifies the weights applying the delta rule:

\[ \Delta w_A = 0,1 \times (0-1) \times 1 = -0,1 \]
\[ \Delta w_B = 0,1 \times (0-1) \times 0 = 0 \]
\[ \Delta w_C = 0,1 \times (0-1) \times 0 = 0 \]

In this case, the net removes 0,1 from the weight connecting with unit A, maintaining the same weights for B and C. So the new weights are:

\[ w_A = 0,1 + \Delta w_A = -0,1 \]
\[ w_B = 0,1 + \Delta w_B = 0 \]
\[ w_C = 0,1 + \Delta w_C = 0 \]

With these new weights the net will be able to perform right the task.

**3rd task:** Choose the sequence you want. Let’s take 110, which implies that the output unit will be on (1). Let’s make the algorithms with these new weights:

\[ (0 \times 1) + (0,1 \times 1) + (0,1 \times 0) = 0,1 \]

This value reaches the threshold, so the output unit is on (1). As there is no error now and the delta value is 0, the net has at last learnt. Let’s check this out with the sequence 111, with which it made a mistake before:

\[ (0 \times 1) + (0,1 \times 1) + (0,1 \times 1) = 0,2 \]

This value reaches the threshold, so the output unit is on (1).

**Practice 1:**

Here you have one net. Next to it there is a table with three columns. In the first two columns all the possible activation states of the input units are represented. So you have to indicate in the third column the answer of the output unit (1 or 0), taking into account the weights and the thresholds indicated for each unit.
Practice 2:

Let’s apply the delta rule to a net able to categorize the faces represented in the following page. As you can see, the Rodríguez and the Fernández have some common features, so they could have had common relatives as well. Our net is going to learn to discriminate the members of each family.

The face of each member is composed by 5 characteristics: hair, eyes, nose, mouth and ears. Each of these characteristics can take the value +1 or -1: the former indicates the presence of a characteristic (a kind of hair, eyes, etc.) and the latter indicates that the alternative characteristic is the present one. 0 indicates that the unit is off and thus doesn’t recognize any characteristic. So a sequence like this:

+1+1+1-1-1

means that the person has straight hair (+1), lively eyes (+1), round nose (+1), wears lipstick (-1) and earrings (-1). So if we want an ANN to discriminate among the members of both families from these characteristics, the ANN will have 5 input units. Let’s suppose that (n) has always the value of 1 and the threshold of the output unit is 0.

1st task: Draw the net with 5 input units (A,B,C,D and E) connecting with an output unit with a 0 threshold.
2nd task: Let’s fill the following table, which constitutes the task that the net must learn. We will use just one output unit with a binary activation; so for a Rodríguez the output value will be (1) and for a Fernández, (0). We have just written the result for the first face for you.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>

3rd task: Which are the necessary weights for the net to discriminate the faces? To find them out, you must carry out a delta rule. We are going to provide you with the first repetition. The exercise is prepared so that in the third repetition you obtain a weight increase and in the fourth one you check that this weight increase is right to solve the task. The definitive weights will be \( w = (+1+1+1+1+1) \). So the important thing is to understand how you obtain them and why the net applies them.

The exercise is to be done in the table below. This table is organized in the following way. The first column shows the number of repetitions needed for the learning (only three will be necessary, as the fourth one is to check that the net learnt the task). The second column shows the weights connecting each input unit with the output one \((w)\); following the values of the input units \((x)\); \(i\) corresponds to the input value that the output unit receives; \(t\) is the desired value and \(^\wedge t\) the real answer of the output unit. Finally, \(n\) is the learning constant, which would be 1.

<table>
<thead>
<tr>
<th>Repetition</th>
<th>Weights</th>
<th>Input values</th>
<th>Desired value of o.u.</th>
<th>Real value of o.u.</th>
<th>Error</th>
<th>(\Delta n)</th>
<th>(\Delta A)</th>
<th>(\Delta B)</th>
<th>(\Delta C)</th>
<th>(\Delta D)</th>
<th>(\Delta E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>+1</td>
<td>+1</td>
</tr>
<tr>
<td>2</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>+1</td>
<td>+1</td>
</tr>
<tr>
<td>3</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Definitive weights

\( o.u. = \text{output unit} \)

Let’s describe the first repetition. At the beginning, all weights are 0 because the net is naïve and doesn’t know anything. We choose one face, for instance, the first face of the Rodriguez \((+1+1+1+1+1)\) and will write these values in the \((x)\) boxes. Let’s see how the output unit answers:

\[ i = (0 \times +1) + (0 \times +1) + (0 \times +1) + (0 \times +1) + (0 \times -1) = 0 \]

The obtained value 0 doesn’t reach the threshold, so the output unit will be off. Let’s fill in the table above. What is value with which the output unit should have answered? It’s a Rodriguez, so it should have been (1). Which is the real value of \(^\wedge t\)? It’s 0 because it didn’t reach the
threshold. Which is the error of these two last values? (1-0)= 1. Perfect! ... we can now calculate the increase to be applied to each weight following the delta rule:

\[
\Delta w_A = n (t-^t)x_A = 1 x (1-0) x +1 = +1
\]
\[
\Delta w_B = n (t-^t)x_B = 1 x (1-0) x +1 = +1
\]
\[
\Delta w_C = n (t-^t)x_C = 1 x (1-0) x +1 = +1
\]
\[
\Delta w_D = n (t-^t)x_D = 1 x (1-0) x -1 = -1
\]
\[
\Delta w_E = n (t-^t)x_E = 1 x (1-0) x -1 = -1
\]

So we write the increase (Δw) in the last five columns to calculate the new weights. Which are the new weights from Δw? We add the increase weights (+1+1+1-1-1) to the old weights (0 0 0 0 0):

\[
\begin{align*}
w_A &= 0 + \Delta w_A = +1 & w_A &= +1 \\
w_B &= 0 + \Delta w_B = +1 & w_B &= +1 \\
w_C &= 0 + \Delta w_C = +1 & w_C &= +1 \\
w_D &= 0 + \Delta w_D = -1 & w_D &= -1 \\
w_E &= 0 + \Delta w_E = -1 & w_E &= -1
\end{align*}
\]

So the new weights will be (+1+1+1+1-1) and we will write them in the (w) columns to check them in the second repetition. So the other values are for you to calculate them. The input values in the following repetitions are written for you. They will be:

2\textsuperscript{nd} repetition: +1-1+1+1+1 (corresponds to a Rodríguez)
3\textsuperscript{rd} repetition: +1-1+1-1-1 (corresponds to a Fernández)
4\textsuperscript{th} repetition: -1+1-1+1+1 (corresponds to a Rodríguez)

If your calculus is right in the fourth repetition the weight increase will be (0 0 0 0 0), as written in the table. The final weights to be obtained are (+1+1+1+1+1).

\textbf{4\textsuperscript{th} task:} Make sure that the net can discriminate among the 6 faces. In the following table, we provide you with the calculus for the first face. If your operations are right, you will see that the net perfectly discriminates the members of both families. What is the logic underlying the net? Simply looking at the positive or negative characteristics. If there are more positive ones the face will be considered Rodriguez, whereas the opposite means it is a Fernández.

| (+1x+1) | (+1x+1) | (+1x+1) | (+1x-1) | (+1x-1) = 1 (higher than the threshold, so output unit answers with 1 (Rodríguez). |
| (+1x ) | (+1x ) | (+1x ) | (+1x ) | (+1x ) = |
| (+1x ) | (+1x ) | (+1x ) | (+1x ) | (+1x ) = |
| (+1x ) | (+1x ) | (+1x ) | (+1x ) | (+1x ) = |
| (+1x ) | (+1x ) | (+1x ) | (+1x ) | (+1x ) = |
| (+1x ) | (+1x ) | (+1x ) | (+1x ) | (+1x ) = |

___ (higher than the threshold, so output unit answers with 1 (Rodríguez).
___ (higher than the threshold, so output unit answers with 1 (Rodríguez).
___ (lower than the threshold, so output unit answers with 0 (Rodríguez).
___ (lower than the threshold, so output unit answers with 0 (Rodríguez).
___ (lower than the threshold, so output unit answers with 0 (Rodríguez).
___ (lower than the threshold, so output unit answers with 0 (Rodríguez).
7. Information processing: selective control and motor activity

7.1. Previous setting: homunculi observing the world?

The great ghost of cognition has been the homunculus. Human cognition has been explained throughout the times by means of a homunculus or small man inhabiting in our mind, and perceiving and interpreting the world. The homunculus would select the information deciding what to pay attention to, what operations to apply during its manipulation and what answers to provide. The homunculus would be the main organ of our cognition.

Everything seems easy to explain if we admit the existence of a homunculus, but the problem arises when we ask ourselves who controls the homunculus, another homunculus inhabiting inside who in turn is controlled by another homunculus...? Definitely, cognition should run away from this type of hypotheses \textit{ad infinitum}.

In cognitive science stating that a theory is homuncular means that to explain a specific cognitive mechanism we appeal to that mechanism. How do we perceive the world? By means of a homunculus in our minds? Then how does he or she perceive what we perceive? By means of another homunculus in their mind? ... Notice that in this way we try to explain perception by means of perception itself: we perceive because another perceives for us, who in turn perceives by means of another, etc. Stating that knowledge is stored as a representation has made philosophers ask themselves about whether cognitive theories are homuncular. Knowledge is stored in our mind to be recovered; thus we infer that our conscious mind represents the external world. The idea of a computer that manipulates symbols has been considered a solution to this problem. Admitting mental representations on which to apply a set of cognitive operations, as the IP model proposes, is a hypothesis far away from the homuncular interpretation.

Recursive decomposition implies analyzing the IP flow to the level of the primitive elements, which Dennett called the \textit{army of idiots}. This army carries out some basic computational operations (yes/no; on/off, for instance, in an ANN- artificial neural network-) and does not
have any semantic knowledge. Thus, understanding the mind as a system processing, encoding, storing, recovering and manipulating information does not imply to admit a homunculus. According to Dennett, the mind is described, from a logical or functional perspective, as a software, independent from any physical substrate: a brain or any other artificial system that could reproduce it functionally, as a computer. Thus, the mind works like a computer program or an algorithm that decomposes progressively a problem in basic elements and that can be implemented in a physical substrate like the neurons or hardware like a chip.

In the following pages, we will study the implications of the processing control, that is to say, what systems or mechanisms are responsible for our cognitive activity. In a metaphorical way, the homunculus has been called the **eye of the mind** and other times it has been compared to a CPU or **mental** processor. We are going to study how we store information to provide the right answer in a given moment, so that our performance is the most correct one. Hence, we will consider first the idea of processing control, i.e. who controls, what is controlled and how is controlled. Then, we will show how some activities do not require control, allowing the participation of automatic mechanisms. Later, we will study the selective control of information, paying attention to attentional, mainly to how it is redirected and controls attention in space. These contents will lead us directly to the relation between cognition and action, analyzing the participation of the sensorial systems in information storing and its relation to the IP mechanisms. As most of the research conducted in this field lies within the vision domain, we have selected eye movements and pupil changes to study information processing.

### 7.2. Is all processing controlled?

According to classical cognitivism, our mind manipulates information by applying a set of cognitive operations on the information symbolically represented. The task is to discover which system or mechanism acts analogously in our mind. Before analyzing the information processing control, we should consider the following three questions: who controls IP, what has to be controlled and how it should be controlled.

The attention mechanism is essential during the IP control, especially when selectively used to identify relevant information and filter out the irrelevant one. The study of attention is not only complex because of the term and conceptual dispersion, but also because nowadays no one knows exactly what attention is or which cognitive systems or mechanisms we refer to when we say that we have paid attention in class or to an advertisement, that using the cell while driving catches our attention or that a subject in an experiment pays attention to an area of the visual field without moving the eyes nor the head. Some scholars claim that attention is not a single concept but a set of psychological phenomena (Neumann, 1996; Roselló, 1999; Styles, 1997); other researchers consider it as a Unitarian mechanism but composed by different modules (Tudela, 1991); some others even see it as a construct applicable to several characteristics of the human cognitive system (Parkin, 2001); finally, others try to avoid the use of the term **attention** as a psychological construct due to the diversity of phenomena it includes (Barsalou, 1992).

**What should be controlled?**

It is necessary to control the sensory systems which allow us to perceive the way in which the information is stored (sight, smell, touch, hearing or taste), as well as the proprioceptors which send us internal signs from our organism. It is also necessary to control the motor activities that ease the processing – such as head movements, eye movements, etc. – to the relevant information. A third domain where control should be exerted is the way in which information is stored in our memory and how it is then recovered. Finally, it is important to control the thought, i.e. the processing activity that allows us to reason and solve problems by
manipulating information.

7.1. Cognitive and motor dimensions to be controlled during processing (Barsalou, 1992).

<table>
<thead>
<tr>
<th>COGNITIVE CONTROL</th>
<th>PERCEPTION</th>
<th>ACTION</th>
<th>MEMORY</th>
<th>THOUGHT</th>
</tr>
</thead>
</table>

Thus, control requires the performance of mechanisms acting on perception, motor activities, memory or thought for information to be correctly selected and for the proposed task to be rightly carried out.

**How to control IP?**

IP requires the following:

- To select the relevant information for the task to be performed. This selection does not only imply to collect events or information from the external environment (sensory perception), but also from the inner one (information stored in the memory).
- To perform cognitive operations allowing information manipulation.
- To perform precisely and quickly the most appropriate motor answers in different environments.

Therefore, processing control does not only imply selecting information but also inner and motor activity.

Regarding the question of whether all processing is controlled, many scholars deny it. The dissociation between controlled and an automatic information processing is a classical one. In the former, the IP depends on the performances of an attention system which decides which information to select, how to manipulate it and which answers are the most appropriate, as it happens when we want to find a phone number in the phone book: We first search the surname alphabetically and then find the number. We remember it temporarily in the memory until we dial it. However, let us think for a minute in the following situation: for sure many days you tie or zip your shoes. The surprising thing is that, whenever you do it in the morning, your attention is directed to more relevant information than just tying or zipping your shoes, such as *I'm late, I should have got up earlier, I wish today were Saturday.*

Shiffrin y Schneider (1977) contrasted controlled and automatic human information processing. The main distinguishing characteristics between both of them are two: 1) automaticity does not require the participation of control nor attention; thus an automatic process "nearly always becomes active in response to a particular input configuration" and "is activated automatically without the necessity for active control or attention by the subject" (Schneider and Shiffrin, 1977:2). However, other activities (such as reading and understanding this text, finding a phone number in a phone book or doing an arithmetic operation) require the attention of a subject. 2) An automated activity cannot be interfered by other mental operations (Pashler, Johnston, Ruthruff, 2001).

One of the most traditionally used experimental tasks to show the dissociation of automatic and controlled processing is the *Stroop task,* not only used in experimental domains but also in many applied contexts such as the study of schizophrenia (Cabaco, Crespo, Castro y Pires, 2000) and of several emotional impairments (Fernández-Rivas y Cabaco, 2002; Martínez-Sánchez y Marín, 1997). A simplification of the *Stroop* task is the one represented in table 7.2. The task is apparently very easy. It consists in just saying the color of the words, not what the word says. Naming the color of the words in the control and congruent conditions does not generate any difficulties. However, the number of mistakes or reaction time in the incongruent condition...
increases due to the existing contradiction between the color and the word denotation.

7.2. Stroop effect

<table>
<thead>
<tr>
<th>Control</th>
<th>Congruent</th>
<th>Incongruent</th>
<th>Incongruent en Swahili</th>
</tr>
</thead>
<tbody>
<tr>
<td>XXXX</td>
<td>RED</td>
<td>YELLOW</td>
<td>JANI</td>
</tr>
<tr>
<td>XXXX</td>
<td>BLUE</td>
<td>GREEN</td>
<td>MANJANO</td>
</tr>
<tr>
<td>XXXX</td>
<td>YELLOW</td>
<td>RED</td>
<td>BULUU</td>
</tr>
<tr>
<td>XXXX</td>
<td>RED</td>
<td>BLUE</td>
<td>EKUNDE</td>
</tr>
<tr>
<td>XXXX</td>
<td>GREEN</td>
<td>YELLOW</td>
<td>MANJANO</td>
</tr>
<tr>
<td>XXXX</td>
<td>BLUE</td>
<td>GREEN</td>
<td>BULUU</td>
</tr>
</tbody>
</table>

Some theories have claimed that the difficulty of the phenomenon lies in the interference of an automatic process in a controlled one. Cognitively, the subject must pay attention to the color, requiring this dimension a controlled processing. Simultaneously, they must try to ignore the content of the word, something really difficult, as our reading habits are so automated that they interfere with the naming of the color. The negative effects of reading the words on color naming are drastically eliminated, when the language in which colors are presented is changed to an unknown one (such as Swahili). In this last condition, the automated word reading habits are absent because the reader does not know the language.

**Practice and automaticity**

Repeated practice enables that many behaviors demanding at the beginning a controlled information processing end up being automatic. Learning to swim is a behavior that begins being controlled to be progressively automated. Imagine that every time you walked your attention should be controlling how to situate your feet. Something similar happens with reading, writing or driving a car. The transition from a controlled to an automatic processing suggests that both types of processing are not as excluding as it could be thought at first. It is true that some tasks require an absolute control, but others are automatic and we are not aware that relevant information is being processed to carry them out. However, it is also true that control and automatic processing are simultaneously present in many behaviors in different degrees. Thus, some researchers have argued for the necessity of considering automatic and controlled processes two extremes of a continuum rather than exclusive domains (Pashler et al., 2001).

Nevertheless, there are circumstances where repeated practice, even in apparently simple tasks, does not facilitate their automatization, as there are IP limits which impede to pay attention to more than one stimuli simultaneously (this phenomenon is called bottleneck). One of them is, for instance, to control both hands independently, something really difficult. If you are ambidextrous, try to write the alphabet with one hand and the numbers from 1 to 20 with the other. Surely, you will find it impossible to do simultaneously because your attention will direct either to the letters or to the numbers. Both activities require the same motor mechanisms and so must share the same IP channel. Manual coordination requires that both hands are involved in the same task, as when you write with the keyboard: in this case the task is singular rather than dual, but it does not eliminate the so-called bottleneck.

**PDP and automaticity**

The connectionist models offer an alternative explanation to automaticity and control. The activity is explained by connectivity and distributed activation patterns. Let us have a look at table 7.3. It is a multilayer feed-forward network with 6 input, 4 hidden and 2 output units.

<table>
<thead>
<tr>
<th>red</th>
<th>white</th>
<th>answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>red</td>
<td>white</td>
<td>answer</td>
</tr>
</tbody>
</table>
Once the network has learnt, it determines the connection weights to simulate human behavior, making the same mistakes as human beings. In table 7.3., there is a multilayer feed-forward network which has learnt to establish the appropriate weights to simulate human behavior in this task. The input units must combine the ink color and the written word (a unit in red is activated/ON). Observe the network on the left: it corresponds to the stimulus RED. The right answer is WHITE (ink color) but the net answers activating the output unit RED. On the right, the stimulus is RED, which is the right answer now, as there is congruence between the color and the word. Therefore, this network has been able to learn to establish the appropriate excitatory (in red) and inhibitory (in blue) connections among the units to simulate human behavior. How has it achieved it? Simply, applying a differential degree of learning to the connections linked to the word rather than to the ones linked to color. Thus, we are simulating human differential learning, supposedly automated by the word reading task but requiring control of color naming.

7.3. Selective control

Our sensory systems have the ability to perceive and briefly retain an enormous quantity of information, even when presented at brief intervals of milliseconds. However, we all experience in a specific temporal moment that our conscience has only a minimal part of the information affecting our sensory systems or of the information stored in our memory.

Selection is very important during IP control. There are two basic positions: the most traditional one considers that selection is compulsory because there are limitations either in an online processing system or in the cognitive resources of the subject. The second position, the most recent one, claims that IP limits are a fallacy and considers that the function of selective control is to control behavior.

1) Selection for limitations in information processing

From the 20th century onwards, consciously paying attention to something implies recovering the relevant information and filtering out the irrelevant one. The underlying idea is that the information processing system has a limited capacity, because it must perform online operations, as a digital computer. Thus, it is necessary to assume the existence of a CPU that selects the relevant information and filters out the irrelevant one. This selection is called selection for processing (van der Heijden, 1996) or selection for perception (Roselló, 1999).

In this theory, selective control has been compared to a spotlight (spotlight analogy), which projects light to the areas to be visualized, leaving the rest in darkness. Hence, during IP the cognitive system selects the information relevant for the task. The selection is not only external, that is to say, our filtering mechanisms are not exclusively directed to the information from the external environment, as it happens when we pay attention to a traffic light in yellow or when we search in our messy wardrobe for a specific pair of pants. There is also an inner selection, as we only recover from our memory the information relevant for a task and perform a series of cognitive operations to get it, for instance the multiplication table for 7 in descending order. The normal thing is to combine both external and inner control, as it happens when we
make a puzzle: we do not only select the shape of the piece or its color, but also perform certain operations simultaneously to achieve the perfect adjustment of the pieces.

Another analogy to explain attention selection is the zoom, which focuses on a wider area by losing resolution of the details; and the other way around, the smaller the area to select the better we will perceive the fine details.

2) Selection for action/ action selection
These theories defend the existence of a selective control but consider that this control is not due to IP limitations. Their goal is directed to optimizing human performance by means of the activation and performance of the appropriate program and motor mechanisms. The absence of a selective control for action is shown when, lost in our thoughts, we do not realize that we have left behind our bus/ train stop or that we have addressed the email to ourselves rather than to the person we wanted to send it to. These theories have been an interesting change compared with the previous ones, because they do not try to explain the limits in the IP capacity of human beings, but to direct the interest towards action and performance by means of more appropriate answers.

The sensory systems have an enormous encoding capacity. However, action is limited and, usually, only one behavior can be performed by one motor organ. Thus, a selective control is performed on the motor program that starts a specific situation. Such an election will allow an optimal action, but if any problem arises during its selection it may happen that we put dirty clothes in the rubbish bin instead of in the washing machine. Thus, selective control requires establishing an appropriate correspondence between the sensory input and the motor output.

Is selective control strictly selective?

Distinguishing relevant from irrelevant information can be an easy or a difficult task depending on the situation. The dissociation between separable and integral dimensions has been studied for years (Garner, 1974). Notice table 7.4., in which there are four stimuli created from the combination of two dimensions: color and shape. Let us suppose that they are presented visually for subjects to discriminate the color by pressing a button when the figure is red and another when it is blue– or the shape by pressing a button if it is a square or another if a circle. These studies have proved that the controlled dimension is not interfered by the ignore one, i.e. one can classify shape ignoring color, or the other way around; so subjects are able to selectively pay attention to the display dimension and ignore the alternative. When this happens, dimensions are separable because they can be perceived independently from one another.

However, let us observe saturation and gloss in the lower part of table 7.4. When subjects are asked to make separate judgments about each dimension, they have great difficulties in classifying them independently, because both dimensions are perceived holistically, i.e. together. When this happens, dimensions are integral.

7.4. Selective attention is possible with shape and color, not with saturation and gloss
Separable and integral dimensions are two poles of a continuum in which sensory dimensions can be located according to different degrees.

### 7.4. Attentional control

Now we are going to study different ways in which attention can vary across space, how attention can be stimuli- or goal-driven.

#### 7.4.1. Overt and covert attentional orientation

During IP, the visual receptors are not always directed to the display stimulus. If you are walking along the street and hear a loud noise on your back, you may probably turn around and direct your attention to the event. However, if you are chatting with your friend and staring at her/him, your attention can be directed to her/his child and the naughtiness he/she is about to make. In the first example, your attention is overt while in the second one it is covert. In other words, information processing does not always require the sensory receptors to be oriented to the relevant event, because attention can be covertly directed to the relevant stimulus. Thus, movement of visual attention and eye movement are dissociable.

#### 7.4.2. Stimuli- and goal-driven attentional control

The distinction between a top-down and a bottom-up processing is that, in the former, the subject processes the information according to the environmental stimuli relevant for their goals; whereas in the latter the stimuli affecting the sensory receptors control and direct the subject’s attention without activating any intentional mechanism. Top-down processing is goal-driven control while bottom-up processing is a stimulus-driven control. Behavior can be set along this continuum which goes from selecting a thought without being directly elicited by an external stimulus to a simple reflex.

<table>
<thead>
<tr>
<th>Bottom-up (direct signs)</th>
<th>Exogenous, involuntary extrinsic attention control or Stimulus-driven</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top-down (symbolic signs)</td>
<td>Endogenous, voluntary intrinsic attention control or Goal-driven</td>
</tr>
</tbody>
</table>

The distinction between these two types of processing has been studied in different ways (Jonides, 1981; Posner, 1978):

1) Exogenous vs. endogenous: they refer to the sign that starts an attention movement which cannot be easily inhibited. Let us consider the noise example again, in which the
noise catches our attention in an overt and involuntary way and makes us probably turn around to see what has happened.

2) involuntary or extrinsic vs. voluntary or intrinsic: they refer to the symbolic nature of the sign and, as it requires a semantic interpretation, the attention movement can be modified once it has begun or even eliminated. For instance, it may happen that the noise that we have heard belongs to a car or a petard, but once we recognize the noise we decide not to direct our attention to it and keep on walking.

If the noise is not conceptually identified and it involuntary controls our attention, then the attentional control is **exogenous, involuntary or extrinsic and stimulus-driven**. If the noise is conceptually interpreted and we voluntary decide not to pay attention to it, then the control is **endogenous, voluntary or intrinsic and goal-driven** (Wright y Ward, 1998).

### 7.5. Information processing: the sight

Sight represents the main information source for human beings. Thus, the link between IP control and action will be studied according to the indicators of cognitive activity in the visual sensory system: saccades, visual fixations and pupil size.

The study of visual processing control is closely linked to the analysis of the motor activity shown in the visual behavior. Paying attention to a specific point in the visual field does not necessarily imply an eye movement, because a covert attentional movement with the eyes stable on a fixation point may take place. However, in most situations of daily life, there is a high correlation between directing attention to a specific stimulus and moving the eyes simultaneously. This dissociation between **attentional movements and eye movements** has been carefully analyzed scientifically to find out the relations between both mechanisms during IP control. Three eye parameters have been used to measure the visual activity during IP: saccades, visual fixations and pupil size.

There is a great variety of eye movements (see table 7.6.), though the saccades have been the most frequently used to analyze the relations between eye movements and attentional control during processing. The saccadic system has two different motor parameters: the saccadic movement (or fast movement of the eye) and the visual fixation (i.e., the maintaining of the visual gaze on a single location). Both are closely link. The saccades are quick, simultaneous movements of both eyes in the same direction whereas visual fixations do not only capture information but also decide where to look at next in the visual field. The graphic representation of the exploratory visual behavior of a subject on a stimulus or image is called **scan path** (see image 7.7.).

#### 7.6. Typically eye movements

**CLASSICAL TYPOLOGY OF EYE MOVEMENTS**

**Saccadic movements**
Fast movements of the eyes closely related to attentional control. One reason for the saccadic movement of the human eye is that the central part of the retina—known as the fovea—plays a critical role in resolving objects. By moving the eye in a way that small parts of a scene can be sensed with greater resolution, body resources can be more efficiently used.

**Conjugate eye movements**
Those preserve the angular relationship between the right and left eyes. For example, when you move both eyes left and then right, a conjugate movement is made. Up and down movements and combinations of vertical and lateral movements also fall into the conjugate category. Their goal is to bring stability to the images in the retina due to the head and trunk movements as we walk.

**Vergence or disjunctive eye movements**
Those are the ones where the angle between the eyes changes. For example, hold your index finger in front of your nose. Now move your finger towards and away from your face, all the while fixating on your finger tip.
As you move your finger closer to your face, both eyes will converge; as you move your finger away from your face, your eyes will diverge. Their function is to try to make the image as clear as possible.

Smooth pursuit movements
The eyes move smoothly instead of in jumps. They are called pursuit because this type of eye movement is made when the eyes follow an object. Therefore, to make a pursuit movement, look at your forefinger, at arm's length, and then move your arm left and right while fixating your finger tip. They are not voluntary movements and their goal is to stabilize the visual image in movement in the retina. However, it is possible to control them with training.

Nystagmus
The vestibular system has a profound influence on eye movements. You can prove this to yourself by asking a friend to sit on a rotating chair. Before you spin the chair, ask your friend to look at some object opposite them, perhaps a picture on the wall. Note that your friend's eye will be relatively steady. Now spin your friend for 30 seconds or so and when you stop the chair, tell them to try to fixate the same target again. You will note that your friend's eyes will not be steady rather they will be moving back and forth. This reflex is used to stabilize an image on the surface of the retina during head movements.

7.7. Scan path
As it can be seen in this photo, the continuous lines represent the saccades. The saccadic amplitude is the distance between one fixation and the next one. It is measured in degree visual angle; thus, the longer the line is, the bigger the saccade has been. It shows us the visual field extension. Some studies have proved the relation between the amplitude of the visual field explored and the cognitive requirements of a given task, so an increase in the requirements causes a narrowing of the field (Crespo, 2002). A scan path is a sequence of eye fixations made when someone samples a visual scene. **Put simply, a scan path is the path our eyes follow when looking at an image.** It helps to interpret the exploratory behavior during IP. The controlled eye movement is different depending on the task required: reading, visual search, etc.

### 7.5.1. Saccadic activity, visual fixations and pupillary activity

#### 7.5.1.1. Basis of the saccadic activity

Saccades are fast eye movements whose aim is to dispose stimuli in the fovea, which is the area of greatest visual resolution. The bigger the saccadic amplitude, the faster the saccade will be. For instance, a saccade of 30° can reach a speed of 600º/sg. Maintaining the head steady the area of a saccadic movement is of 30°.

Another parameter related to saccades is their duration, also related with the saccadic amplitude: a saccadic movement of 30° takes 100 milliseconds.

A third parameter is the latency, i.e. the time from the target onset to the beginning of the saccadic movement. The latency in standard conditions (that is, the subject looks at a point but redirects their look with the stimulus onset) oscillates between 175-200 ms.

The saccades can be voluntary or involuntary. An **involuntary saccade** is highly automated and implies directing the eyes to the stimulus. For instance, when hearing a loud noise, you direct your eyes to the place it comes from. A **voluntary saccade** is goal-driven and is the
person who activates the appropriate mechanisms to perform a movement to direct the eyes to the area of interest, as when you look at a picture or a photo.

**Attention and saccadic control**

The relations between attentional changes and eye movements in IP have generated two different theoretical positions: ones defending an independent control system and others proposing a unique control system.

**Independent control models**

It has been proved that the eyes can move to a point in space before the attentional change has taken place. What is more, subjects can maintain the attention in a specific point and perform a saccade in the opposite direction. This has led scholars to think of two independent systems: one involved in maintaining the attention in space and the other in charge of preparing and performing the saccadic activity.

**Shared control models**

These models defend just the opposite: that there is a close relation between saccadic behavior and attentional changes, and claim that attention and eye movement depend on a shared control system as, according to some experimental results, it is not possible to maintain the attention on a specific spatial position while the eyes move to a different one. These models have more supporting experimental results.

**Saccade suppression**

An important aspect of the saccadic movement is that information withdrawal is only produced during the periods of visual fixation, not during the saccade, when the image disappears from the retina and access to new information is impeded. This is called saccade suppression because information withdrawal is not possible during the saccadic movement, allowing meanwhile stimulus changes without the subject being aware of it.

However, the fact that perception is continuous (although information withdrawal is discontinuous) indicates some kind of integration between the information from the preceding and the following saccade. Thus, there must be a storing system that keeps information in the memory.

**Change blindness**

It is the phenomenon that occurs when a person viewing a visual scene apparently fails to detect large changes in the scene. For change blindness to occur, the change in the scene typically has to coincide with some visual disruption such as a saccade (eye movement) or a brief obscuration of the observed scene or image. When looking at still images, a viewer can experience change blindness if part of the image changes. For example, change blindness can occur when a blank screen is inserted between the original and changed image.

Find out the differences between these two images
Three hypotheses have been proposed to explain this phenomenon: 1) Information masking takes place, so image A is masked by image A’, eliminating all its visual characteristics and leaving only an abstract representation. 2) The first impression hypothesis argues that subjects encode A accurately, extracting its abstract meaning. When A’ is presented, it is not necessary to process the details again, because the abstract meaning is maintained. A has generated a first impression which is difficult to modify with A’, as their abstract meanings are the same. 3) It claims that everything is perceived but nothing compared. Both A and A’ are fully processed; however, in order to detect any change, a comparison mechanism must be activated by asking if there is any difference between the two images.

Change blindness proves that the IP system is limited and that only part of the total information to which attention is directed is processed.

7.5.1.2. Visual fixations

Visual fixations (i.e. the maintaining of the visual gaze on a single location) occur between the saccadic movements. The eye is never static and some micro-movements such as flicks, drifts and eye micro-tremor take place. Those micro-movements are determined by two parameters: eye-fixation duration and the number of fixations. Fixations are linked to a point in time and space of focus (called area of interest), as they indicate where is the person looking at and how long they have been looking.
According to some studies, the first image fixations (taking place the first and a half second, approximately) are not directed to the most informative areas of the scene (ie. its semantic aspects), but to specific perceptive characteristics. Probably, this first exploration serves to create a first visual representation of the explored space that will later allow an informative analysis, where fixations will be directed to the most significant areas for the observer. The most informative areas generate longer lasting gazes, related with the conceptual or semantic interpretation that the subject performs.

Maintaining the visual gaze on a single location, the necessary time for visual processing is of 100 ms (milliseconds), although 50 ms allow processing 75% of the information. In exploratory eye movement tasks, the minimal duration of the cognitive processing in the fixation period is of 100-150 ms.

The sequential effects of visual fixations are the following: if in the previous fixation enough information has been gathered to take a cognitive decision, then the duration of the following fixation will be shorter. In the milliseconds of the cognitive processing phase of the fixation, many decisions are taken: a) selecting the relevant information; b) exploring the periphery by gathering the general features; c) planning the next saccadic movement.

### 7.5.1.3. Pupillary activity

The pupil is the dark aperture in the iris that determines how much light is let into the eye. The pupil also has the function of eliminating marginal light rays improving visual quality. Pupil size is between 2 and 7 mm.

There are optical changes such as the pupillary light reflex (a reflex that controls the diameter of the pupil in response to the intensity of light that falls on the retina. Greater intensity light causes the pupil to become smaller (allowing less light in), whereas lower intensity light causes the pupil to become larger (allowing more light in). Thus, the pupillary light reflex regulates the intensity of the light entering the eye) and the accommodation reflex (a reflex action of the eye in response to focusing on a near object, then looking at a distant object and vice versa, comprising coordinated changes in vergence, lens shape and pupil size). But the most interesting variations are related with emotional events, in which the larger size of the pupil is interpreted as an increase in the cognitive processing requirements or a way to cope more efficiently.

In pupilometric studies about different aspects of language processing (such as letters, syllables, words or sentences), about reasoning tasks (such as arithmetic) or visual discrimination, it has been observed that the pupil becomes larger as the task difficulty increases.

<table>
<thead>
<tr>
<th>Pupillary dilatation</th>
<th>Memory</th>
<th>Language</th>
<th>Reasoning</th>
<th>Perception</th>
</tr>
</thead>
<tbody>
<tr>
<td>0,6</td>
<td></td>
<td></td>
<td>Difficult multiplication</td>
<td></td>
</tr>
<tr>
<td>0,5</td>
<td>7 digits</td>
<td>Grammatical reasoning</td>
<td>Average multiplication</td>
<td></td>
</tr>
<tr>
<td>0,4</td>
<td>6 digits</td>
<td>Difficult word match</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0,3</td>
<td>5 digits</td>
<td>Easy word match</td>
<td>Easy multiplication</td>
<td>Difficult discrimination</td>
</tr>
<tr>
<td>0,2</td>
<td>3 digits, 4 digits</td>
<td>Sentence encoding</td>
<td>Auditory discrimination</td>
<td></td>
</tr>
<tr>
<td>0,1</td>
<td>1 digit, 2 digits</td>
<td>Letter match</td>
<td>Remembering the multiplicand</td>
<td>Auditory discrimination and visual detection</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**7.5.2. Cognitive aspects related to visual recognition**
IP is closely linked to memory. Information processing is understood as an analysis continuum between the most superficial and the deepest levels:

- The surface levels of processing correspond to the sensory or physical information analysis while the deep ones correspond to the semantic or conceptual level.
- Analyzing deep level information requires much more time than analyzing it superficially.
- The lasting of information in memory depends on the depth degree of information processing. Thus, the deeper the analysis, the likelier it will be to remember it later on.

7.6 Registros sensoriales en el procesamiento de la información

Sensory register (SR) is the short-term memory store for senses, which holds large quantities of information, but only for milliseconds. Each sense has its own register, although the visual SR has been the most studied one. SRs are closely linked to perceptive and attentional IP mechanisms. Nowadays, many scholars consider that SRs are early forms of information withdrawal related to perception or attention more than memory stores. SRs, specially the iconic one, retain information for less than a second.

7.6.1. The visual sensory register: basic properties

The visual SR seems to have an enormous informative capacity, virtually unlimited, although its quantification can only be determined relatively to the set of stimuli. Experimental work has proved in very brief visual presentations the existence of an icon or visual sensory track, which is the informative persistence of a stimulus once this has disappeared.

Traditionally, this SR has been considered an iconic memory store of great capacity, very short duration and pre-categorical encoding. It was a unique phenomenon whose decay began with the stimulus offset, just like an image fading out onscreen. However, lately the idea that the SR is not a unique and periphery iconic memory store prevails, as there seems to be another component involved during the sensory information processing. This implies a multiple visual SR.

The fact that the subject always remembers an item even in the cases of poor performance points to the idea that, during icon persistence, the subject extracts information and sends it to a new memory system called short-term conceptual visual memory (Potter, 1993). Its characteristics are the following:

- It is a processing and memory system different from the iconic memory system.
- Its main function is to act as a temporal store where information is processed and classified by categories.
- Non-stored information decays and is quickly forgotten, even without the subject being aware of its presentation.

Information mapping between the iconic memory and the short-term conceptual visual memory in the multiple visual SR can be of two kinds: selective and non-selective. A selective mapping is the mapping of relevant information for the task, which facilitates remembering; this mapping takes place once the stimulus has been presented, because the subject directs their attention to the information to be sent to the short-term visual memory. A non-selective mapping is the mapping performed at the stimulus onset, as the subject maps information randomly.

Functionality of the visual sensory register
The iconic memory is the store retaining information to preserve the phenomenal continuity that characterizes vision, as our perception is discrete, not continuous. Thus, during the saccadic suppression period with the stimulus offset, the iconic memory retains the fixation information until the next fixation. It is undeniable the adaptive function of the visual sensory memory in the mechanisms involved in the animal adaptation to the environment, such as perception, providing continuity to it.
TEST 6: EYE MOVEMENT AND PROCESSING CONTROL

Goals
This practice’s goal is that you familiarize with the oculography register techniques, which allow us to get data to validate IP theories. In this practice you will manage a set of registers from which you will have to get some information and to drop some conclusions.

The practice has two exercises. The first one consists in reproducing an example like the one you are given to acquire some basic knowledge on experimental procedures. In the second exercise you will work with the data from subjects to contrast the assumption that we propose.

Exercise 1
Imagine that you are interested in art and in analyzing specific elements from a picture that can call people’s attention. Imagine that you move to the Prado Museum an eye tracker and select some subjects to look at the picture of Las Hilanderas. Once the data of the eye tracking have been registered, you go back to your lab and begin to work with them.

In this first exercise, you will learn to:
- Represent visual fixations on a stimulus and to draw a scan path.
- Represent the AOI (areas of interest).
- Calculate the number and duration of the fixations for each AOI.
- Calculate the transitions among the different AOI.

In order for you to learn to do all this, we are going to do it previously on an image, so that you can do exactly the same on a different one. Thus, we will now work exclusively with the tables 1 to 7 and the image of Las Hilanderas. Look at this image. The continuous lines on it correspond to the saccadic movements, while the red points represent the visual fixations. This visual representation of the eye behavior on an image is called scan path. The lines of the scan path are a simple visual representation: the subject begins her exploration paying attention to the center of the image (fixation 1). Then, moving her eyes almost horizontally, she pays attention to the two spinners on the left of the picture. Later on, she feels curiosity for the contrasting figure on the upper part of the picture, the circular window (fixation 4) to then look at the distaff (fixations 5 & 6). Again she has looked at the spinners on the left to then pay attention to the spinners on the right (fixations 8, 9 & 10). Afterwards, she has interested herself in the background of the scene (fixation 11). Finally, she has returned to the foreground and fixed her eyes on the spinner in the middle ground (fixation 12). The analysis of a subject’s scan path gives us very interesting information to interpret the exploration process and why it is done so.

Have a look at Table 1 (Appendix). Once you have looked at the scan path its data are intuitive. The 1st column is the fixation order, from 1 to 12. The 2nd is the duration of each fixation. In order to determine the existence of a fixation, the eye must be fixed for at least 100 milliseconds. The average duration of an ocular fixation is between 250 and 300 milliseconds, though it can be longer or shorter. The 3rd and 4th columns indicate the horizontal (x) and vertical (y) position respectively where the fixation has taken place. In order to represent it, a squared pattern has been drawn on the image. Observe which line represents each coordinate (the nearest to the number; horizontally, the line underneath and vertically, the one on the left). The real squared pattern is much wider; so, to simplify the practice, we have transformed the values in an 22x15 interval (notice that we begin with [0.0] in the origin) for this example. If you happen to find a value such as 12.5, you will have to understand that the fixation is located in the center of the square, between the values 12 and 13.

Notice that in the image there are also 6 discontinuous squares in yellow color. These squares define the areas of interest: one subject may pay attention to the figures in the background; another may be interested in the role of the human figures compared to the objects to perceive a scene; or a painter may show interest in way in which colors call subjects’
attention. Those areas are called **areas of interest (AOI)**. Their numeric representation is described in Table 2. As it is an area, more than just one point will be necessary to define it, so that each of the 6 areas will be defined by the coordinate \((x,y)\) of the upper left corner and the coordinate \((x,y)\) of the lower right corner. When the AOI are defined, we also define by exclusion the off areas, i.e. the rest of the possible locations not included in any of the defined areas. These areas indicate when the ocular events have happened off any of the AOI. Notice how the look of the subject has fixed on the AOI that we have defined, either by their informative value, either by the kind of color used, etc.

Table 3 is a summary table with the fixations for each AOI. It shows us how many fixations have been produced and how much time the subject has been looking at the corresponding area. It is calculated by adding the duration of each of the fixations produced in that area. For instance, in the area A, there are 3 fixations: 2, 3 & 7. Hence, to calculate the total duration on the area, the duration of each of these fixations has been added.

Another interesting aspect can be how the information of an AOI directs the look to another specific area. This information is given by Tables 4 & 5, called **tabla de residencia de fijaciones** y **tabla de transición** respectively. The **tabla de residencia de fijaciones** or fixation residence table represents the fixation order and the AOI that given fixation is in. In the next column the transition is represented. A transition consists in the movement from one area to another in which the fixations are produced. For example, if fixation 4 is produced in the area C and the next fixation is produced in area B, then a transition has been produced between the areas C & B (CB). In Table 5 this information about transitions is represented in an orthogonal table. As the 1st fixation does not have a previous referent, the total transitions will always be the fixation number minus one.

Table 6 exemplifies the calculus of the **combined probability**. The combined probability is the relative transition frequency between two areas compared to the transitions between any two other areas. It shows us the relative transitional value between two areas. Somehow, this value informs us of the transition hierarchy, letting us know the most repeated transitional value from one area to another. In our example, the subject has a combined probability of 0.182 on the area F. These transitions have been done within the same area, which may indicate that this area requires a deeper exploration to be interpreted, as it is intensively processed before moving onto another area. The formula for the probability is the following:

\[
P(i\cdot j) = \frac{\text{Number of transitions between area } i \text{ and area } j}{\text{Number of total fixations } - 1}
\]

For instance, the combined probability from moving from area C to area B is done as follows:

I look at Table 5 and get the transition from area C (column) to area B (row) = 1. Then I look at the total sum in Table 5 = 11 (which is the same as resting 1 to the number of total fixations \([12-1=11]\)).

\[
P(C\cdot B) = \frac{1}{11} = 0.091
\]

Table 7 shows the calculus of the **conditional probability**, which is the probability that, given a fixation in a specific area \((i)\), the next fixation will occur in another specific area \((j)\). It is calculated as follows:

\[
P(i\cdot j) = \frac{\text{Number of transitions from area } i \text{ to area } j}{\text{Number of total fixations in the area}}
\]

So the conditional probability of moving from area F to area D is calculated by looking at Table 5 and looking at the transition from area F (column) to area D (row) = 1. Then I look at the sum of this row = 3.

\[
P(F\cdot D) = \frac{1}{3} = 0.333
\]
We have to be careful with how we interpret the data of the conditional probability. Let us imagine a scene with 5 AOI. Let us suppose that the subject has made 200 fixations all together. From these fixations she made just one fixation in area 2 and the next fixation is in area 3. Then the conditional probability equals 1 (or 100%). But this does not imply that there is much visual movement between area 2 & 3. It only means that whenever there has been a fixation in area 2, the next one has been in area 3. This parameter indicates us indirectly the informative relation that two areas of a visual scene may have. In a newspaper, for instance, the area at the foot of the image and the image to which this refers will tend to have a high conditional probability, as after observing the image, it is frequent to produce fixations on the text to better understand the visually processed information.

Now it’s your go!

Up to now we have described the basic concepts of ocular movements by means of the ocular registers. Now you will have to reproduce all the steps we have described to familiarize yourself with the concepts and the methodology. Suppose that you have gone to the Prado Museum and there you have registered the exploration on the picture *La Fragua de Vulcano*. The data are given in Table 8 and the AOI in Table 9. You also have a squared pattern with the *Fragua de Vulcano* to do a scan path and to draw the AOI on it. Afterwards, you will have to fill in Tables 10 to 14.

Exercise 2

The Art and Sciences Foundation has been given a funding from the European Community to do a I+D project on scientific conclusions about the perceptual aspects of Velázquez picture *La Fragua de Vulcano*. One of the first tasks will be to determine the informative relevance of the picture areas. You are hired as a specialist and decide to apply oculography and pupilography techniques on the picture exploration. Thus, you propose the following hypothesis:

*If the right area of the picture is more informative to understand the story represented in the picture, this will be visually more intensively explored and will receive:*

1. A higher number of fixations.
2. The fixations will be longer.
3. The average pupil diameter in that area will be wider.

To verify your hypothesis, you have registered the ocular parameters of 35 subjects which saw the aforementioned picture. The register shows 3 dependent variables: the fixation number, the fixation duration and the pupil diameter. **The division between left and right AOI is simple, as the pictured is divided into two halves. The left AOI will be called A and the right one B.** Table 15 shows these data. Thus, you will have to draw conclusions about whether there are significant exploration differences (in the fixation number, fixation duration and pupil size) between the two AOI.

**Example:** You are given the data from 10 subjects whose fixation number exploring two AOI (x & y) has been registered:

<table>
<thead>
<tr>
<th>Subject</th>
<th>Number of fixations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Area X</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>10</td>
<td>17</td>
</tr>
</tbody>
</table>
In order to test that there are differences in the explorations of areas X & Y, we need to compare the number of fixations carried out by the subjects in each of the areas by means of a test T, according to the formula:

\[ t = \frac{\overline{D}}{\hat{S}_D/\sqrt{n}} \]

donde

\[ \overline{D} = \bar{y}_1 - \bar{y}_2 = \sum \frac{D_j}{n} \]

\[ \hat{S}_D = \sqrt{\sum (D_j - \overline{D})^2 / n-1} \]

Let us see how the calculi would be done:

<table>
<thead>
<tr>
<th>Subject</th>
<th>Area X</th>
<th>Area Y</th>
<th>Dj (area X-Y)</th>
<th>Dj- \overline{D}</th>
<th>(Dj- \overline{D})^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17</td>
<td>24</td>
<td>-7</td>
<td>2.5</td>
<td>6.25</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>25</td>
<td>-10</td>
<td>-0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>3</td>
<td>19</td>
<td>22</td>
<td>-3</td>
<td>6.5</td>
<td>42.25</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>30</td>
<td>-15</td>
<td>-5.5</td>
<td>30.25</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>27</td>
<td>-12</td>
<td>-2.5</td>
<td>6.25</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>28</td>
<td>-13</td>
<td>-3.5</td>
<td>12.25</td>
</tr>
<tr>
<td>7</td>
<td>18</td>
<td>22</td>
<td>-4</td>
<td>5.5</td>
<td>30.25</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>29</td>
<td>-13</td>
<td>-3.5</td>
<td>12.25</td>
</tr>
<tr>
<td>9</td>
<td>19</td>
<td>30</td>
<td>-11</td>
<td>-1.5</td>
<td>2.25</td>
</tr>
<tr>
<td>10</td>
<td>17</td>
<td>24</td>
<td>-7</td>
<td>2.5</td>
<td>6.25</td>
</tr>
<tr>
<td>\Sigma</td>
<td>166</td>
<td>261</td>
<td>-95</td>
<td>0</td>
<td>148.5</td>
</tr>
<tr>
<td>Medias</td>
<td>16.6</td>
<td>26.1</td>
<td>-9.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Once we have filled the table in, we can calculate \( t \):

\[ \overline{D} = \bar{y}_1 - \bar{y}_2 = \sum \frac{D_j}{n} = -95/10 = -9.5 \]

\[ \hat{S}_D = \sqrt{\sum (D_j - \overline{D})^2 / n-1} = \sqrt{148.5/9} = 4.062 \]

\[ t = \frac{\overline{D}}{\hat{S}_D/\sqrt{n}} = \frac{-9.5}{4.062/\sqrt{10}} = \frac{-9.5}{1.284} = -7.396 \]

We reject the hypothesis if \( t \leq -3.25 \) ó si \( t \geq 3.25 \). As the obtained value is -7.396 is not within the acceptable area of the hypothesis 0, we can conclude that there are differences in the exploration of the two AOI.

Now it’s your go!

You should do the same with the 35 subjects of Table 15. In order to make things easier, we have solved the \( t \) test both for the pupil size (Table 16) and for the fixation duration (Table 17) and have represented the average results in a histogram. Hence, you just have to calculate the number of fixations, completing Tables 18 & 19 with the corresponding data, as well as the histogram of the fixation number.
References


BROOK, A. & STAINTON, R. J. (2000): Knowledge and Mind, MIT, Bradford


