HUMAN VERSUS ARTIFICIAL INTELLIGENCE

Why the present-day systems of Artificial Intelligence are unable to recognize their environment

By John B. Kioustelidis,
jqstel@math.ntua.gr, jqstel@gmail.com

Abstract

Newspapers and Journals are presenting ever more achievements of the, so called, "intelligent systems". However, the robot-servant will not appear soon and the highly intelligent, the creative robot exists only science fiction.

Here we will investigate briefly the basic mechanisms on which human thinking relies and we will explain, thus, what today's "intelligent artificial systems" can and cannot do and why it is so difficult for them to simulate human thinking and in particular the human ability to improvise using objects in an unconventional way in order to achieve some goal. In other words, we will see why people can deal with the unexpected while artificial intelligence cannot deal with it.

This book comes as a further development of a previous one [1] and many specific subjects discussed briefly here are presented more extensively there.

KEYWORDS

Intuition, Improvisation, Sensorimotor mechanisms, Associative Learning, Artificial Intelligence, Human Intelligence, Language, Logic, Mental Models, Robots, Qualia, Preconcepts, Pseudo-concepts, Creative Haziness.

Introduction

What distinguishes the human mind from artificial systems of information processing? It is not logic, because it exists already in all electronic computers. What distinguishes us is rather our ability to improvise, to reinterpret things in order to achieve some goal. Indeed, it is what already very young children often do; they interpret things in unusual and unexpected ways. A stick can become eventually a provisional horse, sword, or rifle. But where does this ability come from?

The main factor which gives rise to this ability seems to be merely growth; the gradual formation of the body and the mind.

When we are born, we continue to gradually learn how to use our body. However, we also learn at the same time gradually how to use the things around us as it suits us each time and regardless of their established use. So, we learn countless imperceptible, implicit uses of things without being told about them and no matter what the grownups show us to do with them.

Later, we learn together with the language also the official, established, uses and names of things. However, what we learned pre-lingually is not erased from the mind but remains anonymous and usually unconscious until we find ourselves in need of its use. The same is true for the skills we acquire later non-verbally.

These constitute the unconscious knowledge which is commonly called "intuition" or "intuitive knowledge". E.g., if necessary, we use without much thought any heavy compact object instead of a hammer or we use a newspaper as a sweeper for crumbles. Daily, we make countless such uses automatically, without thinking about it or realizing it and that's what makes our thinking creative.

On the other hand, the artificial intelligence initially attempted to find rules that mimic intelligent behavior. In recent decades, however, it has limited the reduction to rules and started building neural networks mimicking the animal visual system. It has built so "taught" devices that learn to recognize objects within images. Yet, those things they can "see" are limited to a selected and prepared by US repertoire. Anything else is not recognized or misinterpreted.

However, while the artificial intelligence is finding ever new applications of somehow imitating human beings, it will need huge efforts until it learns to incorporate the intangible properties of the objects into its perception, because it lacks the direct experience gained with the continuous multifaceted use of our body. The future "intelligent systems" will be rather robotic, with artificial limbs and will learn the possibilities of things gradually, playing, like a small child. Only in this way will they think, somewhat, like us, and partially "see" the world like us.

In the following text we will discuss all these subjects and present psychological and neurological evidence which supports this view. The text is a far shorter and more precise and clear exposition of the views presented in a previous book [1]. So, many specific subjects discussed briefly here are presented more extensively there.

1. Improvisation versus Imitation: Why Artificial Intelligence-systems cannot make sense of their environment

A recent book published by the publishers of the journal New Scientist hat the title MACHINES THAT THINK [2]. It describes the attempts to create genuine Artificial Intelligence. However, this title, in spite of being catchy, is inaccurate, because no kind of present-day machines can approach human intelligence, while the possibility of creating such machines is still very remote. Here I will try to explain why this is so and to make clear what difficulties have to be overcome in order to achieve that goal.

What are the activities which the, so called, "intelligent artificial systems" find very difficult or impossible to perform without guidance?

Surprisingly, they are the most commonplace actions we do every day without realizing it, because they are almost automatically done:

We are in the country and wish to crack a nut. What do we do? Almost automatically we look for a rock to place the nut and a stone to hit and crack it. Sitting at the table after dinner we wish to remove the crumbles from its top in order to place there our notebook and write. What do we do? We may use any journal as a sweeper to push the crumbles away. Noticing that a screw from our eyeglasses is coming loose what do we do? We may open our penknife and try to use its tip as a screwdriver. We want to close a bottle but do not have a cork. What do we do? We may make a temporary stopper using a carrot.

In order to scare a dog away we can throw at it a stone, but also an apple or a dry onion or any other object our hand can lift. We can stabilize a shaking table by pushing a folded piece of paper under one of its legs. In order to reach a skylight high on a wall we may put a chair on a table and climb on it.

This is the kind of intuitive thinking we use all the time automatically and without noticing it. In contrast to computer-based systems, who usually can only follow precise instructions people are good improvisators. How many uses can we find for a broken plate and how many a robot directed by a computational system?

Why are we usually unaware of these almost automatic actions or reactions? Exactly because we perform them automatically, without intervening verbal-logical analysis of the problem. For many of them we do not even have an appropriate verbal description. When we stumble and fall we automatically bring our hands forward to avoid the impact of the body with the ground.

On what kind of logical analysis does this reaction rely? None! Such a reaction is usually called "reflex reaction", because it is performed without any intervening conscious thought. There is no time for thinking "I am falling. What shall I do?". This is a reaction which we gradually learned as infants, when we were much more awkward. It is a combination of visual perception, but also of the inward awareness of the loss of balance of the body and of the kinetic reaction of bringing forward our hands.

2. Sensorimotor Understanding

How do we learn this kind of thinking even before going to school?

By mere play, not any specific play governed by rules, but the child's interaction with his/her environment.

This is what children start doing from their first day of life. They do not only exercise the use of their body, they also learn in play the ways objects can influence each other. There are countless such interactions which we can induce between the objects of our environment and we learn them gradually from our birth on. They form the basis not only of everyday actions, but even of great inspirations and inventions. The first steam engine may have been a result of observing the shaking top of a pot with boiling water. It shows that steam can move things.

It seems, thus, that there are two different classifications of things we learn to use.

One is the logical classification in logical categories we learn together with language. It is necessary for communicating between us, because without appropriate standardization of meanings we do not know what another speaker means. The phrase "a cat is barking" makes no sense.

However, the earliest, and even more important, is the combined sensory and motor (or 'sensorimotor') classification, not in logical categories, but in families of objects and actions with implicit similarities in perceiving and handling them. It is based on direct interconnection of perceptive stimuli with movements of the body necessary for observing or catching and handling something. In this way standardized neural mechanisms for acting or reacting are formed without the mediation of logical processing. Immediate handling shows us that a thimble, a water-glass and an empty nutshell are all hollow objects. So we may substitute one for the other. This is why a young child may use a thimble or a nutshell pretending to give water to a doll. The child does this automatically and without logical elaboration. By putting things in hollow objects, he knows that they can be used as carriers, e.g. of fluids.

We call this classification sensorimotor (following the swiss child psychologist Jean Piaget [3]) because it consists of combining in our mind sensory stimuli with motor action we may perform in order to use things. An important part of all thought processes is sensorimotor, because all the images that participate in our mental models of a situation have a sensorimotor background, i.e., 'definition'.

Lexical definitions are based on logical categories, but perceptual understanding is based on composite sensory-motor procedures, which Piaget calls "sensorimotor schemata".

The three-dimensional perception of objects is based on our ability to move and see an object from many sides. As the great mathematician Henri Poincare says, [4, chapter III 5.] "For an absolutely immovable being, there would be no space nor geometry". He would see the world as a flat screen and the changes in the appearance of an object would be interpreted, not as movements, but as changes of the object itself, similar to the variation of oil spots on water.

Lexical definitions are totally different from the sensorimotor encoding of concepts. E.g. the concept "to throw" refers to a movement which is directed by a whole neural network and cannot be fully described by means of words. The same is true for walking or catching.

An indication of the fact that naming is not the same as recognizing, i.e. that the words are only tags, is that we often forget the name of someone or something without forgetting their features or properties.[1] Besides, if we see someone make a move, we usually understand immediately and without words what he is doing. For example, if he puts something small in his mouth, we assume that he will eat it. But if he just tells us 'Τροφή' we do not know what he wants unless he know Greek. Sensorimotor encoding is panhuman and automatically understandable, while verbal encoding is neither pan-human nor automatically understood. Note that we do not need to know how a car or a computer works in order to be able to use them. Our knowledge of them is utilitarian -sensorimotor and not analytic.

[1] This kind of knowledge is 'written' in our body, our nervous system, and not in declarative form only in our brain. The classifications it makes are due to perceptual and functional similarities and not to logical categorization. It remains mostly unconscious, because we have no names for all these features.

In a sensorimotor schema, what is "written" in our mind and our body is not WHAT we do, but HOW we do it. What is more, this happens without intervention of language, which in any case is not developed in the first year of life when the first important sensorimotor mechanisms are formed.

A sensorimotor schema is an automatic mechanism of acting or reacting to something we perceive, directed by some motive. The rest of the body participates in it both through the sensory neurons as well as the neuro-muscular system and through reflex mechanisms performed automatically, without direct intervention of the brain. Every sensorimotor process begins at sensory neurons in the body and ends up in activating motor neurons again in the body. Thus, we somehow "think" with our whole body and not only with the brain.

What is intervening between sensory stimulation and kinetic reaction?

This depends on the kind of the sensorimotor mechanism.

Reflexive reactions, such as maintaining our balance, result from sensory connections to motor neurons in the spinal cord within the spine.

Automatic movements we have learned, like walking or riding a bicycle are controlled by one of the basal nuclei, the putamen, and the cerebellum, which coordinates movements. The basal nuclei (or basal ganglia) are lumps of cooperating neurons and they are called basal, because they are situated at the bottom, the basis, of the brain.

Similarly, another of the basal nuclei, the caudate nucleus takes care of certain aspects of automatic thinking. It cares for the automatic performance of basic habits like washing or wiping our hands when they are dirty and warns us and focuses our attention on something that does not work or function properly.

Although they communicate with the cerebral cortex, the center of general information processing, these areas tell the body how to realize the kind of movements we do automatically, without declarative thinking: e.g., putting one foot in front of the other to walk or stretching our body to reach something (see Rita Carter [28, p.87]). Their role seems to be to avoid the employment of the cortex for stereotypical data.

Only basic decisions, such as where we are going (which way to turn the wheel of the bike) are made with immediate use and the cerebral cortex (in cooperation with these nuclei). I.e., to the cortex are referred only decisions and non-repetitive manipulations, such as jumping over a ditch.[2]

Correspondingly, the interpretation of a motion we observe is done using the same mechanisms and not using the cerebral cortex.

A sensorimotor mechanism is not a set of data written somewhere in our memory. We, simply, "think", i.e., we react by activating directly connections of sensory perceptions with motor reactions of the body, and not only by using verbally expressible information. Note that all animals continually use sensorimotor thinking without any words. Otherwise they could not find their way and survive in the forest or in the sea.

[2] How do we know all these? The basal nuclei are activated mainly by the neurotransmitter dopamine. If dopamine is prevented somehow from acting, we observe difficulties in the control of movement, as for instance in Parkinson's disease.

We will never learn how to use a bicycle simply by memorizing WHAT to do, i.e., the description of the necessary movements from some book on Physiology. We must exercise both our neuronal and our muscular system in using them. Only in this way do we learn to coordinate visual data with the movements that need each time be done. The development of our neuro-muscular system makes us more dexterous. Thus, it gives us abilities which we didn't have before, like the ability to climb high or to jump over a ditch.

The classifications that are made by neuro-muscular sensorimotor knowledge are due to perceptual and functional similarities and not to some logical categorization. They remain usually unconscious, because we don't have names for all these mechanisms and, thus, we are unable to point them out verbally. Note that this kind of "unconscious" is not related to the emotional unconscious, i.e., suppressed feelings and memories as they are described by Sigmund Freud.

The brain often does not cause, but reevaluates our reactions, as in the case of reflexes, automatic reactions of the body to experiences like touching something hot or losing our balance. These have their initial centers of operation in the spine and only in hindsight receive commands from the brain.

Besides, all our movements need both muscle growth and flexibility. Continuous exercise, in fact, seems to create other sensory connections with motor neurons in the spine, in order to produce rapid and flexible reactions. This phenomenon is called "muscle synergy" and characterizes all skillful behavior.

All this, exercise in automatic reactions but also muscle growth is a kind of imperceptible physical memory.

Let's not forget, that a physical ability that is not exercised is gradually lost. If our hand or foot stays for some time motionless in a plaster we must then practice it for a long time in order to regain its abilities, because e.g. an arm that does not move paralyzes. Not only the neurons, but even the muscles atrophy.

Moreover, if the body does not move from the outset, we do not learn certain aspects of reality, such as depth, i.e., the third dimension of each perceptual impression. This is shown not only the above reasoning of Poincare, but also by the kitten carousel experiment (1960) of Richard Held and Alan Hein [5]. It involves two very young kittens which were kept in dark except at the periods of the experiment. One of them moved circularly a car in which the other sat still. However, although both had the same visual impressions, only the moving one learned to recognize the depth in the environment. The other one was still very awkward.

3. Sensorimotor Thought versus Verbal Thought

The neural action-reaction mechanisms acquire a verbal designation, i.e., word denotation, only at their latest, most specialized level. Their earliest levels are perceived only as 'feelings' or intuitions.

When we use the sensorimotor memory system we may end up to perceiving a specific object having a name, but at the same time we implicitly "see" all potentialities of this object and not only the standardized /typical ones.

Sensorimotor action-reaction schemata constitute also the first child communication system before the child learns to speak. It is a silent talk by means of movements and postures of the body and face expressions as David Lewis describes [6]. E.g., in order to be taken up to his mother's embrace the child stretches the hands towards her. As Lewis shows by means of a series of photographs of young children playing, a whole dialogue can be conducted by children by means of a combination of (defiant or compliant) movements of the body and facial expressions. Every child's behavior makes obvious his wishes and intensions as he reacts each time to the changes of posture or facial expression of some other child.

However, sensorimotor thinking is predominant even in everyday activities of adults. How does someone catch a ball? Does he instantly calculate in mind the possible trajectory? There is not enough time available for logical or mathematical processing of the data. The almost instant reaction of sensorimotor mechanisms precludes any logical or symbolic processing in the brain.

Something that we hardly realize is that we often DO NOT think in words but with mental representations, or rather, by modifying mental representations until a desirable result is seen. This is documented by the fact that what we call "inward dialogue" is carried out by congenital deaf people in exactly the same way as by those having hearing. Mrs. Michele Westfall, writes on April 2,2018 in QUORA, answering the question If a person is born deaf, which language do they think in?: "I was born Deaf and have been Deaf my whole life. I do not wear hearing aids or cochlear implants (and have no desire to wear either). I speak American sign language (ASL) and it is my primary language. I am a mother of two born-Deaf children, so our being Deaf is genetic for us.

I have a voice in my head, but it is not sound-based. I am a visual being, so in my head, I either see ASL signs, or pictures, or sometimes printed words. My inner voice does have words, concepts, and thoughts. My mind is not blank, nor is it silent.

I process information through my brain, my eyes, my nose, my tongue, and my touch, all in the same way anyone would process their information. Sound just isn't part of my thought process, and because it's not part of my thought process does not mean I don't have an inner voice. I do.

I'm a conscious, sentient being who thinks and reasons. :-)".

4. Mirror Neurons

A direct demonstration of the fact that much of our thinking is based on sensorimotor reactions rather than rational processing provides the discovery of mirror neurons by Rizzolatti, Fogassi, Galleze their coworkers in the 1990s [7].

According to experiments of this research group in the frontal and parietal lobe of monkey brains (macaques) there are groups of neurons which are activated, i.e. send signals (discharge or fire), both when the monkey makes an aimed movement, e.g. to catch a piece of food and bring it to his mouth, as well as when somebody else does it. I.e., they recognize directly what somebody else does as well as his intention *without* intervening use of lingual centers, which are anyway not developed in monkeys. For course, this does not mean that such a group of neurons performs the whole process of recognizing by itself, but only that, in some way, it is the point of convergence of thousands of neurons which participate in this process. Recognition means here the association of the new visual stimulus with earlier personal experiences.

However, as it was established, such neurons exist also in the human cortex. This was done by using functional magnetic tomography (fMRI) of the activation of various regions in the brain, both when someone performs a simple movement as well as when he observes others performing the same movement. Therefore, these neurons mirror in the mind the actions of somebody else.

The instant reaction of mirror-neurons indicates that, in understanding the actions of somebody else and their purpose, there is no intervening logical analysis such as "What is this person doing?" when he is directly observable. The reaction of the neurons and understanding are directly sensorimotor.

These researchers write in particular that they have located "a surprising class of neurons in the monkey brain that fire when an individual performs simple goal-directed motor actions, such as grasping a piece of fruit. The surprising part was that these same neurons also fire when the individual sees someone else perform the same act. Because this newly discovered subset of cells seemed to directly reflect acts performed by another in the observer's brain, we named them mirror neurons". They also say: "Much as circuits of neurons are believed to store specific memories within the brain, sets of mirror neurons appear to encode templates for specific actions. This property may allow an individual not only to perform basic motor procedures without thinking about them but also to comprehend those acts when they are observed, without any need for explicit reasoning about them". One grasps some other's action because even as it is happening before his eyes, it is also happening, in effect, inside his head. They also note: "We realized that the pattern of neuron activity associated with the observed action was a true representation in the brain of the act itself, regardless of who was performing it". Thus, the activation of mirror neurons causes in someone an *internal representation* of somebody else's actions. The responses of mirror neurons reflected also comprehension by the observer of a movement's final goal. If somebody grasped a piece of food and raised his hand to put it in the mouth the activation was more intense than when he merely moved to another container. Thus, the authors concluded that action understanding is a primary purpose of the mirror mechanism.

As we have already noted, use of functional magnetic resonance imaging (fMRI) have also shown that corresponding neurons exist also in the human cortex and they function just as those in the monkey's brain, i.e., responding directly without any verbal mental processing.

However, sensorimotor thinking is far more extensive than this mechanism indicates. For instance, how do we walk? Do we mentally calculate each step (feet movements etc.)? Do we apply some set of verbal-logical instructions? But the young child learns to walk before learning how to speak.

Such a sensorimotor mechanism may seem to some people inacceptable because it is apparently non-transparent. However, we should not forget that the artificial neural networks we use with great success, e.g., for pattern recognition are non-transparent (undurchsichtig). As Pedro Domingos professor of Informatics in Washington University in Seattle, U.S.A. said in a recent interview to the magazine DER SPIEGEL [8]: The best learning algorithms are imitations of neuronal networks, which are inspired by neuronal systems which occur in man or animals. These algorithms are accurate to a very high degree, because they understand better than we do a section of the world based on immense data sets. But they are totally non-transparent. Even we the experts do not understand exactly how they operate. We only know that they function. Therefore, we should not pose rules which allow only fully explainable algorithms. It is difficult to capture the complexity of reality and keep things simple.

Similarly, Nir Ben Zrihem of the Techological Institute of Israel in Haifa says about today's neuronal systems of "deep learning": The complexity of their connections means that it may be impossible to trace the steps followed by a deep leaning algorithm in order to reach a particular result [2, p.47].

On the other hand, natural neural networks have a complex, but partially discernible form because they are gradually built up in successive developmental stages and their preliminary stages are common to many concepts, as it is revealed by the use of the same word for all by young children when they acquire their first words. For instance, they initially say "doggie" in order to refer to a dog, but soon they refer by it also to a cat or some other four-legged animal. Only gradually they restrict their usage to the established meaning of the word as they learn the names of other animals. (See e.g. [Stern, 9, 170-181] or [Moskowitz, [10]).

5. The Cyclicity of Definitions

Perhaps because they identify thought with logical analysis many people think that everything or almost everything has a verbal description in our mind, i.e., that it is stored so that we could describe it with words. But this is not so. We should not forget the cyclicity of verbal definitions, i.e., the fact that often when we explain concept A we use concept B, but then, explaining concept B we are obliged to use concept A. There are no lingual definitions without recourse to some concepts. For example, consider the following definitions given by the CAMBRIDGE ENGLISH DICTIONARY:

Branch: One of the parts of a tree that grows out from the main trunk and has leaves, flowers, or fruit on it.

Tree: a tall plant that has a wooden trunk and branches that grow from its upper part.

Leaf: one of the flat, usually green parts of a plant that are joined at one end to the stem or branch

Here we have the circular definitions:

Branch ⇌ Tree, Leaf⇌ Branch

Another example of indirect self-reference from the same lexicon is the following:

Parent: a mother or father of a person or an animal, a person who gives birth to or raises a child.

Father: A male parent. **Mother**: A female parent

Thus we have the circular definitions: Father ⇌Parent, Mother ⇌Parent

Such definitions are not very illuminating because, in the end, they are self-referential. They explain a concept in terms of itself. Such definitions are therefore usually supported in some lexica by a picture, while there are also lexica without definitions but only with depictions-images like the VISUAL DICTIONARY and the OXFORD-DUDEN PICTORIAL ENGLISH DICTIONARY, Oxford U.P. (They show e.g. a boat or a machine pointing to each part of it and giving its name). This is an indication that ultimately reference to sensorimotor perception is needed in order to clarify what the defined object is.

6. Why are the words created and how far do they determine sensorimotor mechanisms.

Word creation is actually always due to the need for communication. Words, i.e., concept-names are attached only to such objects or actions important for communication, i.e. to concepts to which we might refer directly when speaking. There are whole classes of perceptions and motions which have no verbal expression. Most sensorimotor mechanisms or "sensorimotor schemas" in our mind have no names. They are not expressible in words. Like music, sensorimotor mechanisms are NOT translatable into language. We may give them a name, but we cannot describe them by means of words. For instance, we recognize familiar faces or landscapes, but most of us would not be able to describe them precisely to an artist so that he could draw them. Most of us cannot even describe precisely the entrance of our home. We may ride a bicycle but we cannot tell precisely what we do when we use it. We can only show it by direct performance. How can we describe sea-waves or flames? How do we judge how soft or how hard something is without measuring instruments? How do we fit the pieces of a puzzle into a whole picture?

Many judgements are made not logically, but sensorimotorically. Only so can we tell e.g. (a) if a stopper fits into the opening of a bottle, (b) if a broken plate or glass can be used to scratch, cut, or carry something, (c) how the pieces of a broken vase fit together in order to be glued.

There is a general impression that this is done by a computational system very successfully. However, as Margaret Boden [13, p.34] says, the recognition of shapes (pattern recognition) is successful only if they are seen from a predetermined angle: "But the programs (many are neural networks) usually have to know exactly what they're looking for: for example, a face which is *not* upside down, *not* in profile, *not* partly hidden behind something else, and (for 98 per cent success) lit in a particular way". Can a computer system interpret Picasso's paintings?

In many of the above activities we process in our mind visually and kinetically perceptual images and do not make any kind of logical analysis of sentences or symbolic expressions.

7. Why a robot finds it difficult to imitate us

This tacit and automatic knowledge of similarities between objects and the ensuing aptitude in using them is what is lacking in a robot. We teach it to adapt its performance towards achieving a better execution of a given task but do not let it learn all implicit properties of things by playing with them. What is more, an aimless play will never teach it anything, because it is not guided by evaluation of these implicit properties.

We learn to discern properties because they have a certain survival value for us.

We tend to ignore that all learning is directed towards surviving because initially our parents take care of us and protect us very well, especially in today's urban society. But children living in a less protected environment, e.g., in the country, must become very soon aware about potential dangers. Otherwise they cannot survive. A nice-looking fruit on a plant is not necessarily edible. It may be poisonous. A wasp or a dog may cause painful wounds.

Seduced by the omnipresent use of language after the childhood many people miss also the fact that the learning of implicit properties of things is not restricted to early childhood but persists throughout our life. The use of a screwdriver is not learned in early childhood. We learn to ride a bicycle or drive a car by experience and not by reading books. We also learn all kinds of sports by exercising, perhaps by following instructions, but mostly by developing skills. Similarly, we learn how to use a keyboard without noticing how the letters are arranged on it. Our mind finds each time the appropriate key unconsciously, but we cannot tell beforehand where it lies.

8. The role of motives

A component of this classification into families of objects with similar implicit associative properties is this steadily performed evaluation, which some philosophers call 'qualia' (the perceived quality of these features for us). We do not merely note possibilities of mutual interactions of things, but also note the survival value they may have for us. People learn by experience to avoid bitter food (it is often poisonous), while they seek sweet things because they have a high nutritional value. They also learn colors, because they let one distinguish the contours of objects in the environment, but also ripe fruits. They are usually red while not yet ripe ones are usually green.

Note that the discrimination of stimuli is always due to motivation. What we choose to consider important and study more carefully depends on our motivation. We may notice the plate and not the food in it if the plate is decorated while in a costume presentation we notice the clothes more than the mannequins.

This early classification of all stimuli in functional families is what most AI researchers overlook, because it is far harder to realize technically (artificially) than mere instruction-following. As long as this kind of learning is inaccessible to robots, they will remain good simulators, but not improvisators.

One thing should be noted before going further: what I call 'functional families' is only a superficial verbal denotation. All objects (or activities) are perceived by forming neural sensorimotor hierarchies, which develop from very unspecific to ever more specialized mechanisms, and may have varying functional similarities with other objects or activities. We use a thimble as a container or vessel in order to carry something small, but at another instance, we use it to cover our finger in order to be able to press a needle when we are sewing. There is no single functional family to which a thimble belongs. Similarly, a water-glass is used for drinking water, but also as a flower pot, a capture instrument for small insects and even as an instrument to press a tack on a board.

We perceive things interpreting them according to our needs and not always according to their official use and naming. We may perceive an object with a "creative haziness or ambiguousness" interpreting it in various ways depending on the use needed each time. A newspaper can be read, used as a sweeper, or to wipe spilled water, folded into a hat, turned into a funnel for carrying cherries or fish, or torn and used as a stopper, a plug). Folded it can also be put under the foot of a table in order to prevent it from shaking. A rag-doll is perceived already by very young children as a toy-baby. Cardboard boxes are seen as miniature houses or rooms, while a stick seems to be a substitute -sword or -rifle or even -horse.

One can certainly object: "We know that the newspaper can be folded. Therefore, logic tells us that we can turn it into a hat. But who tells us that the newspaper can be forded? Only sensorimotor mechanisms for interacting with newspapers. Even the fact that folding a newspaper can form some kind of hat may result, not from logical analysis, but from immediate evaluation of trials like the fitting of a puzzle.

Some people believe that the haziness of the nature of objects can be overcome mathematically by means of the, so called, Fuzzy Set Theory. Fuzzy sets attach to all objects a probability (instead of certainty) of belonging to some given set of objects (logical categories) A, B, C, … Correspondingly, the logical conclusions are not certain but have some probability of being true. However, this theory is not adequate, e.g., as a means of recognizing forms or tacit properties, because it assumes that the categories to which an object may belong are predetermined. It also presupposes that the probabilities that the object belongs to category A or B are also somehow given.

This theory cannot reveal new unexpected classification categories as well as the probabilities that an object belongs to them. However, people are not restricted by their imagination to predetermined categories, but invent steadily new ones. The pot of boiling water may become a steam engine. How do we decide if some cloth is an undershirt or dusting cloth or a mop? What determines its use is not its shape but NECESSITY. The same is true for a newspaper, which has dozens of uses besides informing.

9. Learning to improvise

How do we classify objects and actions according to their implicit properties?

This classification is acquired together with the aptitude in handling them. Round things may roll away, but things with flat surfaces may be used to build various structures. This kind of classification may be noticed by observing the gradual progress of the child in handling things. It is also revealed by the inappropriate use of the first learned words for objects which have no perceptible similarities for adults. A child learns that "moon" is a round object that shines in the sky at night, but soon he/she may use this word referring to father's cufflinks, which are shiny, to round marks on a window and round shapes in a book. He learns the word "doggie" when referring to his pet dog, but soon will call "doggie" cats, sheep, and all other kinds of four legged animals [9], [10]. Does the child think that all these things are identical? This is very improbable. He merely uses the only words available to him in order to designate an object which catches his attention on the basis of superficial perceptual similarities with the original to which adults refer by this word.

This kind of indiscriminately used words are often called "preconcepts" or "pseudoconcepts" because they do not yet have satisfactory referential value. This is no peculiarity of early childhood. Actually, even adults may use a word referring to perceptual or functional similarities in order to name something for which they lack the appropriate word. "Boat" is often called a rowing boat, but also a yacht, a steam boat, and even a tanker. People also speak of "herbal *tea* " instead of the more precise "herbal decoction".

10. Improvisation versus Common Sense

Of course, there have been great efforts to capture basic facts about nature into a data base providing means of understanding common speech which usually omits certain facts considered to be common sense. Thus, it goes without saying that a cup containing something is placed with its opening up and that trees usually grow in open air and not in a house. Such facts are due to accumulated experience and deemed unworthy of mentioning when we speak.

The researcher of Artificial Intelligence Douglas Lenat has developed a data base with millions of such facts that belong to what we call Common Sense, which are not included in any usual lexicon. His steadily expanded system is called Cyc [7, Ch.7].

However, here we do not refer to such trivial facts but to the ability of man (as well as of other animals) to recognize /conceive unexpected relations which are beyond the data of common sense. In order to pour a dry fluid (powder or seeds) into a bottle with narrow opening people automatically form a makeshift funnel using paper, while they may use a sheet of plastic if the fluid is liquid. In order to sip some liquid, one may also form a makeshift straw using e.g. a big leaf.

We can perceive an object with a 'creative haziness' interpreting it (interpreting its role, its use) in various ways depending on the use needed each time. Things are interpreted according to our present needs and not according to their official use and official naming.

Sensorimotor thinking could be called 'animal thinking', because it is really the only kind of thinking which is available to animals since they lack language which is the basic tool of logical thinking. For many kinds of animals, for instance, monkeys and crows it has indeed been shown that they can solve complex problems relating to their environment.

An interesting example of improvisation by animals was the discovery of an easy way of separating edible seeds lying on the ground from the earth mixed with them by dropping the mixture in water so that the seeds float on the water and the earth goes to the bottom. This discovery was made some sixty years ago by a female Japanese macaque ape called 'Imo'. Imo was observed by ethologists using this technique for the first time. She later taught it to all other apes of her tribe, so that they all use it now. See [15, p.494,505]

Some people naively believe that collecting all declarative (i.e., verbally expressible) knowledge in a large data-base is enough in order to produce some kind of artificial super-intelligence. However, such knowledge can never replace our implicit, usually unconscious knowledge about the objects in our environment.

11. How do we think? (Mental Modeling)

We may not notice it, but what we call 'rules of grammar' are also rules of logic. They tell us how to make sense of the words appearing in a sentence. The main purpose of logic is to make communication possible. If a sentence violates syntax we do not understand it. The main purpose of logic is to make communication possible. Without logical standardization of concepts and of the rules of creating composite sentences we do not know what another speaker means. A word-by-word translation from one language to another is not enough in order to make sense of a sentence. For instance, in some languages there is no verb "I have" and possession or ownership is expressed in an indirect way. In Russian we say «U menja kniga» (near to me a book) meaning "I *have* a book". The same meaning is expressed in Turkish by the phrase «Ben-de bir kitap var» (on me a book is), and in Irish by «ta leahhar agam» (is the book to me). In these three languages the bodily proximity serves as a metaphorical indication of possession, while the position of the verb is not between subject and object as in English or Greek. See [Deutscher, 16, p. 130].

Actually, the rules of grammar help us create a rudimentary mental model of the situation referred to by this and the other sentences of a conversation. Such models are the basis of logical thinking as Peter Wason [17] and his pupil Phillip Johnson-Laird [18] have established following an earlier suggestion of K.J.W. Craik [19, p.51]. Craik suggested that a cerebral model of an outward situation allows verifications of truth or falsity because it "has a similar [physical] relation structure to that of the process it imitates".

Based on systematic psychological experiments Peter Wason observed that people perform much better in solving logical problems, when the problems refer to concrete and familiar situations, than when they are posed in an abstract form (see Wason's selection task in Wikipedia). This suggests that, in order to make inferences, people do not apply formal rules of a hypothetical "mental logic" but always try to build up a mental model of the given premises and "look" for valid conclusions by observing it, i.e., mentally scanning or modifying it. More descriptively we might say that thinking is to a great extent mental play acting or miming (expression by means of movements, without words). An often proposed alternative theory of reasoning is that of Lakoff & Johnson [20] that the mind conceives metaphorically conceptual categories as bounded regions or containers containing the concepts of the corresponding category. However, this theory is unable to explain why people do not solve a logical problem just as easily when it is posed in abstract terms instead of a familiar to them form based on everyday experience (see, e.g., the entry 'Wason's Selection Task' in Wikipedia). The "container logic" is the same in both cases. We will discuss this theory more extensively in chapter 3.

The expressions "to catch a cold" and "to catch a ball" show that without an underlying model of the situation we are unable to know what is meant by the verb "to catch".

According to some theorists, the constituents of mental models which Johnson-Laird calls "tokens", i.e., the mental images that represent objects in a scene, should have a more precise description [27]. However, this is not possible. Tokens can be all kinds of sensorimotor conceptualizations of elementary experiences.

Actually, Johnson-Laird does not specify the nature of the "tokens" used in mental models. However, the most plausible assumption is that Tokens are Sensorimotor representations of objects. Mostly, they cannot be properly, but only superficially, described by means of words, i.e. speech.

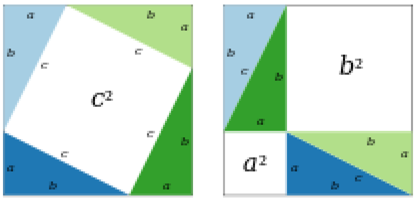
Of course, we can describe many objects and actions by means of mathematics, but this is not what happens in a human brain and does not mediate 'meaning'. How can any description of a face's contour and features make one recognize instantly a smile? Besides, most people have difficulties even with elementary mathematics, a fact which shows that mathematics do not constitute an inborne ability of the mind.

An important part of all thought processes is sensorimotor, because all the images that participate in our mental models of a situation have a sensorimotor background i.e., a sensorimotor determination (in contrast to logic which uses definitions). Like music, sensorimotor mechanisms are not translatable into language. We may give them a name, but we cannot describe them by means of words.

Note also that even when we use Logic in thinking, it is rather a sensorimotor (SeM) logic, because it is based on modifying models of reality rather than applying formal rules of Logic.

Scientific examples of this procedure of modeling and subsequent sensorimotor evaluation are the following:

a) the possibly original, i.e. earliest proof of the Pythagorean theorem (see https://en.wikipedia.org/wiki/Pythagorean_theorem:



$c^2 = a^2 + b^2$

It is made by modifying appropriately a geometrical configuration. Four identical orthogonal triangles lying along the sides of a larger square are simply shifted about like solid objects to form two orthogonal parallelograms. Comparing the original configuration with this new configuration we see then directly that the free space in the first configuration is the square of the hypotenuses while in the second configuration it is the sum of the squares of the other two sides of each triangle. Therefore, the two uncovered areas must be equal.

b) the most important discovery in biology in the 20th century, that of the double helix structure of the DNA molecule by James Watson and Francis Crick. Watson and Crick were trying for some time to fit together in long chains models of the chemical structure of the four bases (Adenin, Guanin,Thymin and Cytosin) which constitute DNA, because this structure was indicated by crystallographic investigations. The final discovery was made by Watson by using cardboard cutouts representing the individual chemical components of these bases. Watson was shifting them around on his desktop to find out the best way they could fit. At some instance, he saw that the connections of Adenin to Thymin and of Guanin to Cytosin gave rise to very similar ring forms. Thus he concluded that these rings were the basic components of DNA (see e.g., James Watson's personal account [21, chapter 26] or https://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397).

c) Einstein's "thought experiments" (Gedankenexperimente) are also due to mental modeling and subsequent study and modification of the models in order to make some physical relation perceptible. In a letter to the great mathematician Jacques Hadamard he says: "The words or the language, as they are written or spoken, do not seem to play any role in my mechanism of thought. The physical entities which seem to serve as elements in thought are certain signs and more or less clear images which can be "voluntarily" reproduced and combined [26, p.142].

Note that even a concept-word like 'bag' which seemingly refers to a concrete object is not enough in order to specify what kind of object is meant unless we see e.g. an image. Alternatively, some scenario referring e.g. to a grocery store or a bureau or office, or a school, or a woman's usual equipment is necessary in order to raise in our mind an approximate understanding of what the word means. Similarly, the word 'box' alone does not tell us whether a hit or a container is meant. The expressions "catch a cold", "catch the ball", "catch the bus" show that without a model of the situation we are unable to know what is meant by the word "catch".

The same is true for expressions like "I got an idea" and "I got the bus". They do not clarify by themselves what "to get" means.

Thus, it seems that thinking has two basic components: logical construction of some model and sensorimotor evaluation of it. Logic alone is usually not enough, because we may need not the logical definitions, but the implicit properties of the objects in order to deal with the whole situation.

As we know, Algebra allows the parsimonious symbolic expression of an arithmetic problem, which allows its solution by an appropriate transformation of the algebraic expressions to equivalent ones. In a similar way, Language allows the parsimonious symbolic description (by means of words) of a mental model, which can then be modified either linguistically by logical thought or associatively by sensorimotor evaluation in order to reach conclusions. Therefore, Language can be seen as the "algebra" of thought.

The single words are like the tags of the boxes in a storeroom. Without them we do not know what is in the box until we open it. On the other hand, by means of the tags (words) we can give instructions for putting together a composite product (some mental scene). That is why we have no recollection of our first two years of life, since we have not yet available words for composing mental models of experiences (but we have also no precise sensorimotor mechanisms of perception).

The suggestion that we constantly use mental models allows us to understand also how we deal with abstract concepts which have no concrete sensorimotor meaning as 'apple' does. An abstract concept like 'truth', 'freedom' or 'democracy' is usually represented in our mind by more than one appropriate mental scenarios (extended mental models) indicating what 'freedom' or 'democracy' is. E.g. we may have "freedom of speech" but also "degrees of freedom" (Physics). Thus, abstract concepts may have great variability depending on how each person prefers to interpret them. There are as many definitions of 'truth', 'freedom' or 'democracy' as there are philosophers. That is why there are so many philosophical studies referring to them and no similar writings referring to what 'apple' or 'orange' means.

Abstract concepts may refer to some scenario, but they often have also a sensorimotor basis. E.g., an initial understanding of 'freedom' is "being bodily unrestrained" ('free like a bird'), a later understanding is "being unrestricted by prohibitions", and finally "having one's human rights respected".

Note that some words refer (or initially referred) directly to a scenario related to the meaning of an abstract word. For instance, 'important' derives from "imported" (in-ported), because in earlier times everything trans-ported by ships, like pepper from the Far East, was expensive and therefore important.

Here it must be also noted that sensorimotor thinking is essentially hazy. The models on which it relies are usually temporary and inexact. They are due to the temporary activation of certain brain centers. In contrast to a computer the brain uses precise descriptions, precise models, only if this is necessary. In all other instances the modelization is hazy and temporary. When we drive we do not perceive whole buildings, trees etc., but usually fleetingly only certain features of them enough for finding our way. There is no time for total recall into memory.

12. How does human perception differ from that of current artificial neural networks?

The current "deep learning" neural networks made for the first time possible something in which the old, based on programming, artificial intelligence failed; the recognition of images of large numbers of objects. However, this achievement is far from recognizing atypical properties of objects, such as those we have described, or even somewhat modified images of these objects. As we will describe in Chapter 4, in humans the recording of experiences in memory is done by strengthening the connections between two neurons if the previous neuron activates frequently the next one and with weakening of such connections if their cooperation ceases for a long time. In this way associations of a new experience with previous ones are formed. When we taste for the first time some fruit, we connect directly its taste with its image. A man forms, thus, associations and modifies them on the basis of new experiences. However, today's neural networks do no such thing. They also do not follow the same method of utilitarian formation of associations (see ch.2):

In them, there is usually a repertoire of signals to which they must learn to respond properly. E.g., for arbitrarily chosen images of objects belonging to a predetermined repertoire, they must respond with the name of each presented object. For every incoming in the system image, the ideal response of the system is therefore known in advance. The difference of the actual system response from the desired one is used here to change the parameters of the system so that after several training tests it learns to respond correctly to each sample from the given repertoire of objects. In some ways such systems "memorize" image names. However, they do not perceive in the incoming images any objects other than those of the repertoire with which they were trained. In order to approach somewhat the way in which people perceive the various objects they should, therefore, learn how to imitate the way a man interprets what he sees.

13. How could a robot imitate or simulate human thinking?

There are two difficulties that need to be overcome:

(a) The formation of sensorimotor mechanisms in a growing child on the basis of experience is gradual and not preprogramed.

(b) The robot must somehow acquire evaluative mechanisms allowing the evaluation of each experience so, that relevant sensorimotor mechanisms are formed /restructured appropriately. A walnut shell and a water glass are thus included sensorimotorically into the same usability family.

(a) The gradual formation of perceptual mechanisms

Why is the gradual formation of perceptual mechanisms necessary?

The gradual formation is necessary, because the initial rough perception babies have, allows an easier discovery of sensorimotor similarities. It makes similarities apparent and differences not so salient. We first develop rough sensorimotor schemas and then, with growing experience, we specialize them in order to perform with greater dexterity more specialized activities. We first learn the general use of the hand and later learn to write or draw figures. We first learn to walk and then to use stairs. We also learn to use not only our bodies, but also other objects in interaction with our bodies, e.g., the banister of the stairs we climb.

The gradual acquirement of sight and movement control [22, ch.9, ch.11] allows the detection of common features. The child first detects similarities and then differences. As Lise Eliot [22, p.210] says: "Yes, newborns can see, but not all that well. The world for Ginna (the newborn daughter of an optometrist) probably looks the way it would to you or me if we had to stare out of a frosted window all the time. Though light makes it through her eyes unimpeded, neither her retinas nor her brain can process its information in a manner sophisticated enough to detect most of the objects, shapes, and colors in the world. Her acuity is poor, meaning that she can discern two different objects only if they are separated by a large distance; everything closely spaced tends to blur together". Without an impaired eyesight we still are blind until we train our eyes to see. This is true not only for young children, but also for adults. A mechanic sees in a car's machine damages, which we may totally miss. Similarly, an architect discovers in a building details which totally escape our attention, if he does not point them to us. The same is true for a doctor. In a computer tomography scan he may see all the health problems we have, while we only see a mess of indistinguishable organs. A machine can also be trained to distinguish certain details in a CT-scan, because such images are not varying. This would be not as easy for varying scenes on a road.

Eliot notes also [22, p.263]: "Particularly in infancy, before a child understands language, motor skills are a critical tool for learning about her social and physical environment. For instance, once a baby can successfully reach out and grasp an object, she can begin exploring physical properties such as shape, weight, and texture. Crawling allows her to actively seek out new encounters with people and things. Every motor milestone broadens a baby's experience and changes her perspective on the world, creating an ever-changing environment that is essential for emotional and cognitive growth".

For a very young child a coffee cup and a water glass are almost equivalent, because he/she cannot yet catch and lift the cup by its handle. A screw driver and a knife seem to be similar hard, cold and sharp instruments, because he/she cannot yet grasp the knife and cut or turn the screwdriver to fix or unfix a screw. These actions require considerable aptitude of the fingers.

One may object "But he/she can observe adults using these objects". However, this does not automatically confer any aptitude in using them. For instance, if we are not familiar with bicycles, we are unable to understand exactly what a bicyclist does even if we observe him for a long time. Likewise, if we don't know anything about a machine, we don't understand what a mechanic who repairs it is doing.

The meaning of some object or action is initially determined not by how it is classified in logical categories, but by how it can be used in each stage of development of our nervous system. I.e., the ‘meaning’ of an object for us is initially how we can use it and not a lexical definition. Thus, the graduations of movement-control correspond to hierarchical levels of perceiving family resemblances. The sensorimotor schemata which gradually develop or evolve in our nervous system are mechanisms of interpreting perceptions which acquire an increasing specialization.

This creative haziness, or ambiguousness, or opaqueness helps us reinterpret perceptions and reach new insights. Initially our mind is like somebody walking in the fog. From afar many objects look alike. As he draws closer they become increasingly distinct. At some stage he may discover differences and give them different names. However, the earlier connections in the mind do not get lost. They are always there as precursors of the final concepts and provide clues for family resemblances between them. This allows eventually the alternative use of one instead of another. Different objects may have a partially common sensorimotor background. The common preliminary stages of sensorimotor encoding allow us to detect functional similarities between objects or actions and to use alternatively one instead of another. All small and heavy objects (a stone, an orange, a ball etc.) can serve as projectiles because we use the same movements in order to catch or throw them. All hollow not very heavy objects of the size of a head can serve as makeshift hats. I.e., we don’t only perceive objects we also perceive potentialities.

Robots will also need to acquire or develop these sensorimotor mechanisms gradually, a task which will be difficult, because it will be necessary to add ever more hardware simulating new neural connections.

Also learning to speak has certain difficulties. Training teaches us the discrimination of stimuli, but we learn which stimuli have names (word designations) by trying to communicate with others, especially our parents and not by some kind of direct encoding. Note that stimulus discrimination is difficult and requires usually many trials before one knows what adults mean when they use a word. There are more than one stimuli we perceive when we hear a word. So, we need to find to which stimulus the speaker refers.

14. Mute Thinking

Sensorimotor in its nature is also the mute thinking, which is different from the inward silent dialogue we often make. We may call ‘mute thinking’, i.e., thinking without words, our ability to think in a succession of mental images, for instance, in order to rearrange objects in a room. We often think about moving something as a sequence if actions without actually performing them. This flows in our minds like a mental video without any use of words. This is also how we understand and enjoy mimesis (gesture play), e.g. the brilliant silent performances of Marcel Marceau. Marceau has called it "The Art of Silence" and explained it this way:" The art of Silence speaks to the soul, like music, making comedy, tragedy and romance, involving you and your life... Creating character and space, making an entire show on stage- showing our life, our dreams and our expectations". Actually, this is ;also the way young children communicate silently by means of gestures and facial expressions [6].

15. The sense of sight and the perception of space

The fact that the sense of space and more generally of the things which are in it is built up gradually during the first years of our life can be seen if we consider the cases of people who lost their sight in early childhood and regained it after 40 or 50 years. Two such cases are the one described by the neuropsychologist Oliver Sachs [23, pp. 185-249] and that of Mike May described in Wikipedia in the entries 'Mike May' and 'Recovery from blindness'.

The previously blind person is unable for a long period of time to understand what he sees and to orient himself in space. As Oliver Sachs explains, our sense of the surrounding space is based for a blind person in a temporally serial groping of objects. Consequently, initially he is unable to perceive it as a simultaneous presentation of all objects that are in it. He is called to recognize each time, not one, but many objects at the same time. He also finds it difficult to interpret the shadows that cover either the ground or other objects. He tries for instance, not to step on them, because he does not know what they are. Likewise, he understands the third dimension deficiently. When he moves, the objects seem to shrink in size and not just to move away. Thus, he finds it difficult to judge distances. This condition continues for years and is often such a cognitive and psychical burden that it influences negatively the general condition of his health. Even after years he does not recognize faces and is unable to distinguish similar objects like a lemon and an apple. This happens because he does not perceive all their features simultaneously.

Here we see that the sense of space, which seems to us self-evident, needs years of audiovisual practice (combination of visual and kinetic experiences) and does not occur automatically as soon as we open our eyes. This is the great difficulty of automatic driving systems that are being developed today. Since they did not have the opportunity to examine the various objects encountered on the street, they do not understand what they are and whether they should be avoided or not.

It is also worth noting that through vision writing makes our eyes "speak" to us and allows even deafmute people to "hear", although what they "hear" (perceive) are rather mental pantomime of concepts and not internal voices. The deafmute born Michele Westfall, who has normal sight wrote: I have a voice in my head, but it is not sound-based. I am a visual being, so in my head, I either see ASL signs, or pictures, or sometimes printed words.

The necessity of evaluative mechanisms or criteria

The initial evaluative mechanisms in man are based on the hormonal system, the other, equally important, communication system in the body besides the neural one. Certain hormones will tell us if we are thirsty or hungry, and enhance the perception of taste or smell. Other hormones (mainly adrenaline and cortisol) trigger fear reactions when we perceive danger and so on.

The main differences of the two communication systems are that neurons transport signals to specific centers within milliseconds, while hormones are carried by our blood, may activate many organs at the same time and their influence is often long lasting [24, pp.127-129]. E.g., hunger is not a momentary feeling, but lasts as long as the blood sugar does not have the appropriate level.

A robot collecting experiences and organizing them in usable sensorimotor associations needs some similar evaluative system.

This system is closely related to emotions and motives. However, any robot that gradually develops its own emotional system will be unpredictable and possibly dangerous. Pretending to have feelings, as a robot might do, is not enough. Feelings (e.g. anger, fear, love, hate) are not only meant to influence others. They do not only determine our reactions, but also filter our experiences co-determining (together with motives) what will be permanently stored in long term memory.

Recent issues are initially recorded in short-term memory and only if they are considered as important they are permanently stored in long-term memory.

Note that the success-failure selection criteria in a computational algorithm or an artificial neuronal network which is taught some aptitude through examples fulfill a similar role as human motivation.

However, the human motives are far wider and versatile (see the corresponding chapter in [1]). They are not restricted to replicating a very specific reaction, but usually judge success or failure on the basis of a general goal of satiating e.g. hunger. Thus, a robot which imitates people successfully must have a multitude of general goals (evaluative criteria) that readjust its behavior just as in people there are dozens of motives doing so (fear, aggressiveness, hunger, thirst, curiosity, communicating, ambition etc.). Such a robot must also learn to which motives it should each time give precedence, because even two motives cannot usually be satisfied at the same time.

17. Conclusions

Referring to the current systems which try to simulate human thinking and perception we can say briefly that:

1. Robots able to simulate any human activity based on symbolic processing are able to operate only in absolutely controlled environments, where every detail is predetermined and every object is marked e.g. by a traceable electronic code (identifier).

2. Systems which are able for successful pattern recognition based on accumulating experiences and not based on rules are the artificial neural networks. But the structure of such networks is totally inscrutable (what is presently called 'deep learning' should rather be called 'inscrutable learning'). Their interconnection with general purpose robots would not result in a mechanism able to move in a

natural environment, because the success of such systems depends on how the stimuli are presented to them. For instance, a face in profile is not recognizable if the system has been trained by perceiving faces looking directly to the camera. Besides, a not well-prepared environment is not recognizable even for them. They may not be able to say what is a branch or a sack fallen on the street.

The functioning of such systems will always be disrupted by unexpected changes in the environment and they will be unable to improvise solutions for a suddenly occurring practical problem. Safe automatic driving is not feasible because the road barrier recognition systems will not know what the unforeseen and unexpected obstacles are. Such obstacles can be, e.g., an open umbrella carried by the wind, a fallen branch, oil or gravel spilled on the road, sacks or boxes fallen from a truck or even a forward fallen man, whose face cannot be seen. Likewise, it is uncertain for them what is a large surface with some inscription extending across the road. It may be a banner of protesters, or a wall or a wagon of a truck, which advertises the company that owns the truck. So, they won't be able to appreciate how the system should react. Moreover, a foggy atmosphere or darkness will make driving for the automatic driving cars even more difficult. As Gary Marcus, an investigator of Artificial Intelligence reports, a neuronal network trained to detect images, considered a mushroom as a bun, a STOP turned mark at 90 degrees as a dumbbell and a STOP with colored stickers affixed on it as an open refrigerator (DER SPIEGEL, 2-11-2019, p. 115-116).

In her recent book ' Artificial Unintelligence ' [25, ch. 8] Meredith Broussard explains that today's automatic driving systems use basically pre-stored details of the road and what is in it and limit the identification of present obstacles to the absolute minimum.

However, nothing prevents the development of semi-automatic (semi-autonomous) cars based on the tests already made by Tesla, Uber and others.

The driver will intervene voluntarily or after a system signal that it is unable to recognize something. But he/she may find it difficult to be vigilant, because the automatic operation of the vehicle will give him a misleading feeling that he does not need to watch as well. A driver will be, anyhow, necessary, because the operation of such systems may be disturbed by unexpected changes in the environment. And it may be unable to improvise solutions to the unexpectedly arising practical problem.

The attempt for a full replication of human thinking must be based on studying in depth how the sensorimotor conceptual system of a young child develops, something which has been done only in general terms by Developmental Psychologists. The psychologists who study the, so called, Embodied Cognition have reached important conclusions, but their research is restricted by the choice of people they study in their psychological experiments. These are in most cases university students. The study of the behavior of young children is greatly hindered by the fact that they do not have yet developed verbal abilities. But the study only of adult people seems like an attempt to find out how a seed grows by investigating only the grown tree. Besides, we still know very little about the mechanisms of the motives which direct the development of our perception about the world and of the system of concepts we use.

The predominant problem of today's artificial neural networks is that they do not have inherent incentives to adapt to their environment (as it is perceived by their sensory devices, e.g. by a digital camera) but achieve exceptional performance only with predetermined "incentives", i.e. criteria for assessing their responses.

Without a prior training to some object’s outlook by the manufacturer they DON'T know what they're observing. In order to have high percentages of successful recognition of forms what they are supposed to "see" must be predetermined (e.g. person) and presented from a previously specified angle, i.e. in a predetermined manner. The so-called today "artificial Intelligence" is still "fictional or fake intelligence".

A system that we can directly test to see the limitations of so-called "intelligent Systems", is automatic translation. Whoever is curious enough can find out for himself these limitations by trying the automatic translation system of Microsoft which cooperates with the recent editions of Office.

The basic commands by means of which a selected text is sent to the system and returned translated are: Review> Language >Translate. If you know a foreign language try translating with this a paragraph of text and judge the result. The translation is almost never satisfactory. However, I already use it extensively to translate texts and only correct every inappropriately translated phrase subsequently. This speeds up considerably the completion of a translation.

3. Another interesting conclusion of these considerations is that the wish to perpetuate our life by copying our thoughts and concepts in some kind of electronic memory cannot be fulfilled. There is no way of copying a mind into an artificial system as some science fiction writers imagine. The mind is based on innumerable neuronal connections and not on verbally expressible contents. How can such connections be copied, and beyond that, how will the hormonal system which motivates the mind be created?

18. The future artificial intelligence

Today’s Artificial Neural Networks can be very successful in their given task and indeed perform better than predesigned algorithms. Nevertheless, they serve only a predetermined and precisely demarcated purpose! They are not adjustable to the varying needs arising in a real environment.

The human mind adjusts our behavior to a multitude of challenges we face in everyday life and it is guided by a multitude of motives. Thus, a really intelligent robot will have to use a hybrid system: algorithms for digital processing of information and a multitude of neural networks serving a multitude of purposes.

An important technical problem they will face is the expandability of their digital memory and of their neural networks. This may be solved by a daily or weekly revision process imitating the daily adjustment of our mind to outward challenges.

Today’s artificial neural networks cannot be industrially replicated in large numbers, because we lack a plan of their internal structure. This structure arises by steadily improving performance by modifying the strength between neural connections (the so called ‘weights’) adjusting them to the outcomes of repeated trials. We can only produce a large number of such neural networks by parallelly training many of them using the same set of training data. In this way we will do what a school does for its pupils.

In any case, a future ‘intelligent’ robot will be a hybrid system, like the human brain. It will combine verbal- rule based processing with sensorimotor processing, i.e. with a multiplicity of neural networks equipped with “motives”, which are the criteria of evaluating all experience.

19. Fictitious (fake) Intelligence

Today's "Artificial intelligence" systems may still seem inadequate to substitute humans, but they can make up for it in many areas. A robot may not be able to arrange a messy room (due to the inability to identify the objects), but may keep company to lonely people or to the elderly by chatting with them either in writing or even verbally. However, such an impressive indication of intelligence is due to skillful use of already pre-existing texts. We already have two examples of systems that respond to what we say:

In the first case, a woman who had an artificial intelligence company developed a program that responds to online messages like a friend of hers who died prematurely. Thus, she can conduct discussions with him to some extent by means of the program, which each time selects from his extensive earlier communications with her and other acquaintances the more appropriate remarks he made to the subject discussed. See Casey Newton, Speak Memory, http://www.theverge.com/a/luka-artificial-intelligence-memorial-roman-mazurenko-bot

The article says that Eugenia Kuyda co-founded in Moscow the company of artificial intelligence Luka, but in 2015 moved it to San Francisco. However, her old friend Roman Mazurenko remained in Moscow and was killed in a car accident.

The first products of the company Luca were bots (internet bots or web robots) “machines” (computer programs) for automatic search in the internet of information in order to make restaurant reservations. Such bots are usually somewhat clumsy impersonations of interlocutors. They respond somewhat awkwardly to the most basic questions relevant to the subject. To Kuyda came now, the inspiration that the messages he had exchanged with Mazurenko could form the basis for a bot that mimics him. Ten friends and relatives of Mazurenko contributed his writings that were more than 8.000 lines of text referring to all kinds of topics. Thus, the company developed a neural network which answers messages in Russian and trained it to use expressions often used by Roman. Roman's father found that this system uses Roman expressions, but sometimes answers wrongly. However, many of his friends found the likeness impressive. The system answers, e.g. giving advice, even to subjects one had never discussed with Roman obviously, deriving texts from his communications with others.

In the second case, an appropriately programmed computer conducted a debate with a human being by choosing from the Internet the most appropriate arguments. See "Man vs AI machine in debating competition", Bloomberg, February 13, 2019, https://businesstech.co.za/news/technology/299114/man-vs-ai-machine-in-debating-competition/ or

"Meet Harish Natarajan, the man who defeated IBM's AI debater in a man-vs-machine faceoff" https://yourstory.com/socialstory/2019/02/harish-natarajan-beats-ibm-project-debater

In both cases the content of the conversation eventually came from a person, but so that this is not evident. That is why I speak of "fictitious or fake intelligence", i.e., intelligence like the apparent intelligence of a student who cheats at an examination. The computer, or the AI system does not understand the dialogue. It only combines in a skillful way information that it finds with the help of a search engine in Internet archives by means of keywords which arise during the dialogue. Both the information and the method of combination-composition are derived from people.

However, these systems are, even as "text replicators", a significant progress since, appropriately prepared, they can help various people. To some extent they can be used for psychotherapy, because, as Joseph Weizenbaum found accidentally in the sixties in MIT using a program called ELISA (see Wikipedia – ELISA) many people tend to consider a keyboard as something impersonal, to which they are tempted to entrust what they usually do not say openly, just as young people used in earlier times to write in a personal diary. A program like ELISA is even more appealing because it is an impersonal conversation partner.

Maybe we think that's not working, because a man has a much deeper understanding of social conditions or psychic reactions than a computer. However, it is already therapeutic to be able to tell someone one's problems while, if some questions exceed the 'understanding' or the capacities of such a system, it can always respond with subterfuge, i.e. with a question like "what do you mean by …" or give an impartial answer like "I never liked … (e.g., poetry, or music)".

Another form of fictitious intelligence, already in use, is consultation systems for patents. Such systems first broaden the search terms of a technical problem and then seek on the Internet similar patents or ideas that we will be able to emulate to achieve a solution. The search is done in patent Office files , but also in social media [2, pp.139-141].

There are infinite possibilities of developing such systems dedicated to some human mental need or to assist an investigation.

Is this "artificial intelligence" genuine intelligence? No, but it is sufficient for some uses.

20. REFERENCES

[1] John Kioustelidis: THE MECHANISM OF THINKING www.math.ntua.gr/~jqstel/book1.pdf a translation of
Ιωάννη Β. Κιουστελίδη: Ο ΜΗΧΑΝΙΣΜΟΣ ΤΗΣ ΝΟΗΣΗΣ, Αθήνα, ΠΑΠΑΣΩΤΗΡΙΟΥ 2002.
[2] Machines that Think, New Scientist 2017
[3] Jean Piaget: The Psychology of Intelligence, Routledge, London & New York, 2002 (first published 1947).
[4] Henri Poincare: The Value of Science https://www3.nd.edu/~powers/ame.60611/poincare.pdf
[5] R. Held – A. Hein: Movement Produced Stimulation in the Development of Visually Guided Behavior, Journal of Comparative and Physiological Psychology, 1963, vol.56, no.5, 872-876.
[6] David Lewis: The Secret Language of your Child, Souvenir Press, London, 1978.
[7] Giacomo Rizzolatti, Leonardo Fogassi, Vittorio Gallese: Mirrors in the Mind, SCIENTIFIC AMERICAN 2006.
[8] K.I. Wettlauf „Wir überlassen den Maschinen die Kontrolle, weil sie so großartig sind" DER SPIEGEL Nr.16, 14-4-2018, 106-109.
[9] Stern, Clara & William: Die Kindersprache, Johann Ambrosius Barth, Leipzig 1922 (First Edition 1907)
[10] Breyne Arlene Moskowitz: The Acquisition of Language, Scientific American, Nov.1978
[11] VISUAL DICTIONARY, Ελληνικό – Αγγλικό (εκδότης Καυκάς, 2005)
[12] OXFORD-DUDEN PICTORIAL ENGLISH DICTIONARY, Oxford U.P.
[13] Margaret Boden: Artificial Intelligence, a very short Introduction, Oxford U.P., 2018
[14] John Horgan: The Undiscovered Mind, Touchstone, N.Y. 1999
[15] S. Hirata, K. Watanabe, M. Kawai: Sweet Potato Washing – Revisited. https://www.japanmonkeycentre.org/pdf/sweet-potato-washing/Hirata-2001-Sweet-potato-washing-revisited.pdf
[16] Guy Deutscher: The Unfolding of Language, Arrow Books, 2006.
[17] Wason, Peter C.: *(1966). "Reasoning". In Foss, B. M. New horizons in psychology. 1. Harmondsworth: Penguin.*
[18] Johnson-Laird, Philip: Mental Models, Cambridge U.P., 1983.
[19] Craik, K.J.W.: The Nature of Explanation, Cambridge U.P., Cambridge, 1943.
[20] George Lakoff – Mark Johnson: Philosophy in the Flesh, BASIC BOOKS, 1999.
[21] Watson, James: Η ΔΙΠΛΗ ΕΛΙΚΑ, Τροχαλία, 1990.
[22] Lise Eliot: Early Intelligence, Penguin Books, London, New York 1999.
[23] Oliver Sachs: Ένας ανθρωπολόγος στον Άρη, εκδ. ΑΓΡΑ 2000.
[24] M.R. Rosenstein-S.M. Breedlove-N.V. Watson, Biological Psychology, 4. Ed., Sinauer Associates, 2005.
[25] Meredith Broussard: Artificial Unintelligence, How Computers Misunderstand the World, The MIT Press, 2018

[26] Jacques Hadamard: The Psychology of Invention in the Mathematical Field, Princeton U.P. 1945, DOVER 1954.
[27] Barsalou, Lawrence W.: Grounded Cognition, Annu. Rev. Psychol. 2008.59: 617-645.
[28] Rita Carter: Mapping the Mind, PHOENIX, London, 1998.