Chapter I

Functionalism vs. psychosubstantialism

Can a machine be conscious?

1. Definition of consciousness

What is consciousness? In what point of biological evolution did it arise? How is it possible to have subjectivity in a physical world? Before proceeding, let us try to define the general concept of "consciousness", in a preliminary way, by following some dictionaries of philosophy.[1]

"Consciousness" is the more or less clear intuition that the subject has of his states and actions (LALANDE, p. 195). The possibility that each one has of paying attention to his own modes of being and to his own actions, of being aware[2] of his own states, perceptions, ideas, feelings, volitions, etc. (ABBAGNANO, p. 217). The conscious mind, as opposed to the unconscious or subconscious mind (RUNES, p. 64).

Consider, however, the classical quotation by the Scottish philosopher William Hamilton[3]:

> Consciousness cannot be defined, – we may be ourselves fully aware what consciousness is, but we cannot, without confusion, convey to others a definition of what we ourselves clearly apprehend. The reason is plain. Consciousness lies at the root of all knowledge.

Scientists and engineers often require one to clearly define the concept of "consciousness" in words, before agreeing to further discussion. But words have been developed for intersubjective communication, and the phenomenon we want to address is subjective, private to each individual. The meaning of the concept of "phenomenal consciousness" is best fixed by ostention (i.e. by pointing to the object): it is that which we are experiencing now. Thomas Nagel put in in the following way: "what is it like to be an X".[4] From this general ostensive definition, we will later seek to distinguish, in § III.1, between different kinds of consciousness.

---

[1] LALANDE, A. (1999), *Vocabulário técnico e crítico da filosofia*, 3ª ed., transl. F.S. Correia, M.E.V. Aguiar, J.E. Torres & M.G. Souza, Martins Fontes, São Paulo. ABBAGNANO, N.(2007), *Dicionário de filosofia*, 5ª ed. revista e ampliada, transl. A. Bosi & I.C. Benedetti, Martins Fontes, São Paulo. RUNES, D.D. (1942), *The dictionary of philosophy*, 4ª ed., Philosophical Library, New York.

[2] In Portuguese, "awareness" might be translated as *ciência* (but of course in a sense different from "science") or *apercebimento* (Sofia Miguens, in her thesis, *Uma teoria fisicalista do conteúdo e da consciência*, U. Porto, 2001). The Italian language has the word *consapevolezza*. On the other hand, in Portuguese one may translate the verb "to experience" by *vivenciar*. An exploration of the terms used in the philosophy of mind, in Portuguese, is presented in the *Arquivos lexicográficos* available in the homepage of this course.

[3] HAMILTON, W. (1877), *Lectures on metaphysics and logic*, 6ª ed., vol. I. Blackwood, Edinburgh, p. 191, (available online). The course on metaphysics (vol. I) was originally written in 1836-37, at the University of Edinburgh.

[4] NAGEL, T. (1974), "What is it like to be a bat?" *Philosophical Review 83*: 435-50. In Portuguese: "Como é ser um morcego?", transl. P. Abrantes & J. Orione, *Cadernos de História e Filosofia da Ciência 15*, 245-62, 2005.

2. The robotic Turing test

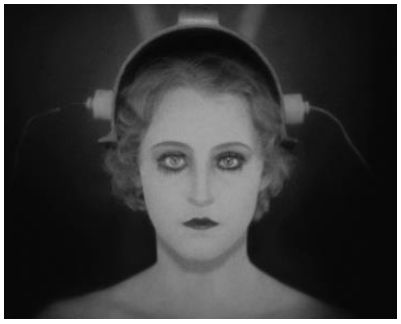

Fig. I.1. Maria, the gynoid.

Consider the situation of the film *Metropolis*, by Fritz Lang (1927), in which a metallic robot is built by a mad scientist and transformed into a gynoid (female android, or fembot), Maria, interpreted by actress Brigitte Helm, and which is indistinguishable from a human being. Let us assume that, inside, Maria is made of valves (or chips of integrated circuits), wire and metallic motors, but judging from her behavior, her facial expressions, and her speech, everyone considers her a normal human. In this sense, one may say that she passes the "robotic Turing test".

Alan Turing was a brilliant mathematician and computer scientist, who in 1950 wrote a paper discussing whether machines could "think". Instead of attempting to give a lexical definition of "thinking", he proposed a game to test whether a machine thinks. Simplifying a bit, the "imitation game" involves a human being who asks questions or establishes a dialogue with a machine or a human hidden behind a wall, and tries to make the correct identification.[5]

> I believe that in about fifty years' time [in the year 2000] it will be possible to programme computers, with a storage capacity of about 10^9 [binary digits; around 10 MB], to make them play the imitation game so well that an average interrogator will not have more than 70 per cent. chance of making the right identification after five minutes of questioning. (TURING, 1950, p. 442)

That is, if a machine can deceive a human being so that he thinks he is talking with another human, then the machine should be considered as thinking or as being intelligent. Turing is not referring directly to "consciousness", but in the discussion of possible criticisms to his approach, he discusses what he calls the "consciousness argument":

> This argument appears to be a denial of the validity of our test. According to the most extreme form of this view the only way by which one could be sure that a machine thinks is to *be* the machine and to feel oneself thinking. One could then describe these feelings to the world, but of course no one would be justified in taking any notice. Likewise according to this view the only way to know that a *man* thinks is to be that particular man. It is in fact the solipsist point of view. (TURING, 1950, p. 446)

Turing concludes that the proponent of the consciousness argument would not like to adopt a solipsist position, i.e. the position that assumes that only my own mind exists. Concluding the discussion of this argument:

> I do not wish to give the impression that I think there is no mystery about consciousness. There is, for instance, something of a paradox connected with any attempt to localise it. But I do not think these mysteries necessarily need to be solved

[5] TURING, A.M. (1950), "Computing machinery and intelligence", *Mind 59*, pp. 433-60. In Portuguese: "Computadores e inteligência", transl. M. Epstein, in EPSTEIN, I. (ed.) (1973), *Cibernética e comunicação*, Cultrix, São Paulo, pp. 45-82. The expression "robotic Turing test" may be found in HARNAD, S. & SCHERZER, P. (2008), "First, scale up to the robotic Turing test, then worry about feeling", *Journal Artificial Intelligence in Medicine 44*, pp. 83-89. On the 2029 prediction: KURZWEIL, R. (2005), *The singularity is near*, Penguin, New York, p. 295.

> before we can answer the question with which we are concerned in this paper. (TURING, 1950, p. 447)

The specific conditions placed on Turing's first quote have not yet been fully met, but each year the Loebner Prize is offered for the best machine performance. Futurologist Ray Kurzweil predicts that the Turing test will be fully surpassed by 2029.

Let us now return to the robotic Turing test involving gynoid Maria. The question to consider is: is the gynoid conscious? By hypothesis, she behaves like a human, speaks like a human, is Machiavellian like us, has the same skin as a human, and smells human, like Ava, from the movie *Ex machina* (2015). But Ava has a brain made of special matter, whereas the ginoid we are discussing is made of integrated circuits. Would Maria have consciousness?

3. Philosophical behaviorism

Behaviorism is the view that the nature of something reduces to its appearance, or to its external behavior, coupled to information about its history and genetic traits. As is well known, this view is associated to a school in psychology, which is divided into many different currents. O'DONAHUE & KITCHENER (1999) describe at least fourteen versions of behaviorism, ten psychological and four philosophical. We will not explore this topic now, but will briefly define "philosophical behaviorism" as the view that states that a mental state can only be attributed to a system based on its observable behavior. This is summarized in Wittgenstein's phrase: "An 'internal process' needs external criteria" (*Philosophical investigations*, §580). This view is also expressed in the following quote by psychologist John Staddon, referring to his position, that he called "theoretical behaviorism":[6]

> Theoretical Behaviorism takes the Turing test view of consciousness. This view is not accepted by everyone, however. John Searle (1992), if I understand him correctly, makes the argument that even if a device were to be found that could pass the Turing test, it would not be conscious. I have three reactions: First, the assumption that such a device *can* be created solely from hardware may be false, in which case we need say no more. Second is the obvious question: Assuming it can be created, *how do you know* it is not conscious? The only answer to this question is: because it does not pass the Turing test, which is contrary to the first assumption. In other words, if the only way we know that someone (or something) is conscious is because it answers our questions appropriately, then, by definition, a machine that can pass the Turing test must be conscious.
>
> My third reaction is simply to wait and see. If a machine is ever created that passes the Turing test, people will soon enough treat it as one of their own. If we are willing to grant consciousness to a dog, or to someone whose ability to communicate is as impaired as Helen Keller's, are we likely to withhold it from a device that speaks and responds indistinguishably from a human being? (STADDON, 1999, p. 230)

[6] STADDON, J.E.R. (1999), "Theoretical behaviorism", in O'DONAHUE, W. & KITCHENER, R. (orgs.), *Handbook of behaviorism*, Academic, San Diego, pp. 217-41. In Portuguese, see ABIB, J.A.D. (2015), "Psicologia, comportamentalismo e subjetividade", in CHITOLINA, C.L.; PEREIRA, J.A. & PINTO, R.H. (eds.), *Mente, cérebro e consciência: um confronto entre filosofia e ciência*, Paco, Jundiaí, pp. 279-96. The short definition of philosophical behaviorism is based on OPPY, G. & DOWE, D. (2016), "The Turing test", *Stanford Encyclopedia of Philosophy* (online). WITTGENSTEIN, L. (1953), *Philosophical investigations*, transl. G.E.M. Anscombe, Macmillan, New York; in Portuguese, in Coleção Os Pensadores, 2nd ed., transl. J.C. Bruni, Abril Cultural, São Paulo, 1979. Mentioned in the quotation: SEARLE, J. (1992), *The rediscovery of the mind*, MIT Press, Cambridge (MA); in Portuguese: *A redescoberta da mente*, transl. E.P. Ferreira, Martins Fontes, São Paulo, 1997.

4. Mentalism

The position that denies that the gynoid has consciousness will be called, in the lack of a better term, "mentalism", in opposition to philosophical behaviorism. Thus according to *mentalism*, there is a subjective or qualitative perspective that characterizes consciousness, and which cannot be externally observed in other people and animal. This thesis is encapsulated in the Hamilton quote of § I.1.

A debate between mentalists and behaviorists happened in the 1980's, concerning animal consciousness. The Gestalt psychologist and zoologist Wolfgang Köhler had gathered evidence in the Tenerife islands, during World War I, of intelligent behavior in chimpanzees, who were able to stack boxes to climb up and pick bananas stuck high in the cage. He concluded that this would have been done not by mere trial and error, but by insight (*Einsicht*), suggesting that chimpanzees have consciousness like us (except for language and other higher cognitive functions). In 1945, Herbert G. Birch showed that prior experience with the instruments was necessary for this intelligent behavior to occur. In 1981, Epstein, Lanza & Skinner followed this lead and also criticized Köhler's conclusion, showing that it is possible to condition pigeons with a repertoire of behaviors so that they can reproduce the sophisticated behavior observed in chimpanzees. For them, this suggested that it was hasty to ascribe a mind to monkeys, as their behavior could be fully explained by their repertoire of conditioning, as Birch suggested. This result is very interesting, but it does not affect the thesis that there is a mental state that serves as an intermediate cause in the causal chain linking conditioning and behavior.[7] Later on we will discuss whether nonhuman animals are conscious, and the evidence of intelligence in birds.

5. Functionalism of mental states

In the philosophy of mind, *functionalism* is the position that defends that mental states may be completely characterized by their *functions*, not by their material constitution.[8] Thus, mental states should be characterized by the causal relations existing between them, besides sensorial inputs and behavioral outputs. The substrate of mind is taken to be irrelevant, be it organic matter, inorganic matter, or event some spiritual substance; the important aspect is the organization of the system, or its informational state.[9] One may compare mental states to the logical states of a computer (its software), that is taken to exist independently of the physical nature of the computer (its hardware).

---

[7] EPSTEIN, R.; LANZA, R.P. & SKINNER, B.F. (1981), "'Self-awareness' in the pigeon", *Science 212*: 695-6. See also the short film by BAXLEY, N. (1982), *Cognition, creativity and behavior: the columban simulations*, available at https://www.youtube.com/watch?v=QKSvu3mj-14 (part 1) and ... v=erhmslcHvaw (part 2), with the comments of Skinner on Köhler. About the latter's work: KÖHLER, W. (1917), *Intelligenzprüfungen an Anthropoiden*, Königlich-Preußische Akademie der Wissenschaften, Berlin; English translation: *The mentality of apes*, transl. Ella Winter, Kegan Paul, Trench & Trubner, London, 1925. For an overview of the topic, see SHETTLEWORTH, Sara J. (2012), "Do animals have insight, and what is insight anyway?", *Canadian Journal of Experimental Psychology 66*: 217-26.

[8] The term "functionalism" in Psychology was used in the beginning of the 20th century to denote "the psychology that examines mental functions with respect to their use for the organism" (BORING, E.G., 1942, *Sensation and perception in the history of experimental psychology*, Appleton-Century, New York, p. 299).

[9] Na outdated example of a functional definition is that of a "carburetor". A carburetor is defined as anything that mixes fuel and air in an engine, using suction to introduce the fuel. In principle, a carburetor can be constructed of any material, as long as it fulfills the function that defines it.

Functionalism was articulated in the 1960's by authors such as Hilary Putnam and Jerry Fodor,[10] but it was already common in neurophysiology after World War II (see next section), and its main thesis goes back to Antiquity (about Aristotle, see Appendix 1, section A1.3).

> The intuition underlying functionalism is that what determines the psychological type to which a mental particular belongs is the causal role of the particular in the mental life of the organism. Functional individuation is differentiation with respect to causal role. A headache, for example, is identified with the type of mental state that among other things causes a disposition for taking aspirin in people who believe aspirin relieves a headache, causes a desire to rid oneself of the pain one is feeling, often causes someone who speaks English to say such things as "I have a headache" and is brought on by overwork, eyestrain and tension. This list is presumably not complete. More will be known about the nature of a headache as psychological and physiological research discovers more about its causal role. (FODOR, 1981, p. 128).

Several mental faculties are defined functionally, such as "calculating ability". An individual with "savant syndrome" has an amazing calculating ability, but a good computer has an even greater ability. In this example, the mental faculty is defined by the efficiency of performing a mathematical calculation, no matter how or what material substrate.

The strong functionalist thesis is that all mental states are definable in functional terms, including consciousness itself. The claim that mechanical machines or computers with silicon integrated circuits may have consciousness is called "machine functionalism" or "strong artificial intelligence" (we've seen, however, that the debate is complicated by the fact that different definitions of consciousness are used). One way to express this is to say that computers not only can "simulate" a human mind, as is being attempted with the *Blue Brain Project* at Lausanne, but also "emulate" consciousness, i.e. make a mind emerge from a highly complicated computation. Surely a computer can only simulate a hurricane, not emulate it, because its interior is not wet; could a computer go beyond simulating a brain, by emulating subjective consciousness? Machine functionalism argues so.

6. The homogeneity thesis

In the 1940's, functionalism was called by neurophysiologists the *thesis of homogeneity*, and was advocated by scientists such as Edgar Adrian, Wilfrid le Gros Clark and Roger Sperry. The idea goes back to the discoveries of a hundred years earlier, with Carlo Matteucci and Emil Du Bois-Reymond, that all nerves carry electricity of the same kind. In 1902, experimental psychologist Wilhelm Wundt[11] defended in detail a functionalist position, in which the simplest psychical content, such as the subjective sensation of redness, would have as its physiological substrate only a complex connection of nerve elements, not a "specific energy" (quality), as Johannes Müller had argued.

---

[10] FODOR, J.A. (1981), "The mind-body problem", *Scientific American 244(1)*: 124-32, 148 (January). See also PUTNAM, H.W. (1967), "Psychological predicates", in CAPITAN, W.H. & MERRILL, D.D. (eds.) (1967), *Art, mind and religion*, U. Pittsburgh Press, pp. 37-48 (republished with the title "The nature of mental states"). A deep discussion of functionalism may be found in BLOCK, N. (ed.) (1980), *Readings in the philosophy of psychology*, vol. 1, Harvard University Press, Cambridge, pp. 171-306.

[11] WUNDT, W.M. (1910), *Principles of physiological psychology*, vol. 1, Sonnenschein, London, pp. 320-31; translation by E.B. Titchener of the 5th German edition, *Grundzüge der physiologischen Psychologie*, 1902. See comments in p. 59 of BRIDGES, J.W. (1912), "Doctrine of specific nerve energies", *Journal of Philosophy, Psychology and Scientific Methods 9*: 57-65.

In 1912, Adrian identified the electrical spikes in the nerves, and this consolidated the rejection of the thesis that the nerves have "qualitative" differences. In his book *The physical backgroun of perception* (1947), he discussed the functionalist thesis over three pages:[12]

> The first consideration is that if all nerve impulses are alike and all messages are made up of them, then it is at least probable that all the different qualities of sensation which we experience must be evoked by a simple type of material change. [...] Impulses travelling to the brain in the fibres of the auditory nerve make us hear sounds and impulses of the same kind arranged in much the same way in the optic nerve makes us see sights. The mental result must differ because a different part of the brain receives the message and not because the message has a different form. [...]
>
> The chief conclusion, however, is that the nerve-fibers carry out their work on a simple and uniform plan, and this suggests that the activity of the brain from moment to moment should be capable of definition as a spatial arrangement and no more. It must be a pattern of excitations highly complex and rapidly fluctuating but built up of the same elements in all its parts, the elements being nerve-cell activity induced by the flow of impulses along the nerve-fibres. As far as we can tell there is no part of the brain and no stage in the elaboration of the patterns [16] where they are likely to depend on some different kind of material change; [...] (ADRIAN, 1947, pp. 14-16)

SPERRY (1952, p. 293) sumarizes well this homogeneity thesis:

> In short, current brain theory encourages us to try to correlate our subjective psychic experience with the activity of relatively homogeneous nerve-cell units conducting essentially homogeneous impulses through roughly homogeneous cerebral tissue. To match the multiple dimensions of mental experience we can only point to a limitless variation in the spatiotemporal patterning of nerve impulses. The difference between one mental state and another is accordingly believed to depend upon variance in the timing and distribution of nerve excitations, not upon differences in quality among the individual impulses.

Functionalism is clearly the dominant position among neuroscientists today, as we will see later when discussing the main theories of consciousness today. For example, in Christian KOCH's excellent popularization book (2012, p. 2):

> So, it was natural for me to wonder during my toothache whether a computer could experience pain. [...] But why not [attribute sentience to a laptop]? Is it because my laptop operates on different physical principles? Instead of positively and negatively charged sodium, potassium, calcium, and chloride ions sloshing into and out of nerves cells, electrons flow onto the gates of transistors, causing them to switch. Is that the critical difference? I don't think so, for it seems to me that, ultimately, it must be the functional relationships of the different parts of the brain to each other that matter. And those can be mimicked, at least in principle, on a computer.

7. The qualitative aspect of the mental

The hardest problem for the functionalist is to give an account of the qualitative aspect of the mind, or what Fodor calls "qualitative content":

[12] ADRIAN, E.D. (1947), *The physical background of perception*, Clarendon, Oxford. SPERRY, R.W. (1952), "Neurology and the mind-brain problem", *American Scientist 40*: 291-312. KOCH, C. (2012), *Consciousness: confessions of a romantic reductionist*, MIT Press, Cambridge (MA).

> It is not easy to say what qualitative content is; indeed, according to some theories, it is not even possible to say what it is because it can be known not by description but only by direct experience. I shall nonetheless attempt to describe it. Try to imagine looking at a blank wall through a red filter. Now change the filter to a green one and leave everything else exactly the way it was. Something about the character of your experience changes when the filter does, and it is this kind of thing that philosophers call qualitative content. [...]
>
> The reason qualitative content is a problem for functionalism is straightforward. Functionalism is committed to defining mental states in terms of their causes and effects. It seems, however, as if two mental states could have all the same causal relations and yet could differ in their qualitative content. (FODOR, 1981, p. 130).

This is a central issue in our course: the "qualitative" nature of subjective conscious experience. Terms like "sense data" and "qualia" will be used in the ensuing discussion. Arguments such as the "inverted spectrum", which goes back to John Locke and which we will examine in Ch. 3, are used to question the validity of mental state functionalism, as Fodor put it in the last sentence quoted above.

8. Thought experiment of replacement of brain cells by chips

A good way to explore positions in philosophy of mind is by presenting thought experiments (*Gedankenexperimenten*), that is, experiments that cannot be performed, at least not today (for technical reasons), or as a matter of principle. In this course, various thought experiments will guide us in the exploration of consciousness and the brain.

We have already examined the thought experiment of the gynoid made of integrated circuits, which defined two positions: philosophical behaviorism and mentalism. Let us now consider the thought experiment of the replacement of brain cells by silicon chips, explored by John SEARLE (1992, pp. 65-66)[13]:

> Imagine that your brain starts to deteriorate in such a way that you are slowly going blind. Imagine that the desperate doctors, anxious to alleviate your condition, try any method to restore your vision. As a last resort, they try plugging silicon chips into your visual cortex. Imagine that to your amazement and theirs, it turns out that the silicon chips restore your vision to its normal state. Now, imagine further that your brain, depressingly, continues to deteriorate and the doctors continue to implant more silicon chips. You can see where the thought experiment is going already: in the end, we imagine that your brain is entirely replaced by silicon chips; that as you shake your head, you can hear the chips rattling around inside your skull.

We will not explore Searle's subsequent discussion, but will suppose that each cell in the brain is replaced by a chip which reproduces the exact causal relations of inputs and outputs with other cells, known today in neuroscience, including plasticity and growth of dendrites. In the end, will the "person" be conscious or not?

We are not sure what would happen, since the experiment is only imagined. But we can define two opposing positions with respect to the expected result of the experiment. The view that consciousness would remain intact characterizes a *functionalism at the level of biological cells*: the organization of the cells (neglecting internal details) is sufficient for subjective consciousness to emerge. This is a mentalist position, since consciousness is not

[13] SEARLE, J. (1992), op. cit. (footnote 4), Ch. 3, § I.

defined by external behavior. In opposition to this type of functionalism, one has the position that Searle calls *biological naturalism*: something that happens in the biological processes within the cells would be essential for the emergence of consciousness (also a mentalist position). Thus, of the three alternatives presented by Staddon, the first one would be considered correct.

9. The functionalist / psychosubstantialist spectrum

Table I.1 presents various scales in which one finds an opposition between, on the one hand, behaviorism and functionalism, and on the other, mentalism and biological naturalism. We represent five positions (rows in the Table), but other intermediate positions are possible. What characterizes the functionalist side of the spectrum is to consider that the consciousness possessed by a system is defined only by the *relations* between parts of the system (i.e. by the organization of the system, the "arrangement" of the atomists, the Aristotelian "form"), or simply by the behavior of the system, no matter what the material constitution of the system (or the spiritual constitution) is.

The most radical metaphysical position has been called "ontological structuralism" (or ontological structural realism), and considers that everything in our universe are relations or structures, and that deep down even matter (and other physical magnitudes) arises from relations without *relata*, that is, from relations of relations of relations, etc., without the presence of basic elements.[14]

THOUGHT EXPERIMEMT	QUESTION	YES	NO
Gynoid with integrated circuits	Is a gynoid conscious?	Philosophical behaviorism	Mentalism (generic sense)
Program that simulates mental states or brain cells (*Blue Brain Project*)	Can a mind or brain simulator be conscious?	Machine functionalism (or of mental states) (strong A.I.)	Anti-functionalist mentalism
Replacement of cerebral cells by chips	Does the replacement of each cell for chips with the same function preserve consciousness?	Biological cell functionalism	Cellular biological naturalism
Replacement of parts of the cells (synthetic biology)	Does the replacement of the cell's organelles, etc. preserve consciousness?	Subcellular funcionalism	Subcellular biological naturalism
– – – – –	Does reality consist only of relations, without *relata*?	Radical ontological structuralism	Minimal substantialism

Table I.1: The functionalist spectrum is represented in the third column, and its complement in the fourth column (the psychosubstantialist spectrum).

The negation of this view will be denoted by the neologism *psychosubstantialism*, i.e. the notion that mind requires a substance, which may be matter (or some other physical

[14] LADYMAN, J. (2014), "Structural realism", in *Stanford encyclopedia of philosophy*, online.

entity), or some other category of entity, irreducible to relations. Aristotelian hylemorphism is a "substantialist" view, since it holds that in Nature there is no form without matter (contrary to Plato's view), but it would not be "psychosubstantialist", if one argues that he tended to accept that human souls could be instantiated in a non-biological substrate, as long as it fulfills its function (see section A1.3). Searle's biological naturalism would also be an expression of psychosubstantialism, so that the complementary spectrum to the functionalist spectrum may be called the "psychosubstantialist spectrum".

10. Can a machine be conscious?

We see that the question of whether "machines can be conscious" may receive several answers. If we define consciousness in the sense of philosophical behaviorism, then it seems very plausible that we will assign consciousness to robots around 2030 (according to Kurzweil's estimate). However, in our course, we will adopt the mentalist definition of consciousness, which considers that we subjectively experience a qualitative state that escapes (in a certain sense) words and which cannot (at least to this day) be observed directly from the external, objective, point of view by other people.

From the mentalist perspective, then, the question arises of how consciousness arises, from the body alone or from two separate substances (the debate between materialism and spiritualism, to be explored in the next chapter)? Is consciousness the result of the organization between the parts, being irrelevant the physical (or spiritual) nature that composes these parts, or is the nature of the parts essential to consciousness? The position that the nature of matter is essential (in opposition to functionalism) has been called "psychosubstantialism".

We have seen that this debate takes place in several scales (Table I.1), and in each of them there is a different answer to the question of whether machines can be conscious. For functionalism of mental states, it would be enough for the machine to reproduce the causal relations between mental states (the so-called "machine functionalism") for it to be conscious. For functionalism of biological cells, it would be necessary to reproduce the causal relationships between brain cells for consciousness to emerge in a robot.

The opposing psychosubstantialist position (called by Searle "biological naturalism") assumes that there is something within the biological cell that is essential to consciousness. But if we found that one part of the cell is essential, then the other parts could be replaced by synthetic parts, as explored today in the field called "synthetic biology". In this case, for this view, an artificial robot could be conscious if it carried in its cells this essential material part. This would be a psychosubstantialist solution.

Functionalism, however, is never completely defeated, a situation that can be called the "matryoshka paradox".[15] At this point, the functionalist could go down one level and argue that that part of the cell, essential for consciousness, is also composed of parts, for example macromolecules (proteins, etc.), and that it would be possible to replace these macromolecules with artificial elements, maintaining the relations between the parts. In this scale, subcellular functionalism would argue that consciousness is the result only of the organization of parts of the system, and the nature of the material substrate is irrelevant. In this case, for this view, it would be possible for a completely artificial machine to have consciousness. But the psychosubstantialist would deny that this is possible, waging a debate on metaphysical grounds.

[15] The analogy with matryohkas, Russian dolls that fit inside each other, originated with the American sociologist Talcott Parsons, who developed functionalism in Sociology.

11. Note concerning two meanings of predicates: V x M

To conclude this chapter, it is worth remembering that a predicate commonly used as a mental attribute, such as "intelligence", can be defined in two distinct ways.

A "verifiable" definition V characterizes intelligence as the ability of a system to efficiently accomplish a complex task in a verifiable intersubjective manner. Such INTELLIGENCE-V can be applied equally to a person or a machine, without one having to worry about the nature of the system.

On the other hand, the term "intelligence" can be understood as "intelligence in the human sense, involving consciousness". In this case, we are defining a mental attribute, INTELLIGENCE-M, and according to the psychosubstantialist view, a silicon machine would not have this predicate, whereas the machine functionalist would say it would.

With respect to philosophical behaviorism, discussed in section I.3, we can say that no distinction is made between the meanings V and M, that is, only type V predicates are taken into account. As for the Turing test, everyone will agree that it provides a good criterion for characterizing INTELLIGENCE-V. At the beginning of his paper, Turing seems to limit himself to this, but then in the section where he discusses the "argument from consciousness", he slides into a philosophical discussion of INTELLIGENCE-M!

Another situation in which this distinction is useful is exemplified by the question: "Does a large tree that falls into an uninhabited forest make a sound as it falls?"[16] If there are no animals in the woods, there is no subjective sensation of sound in anyone's mind, so there is no SOUND-M; but, on the other hand, vibrations are produced in the air, which in physics is considered a "sound". In this case, we can say that yes, there is SOUND-V.

[16] This question is attributed to George Berkeley (1710), but he put it another way. Using our terminology, he argued that if one wants to imagine that "there is sound-V without sound-M", one will have to recognize that this scenario is only in his or her mind, i.e. there is only "there is sound-V without sound-M"-M. An ontological realist will concede that this applies to the scenario imagined in my mind, but will argue that there are unknown instances of "there is sound-V without sound-M"-V. See §23 of BERKELEY, G. (1710), *Treatise on the principles of human knowledge*.