On the description of intelligent systems

Renzo Beltrame[1]

In the framework of cognitive activity, the qualification of a behavior as being intelligent is the result of a particular way of considering it; that is, it is a consequence of a particular categorization. At this point the discussion might seem to be closed, because we have an intelligent behavior when we categorize it in this way, and we do not have an intelligent behavior if we do not categorize it in this way. This kind of assertion is true for any mental activity when we consider it after its occurrence: in fact, after a mental activity is accomplished, we have the mental thing that has that activity as constitutive. However, if we limit ourselves to this approach, then we will bypass the main problem of mental dynamics: to achieve a theory which predicts the mental activity that will occur in certain conditions.

Our discussion of intelligent systems will initially focus on the minimal conditions we want to be satisfied in order to categorize a behavior or a system as being intelligent. In fact, we can show examples of systems that we usually consider as having an intelligent behavior, systems to which we refuse to attribute an intelligent behavior, and systems whose behavior we can consider in both ways: intelligent or not intelligent. Later in the paper we will discuss some features that we nelieve are peculiar to the description of an intelligent system.

Two simple cases, taken from elementary physics can clarify the kind of problems we meet in deciding which requirements must hold to consider a behavior or a system as being intelligent. In simple Newtonian mechanics, a direction and a scalar completely describe the actions on the physical system whose behavior we are studying. In this framework we think of the system as being atomic rather than being composed of parts. Furthermore, we characterize the system by means of one scalar m, the mass.

A vector a completely describes the change in the system that the action F induces at a certain instant of time. The well known relation:

$$F = ma$$

describes the behavior of such systems, the so-called material points. It links the action F, with the vector a that describes the change in the system, and the parameter m that individuates the system. The scalar m must be a constant, otherwise a different relation holds. The vector a is interpreted as an acceleration: that is, as a change in velocity; and the velocity characterizes the state of the system. Mathematically, the vectora is given by the time derivative of the velocity v; that is:

$$a = \frac{dv}{dt}$$

[1] National Research Council of Italy - CNUCE Institute - Via Santa Maria 36 - I 56126 Pisa
e-mail: renzo.beltrame@cnuce.cnr.it

From the two relations written above we have:

$$v(s) = \int_0^s \frac{F}{m} ds$$

and we see that the value of the variable, which characterizes the state of the system, is given mathematically by a functional of the history of the actions on the system until the instant of time we are considering.
We can often substitute the computation of this functional with a physical measure of the vector v. In these cases the result of the measurement is equivalent to the knowledge of the history of the system, at least in order to give predictions of its mechanical behavior. Even in such a simple case, we have a functional of the system history; and this functional is a state variable too.
Usually we do not accept to consider a behavior that is described in this way as being intelligent, at least until we assume that it rigidly depends on the environment actions.
Our second example is a physical system whose dynamics is described by the same variables, but where we will assume that the mass is a scalar function of the time.
The relation that describes the behavior of the system is now:

$$F = \frac{d}{dt}(mv)$$

and it links the action F to the change of the moment mv. We obtain by the derivative:

$$F = m\frac{dv}{dt} + v\frac{dm}{dt}$$

that is:

$$\frac{dv}{dt} = \frac{F}{m} + \frac{v}{m}\dot{m}$$

We can solve the problem when we know the history of the mass rate of change. In this case we can compute the mass as a functional of the history:

$$m(s) = \int_0^s \dot{m} dt$$

and then we can calculate v by the integral written above.
We can still interpret the force F as describing the action of the environment on the system, and the acceleration a as the response of the system to the environment action. The system response was a linear function of the environment action, when the mass m was constant. Now the system response depends on the force F , on two functionals of the history, m and v, and on the rate of change of the mass rate of change; and we have two ways to provoke changes in the system behavior: the action of a force, and the change of the mass. These two modes can be considered as being independent. We can continue to ascribe the action of a force to the actions of the external environment; but we can think of mass variation in different ways.
If we ascribe the change of mass to an environment action, then once again we have a sys-

tem whose behavior rigidly depends on the environment actions. In fact, we can describe the mass change as a function of a further external action Q.

On the other hand, if the changes of mass depend on certain characteristics of the system, then we have a system that can change its behavior both by external, and by internal actions. However this characteristic is only a prerequisite to the system being considered as having an intelligent behavior. Usually we require at least an adaptive behavior; and this in itself may not be enough, since now there are many machines whose performance is controlled by suitable parts of the machine itself.

We shall consider our system with variable mass as being intelligent, when we think of its mass changes as intentional, or even voluntary; that is, when we think that the system can carry out mental activities, and thus it can mentally anticipate its behavior. In these cases the adaptive behavior becomes a consequence, and we expect different instances of such a kind of behavior. We should find many ways of varying a feature, when, as in our case, the system can vary only one of its features. If we only find one way, then we might refuse to consider the system as being intelligent, or we could accept the categorization only with an explanation of this atypical behavior.

However, we are not forced to use this anthropomorphic definition of intelligence. Less constraining definitions might arise and be accepted, like in the case of flight, whose definition was extended to include the flight of planes and helicopters. Therefore the choice of requirements to categorize a system as being intelligent has a certain degree of freedom, and we can find significant differences when we compare different historical moments. We will come back to this degree of freedom; here we will outline a more constraining problem, which the anthropomorphic definition of intelligence clearly shows.

In elementary mechanics we use a mental scheme where the cause of the movement of a body is external to the body itself; that is a physical body cannot be assumed to cause its movement. This assumption must be considered part of the definition of a mechanical body. We find it explicitly stated in Euler's Mechanica[1] and we continue to use it.

In physics the action of a body on another physical body is thought to imply a physical process, that is a change, in the agent. The cause of this process must be external to the agent to agree with the assumption stated above. So the actions among parts of a physical system are conceived as interactions, and in the particular case of interactions between two parts, we must think of the mutual actions as being equal and opposite.

In psychology on the other hand we use a mental scheme where animals and humans may be cause of the activities that they perform, and of their behavior.

When we view animals and humans as biological systems, we may describe how they function with the schemes of physics, and we automatically introduce the assumptions that are implicit in these schemes. We cannot thus maintain in this kind of description the principles of psychology, because we would have to introduce a non physical cause, the acting subject, to explain physical processes, and the contradiction leads to a spiritualistic position.

Yet we cannot base a psychological description of animal and human behavior on the schemes of physics. In these schemes every change has its cause in something that is different from the thing that is changing. Then we lose the subject autonomy, which is a possibility of the psychological description.

We can break down the opposition between the two approaches, and even remove it, by

[1] In Newton's formulation: “Corpus omne perseverare in statu suo quiescendi vel movendi uniformiter in directu, nisi quatenus a viribus impressis cogitur statum illum mutare”, it is not sure whether the cause of a mechanical change must be external to the physical body. No doubt it is possible with regard to Euler’s formulation: “Corpus absolute quiesciens perpetuo in quiete perseverare debet, nisi a causa externa ad motum sollicitetur” [L. Euler, Mechanica sive motus scientia analytice exposita, 1736, Ed. P. Stäckel. Leipzig, 1922, Vol. I, p. 27].

changing one of the points of view. In psychology, for instance, we could continue to see a person as an acting subject, but we might consider her / his actions and behavior as being completely dependent on the physical facts that occurred to the person's body. We should however reconsider large parts of our culture, which are based on the freedom of the acting subject, and, like ethics and criminal law, derive a statement of personal responsibility from this assumption[1].

We can maintain the mental schemes of physics and psychology, and build a theory that encompasses both approaches; but we can use a correspondence between the elements of the two different theories, psychological and physical, instead of grounding one theory on the other. The correspondence requires:

- the construction of two different theories that explain the same experimental facts following the two different approaches;
- an individuation of the experimental facts that are compatible with the implicit assumptions of both approaches;
- a correspondence between the intermediate explanatory elements of the two theories.

The unifying work would often require a deep rearrangement of the two initial theories to obtain a satisfactory correspondence.

When we study as a physical system a system to which we ascribe an intelligent behavior, then the experimental facts must be physical facts, from which we can start to build the theory; and the facts that the theory predicts must have the same character. This choice becomes mandatory for psychological theories too, when we pursue unification.

Nevertheless this requirement is not very constraining. Let the language be employed for conditioning the subject of the experiment, and let the subject’s answer, too, be linguistic. We only require that the interpretation of the sound as words and the correspondence between the words and their meaning be considered part of the theory. That is, only the physical aspect of the facts that we interpret as linguistic facts must be considered as being the experiment data; and the same requisite holds for the movements that we will consider as being gestures or smiles. This choice automatically satisfies the requirement of the scientific activity that experimental data do not have someone's testimony as constitutive[2].

We are proposing a mapping between the occurrences of physical processes, and the occurrences of mental facts. Both kinds of facts require a time localization, and a measure of time intervals between pairs of them, because a theory can require both the ordering of events, and their distance in time. Thus we shall suitably distinguish between: (i) the content of a mental fact; (ii) the localization in a time reference scheme, which characterizes the occurrence of a mental fact, and its different instances too. The terms “mental” and “psychic” may reflect this distinction, when we decide not to use the two adjectives as synonyms[3] .

We can describe the mental facts occurring in the psychological theory in terms of activities, which then become constitutive of mental facts. In this way we simplify the mapping between the physical and the psychological theory of an intelligent system, because the mapping is, to a large extent, between constitutive activities and physical processes.

A scientific approach is not compatible with a statement of total freedom of the individual,

---

1.. The tendency to consider human behavior as being strongly dependent on external conditioning arose again during recent criminal trials in Italy.

2.. If the subject testimony is constitutive, then we cannot compare the results of experiments carried out on different subjects, because an experiment would become another one when we change the subject; and we find the same limit when the testimony of the experimenter is constitutive.

3.. We can claim many differences between physical and psychic processes. Here we will stress that in a physical process we think of objects as being spatially localized and related. When we add the adjective ‘physical’ to a feeling, say, for instance, a ‘physical love’, then we imagine that the subjects involved participate with their body as well, while the feeling, by itself, does not imply this kind of participation.

at least when we consider it equivalent to assert that we cannot predict individual behavior, because in this way we assert that we cannot test any statement about the future behavior of the subject. Clearly both theoretical approaches, psychological and physical, must provide for the same range of different behaviors and with the same probability. If we maintain the differences stated above between the two approaches, then we will explain the different behaviors by the freedom that we give to the subject of mental activity in a psychological approach, and by the complexity of the system in a physical approach.

The oldest and steadiest approach to the correspondence between a psychological and a physiological description is a relation among functions, which arises from the psychological description, and organs, which pertain to an anatomo-physiological description. However, in scientific theories we prefer to introduce only material or efficient causes and to avoid final causes; thus the results predicted by the theory are more easily tested. In scientific experiments too, where only one variable can vary independently, it is convenient to start the process that in theory is considered to be the material cause, and to look for the expected effects. If we use final causes, then we have a rather intriguing situation, because a final cause is thought to be at the end of the process it causes, and so we cannot immediately transfer it into an experiment.

Furthermore, the introduction of organs in our explanatory scheme is useful only when the organs are specific, ie related to a single function, and each organ has a well-delimited place. Clearly we cannot satisfy these requirements when the activity, which we consider as a function, can vary continuously: since an infinite number of organs would be necessary.

Finally, a one to one mapping between organs and functions would exclude mental schemes, which are commonly used in science and technology. For instance, we can obtain a certain number of functions, which are given as elementary in a computer at the level of assembler, by giving different input configurations to the same logical net. In multiprocessor systems we have programs that automatically distribute among the processors the operations by which we accomplish a certain result, for instance the multiplication of two matrices. In each job repetition a different number of processors may be involved, and the tasks may be allocated to processors in different ways.

The relation organ-function also fails when we think of the function as being realized by an integrated activity of different parts, and we think of the integration as being dependent on external factors. A good example is how cats coordinate their movements to land on their paws after they have fallen off a wall. The movements and their coordination are never identical, because they depend on the initial conditions of the fall.

In the physical description of the biological systems to which we attribute intelligent behavior we must be aware of a significant character. The phenomenology shows that a piece of a cat does not behave like a cat, but a reasonably small amount of water behaves like water. The interaction among the parts should thus have considerable differences in the two systems; and this requires very different theoretical models.

In a theory of systems such as water, and particularly crystalline solids, the interaction may involve only the neighbors of each element. A local theory gives good results, and the conditions at the boundary between the system and the environment determine the state of the system.

We must also introduce into the theory interactions among distant parts of the system, when a part, taken in isolation, changes its behavior considerably, as usually observed in a biological system that we consider to be intelligent. Models become useless in these cases, whose global properties and dynamics follow from statistics where we assumed that the elementary events are independent, or, even, equally probable. We expect acceptable results only from models where strong correlations were introduced among the events; but we also expect the related mathematical difficulties.

Today we cannot really propose to build a theory, which explains the complex behavior of an intelligent system, and yet predicts the interactions of this system with its environment: for instance, the interactions which will give rise to perceptions. Therefore a separation between the system and its environment is practically unavoidable, and we must be satisfied with theories that only give predictions that are conditioned by the occurrence of certain interactions between the system and its environment in the interval of time from the moment at which we make a prediction, to the moment the prediction refers to.

We will now come back to a problem we posed at the beginning of the discussion. Mental activity has an extremely wide range of possibilities, because, when we perform a certain mental activity, we automatically have the mental fact which has that activity as constitutive. We never find a limit in contradiction, because contradiction too is a mental fact. Nevertheless, we consider a behavior as being anomalous and rather psychotic, when we observe a flow of small and disconnected pieces of mental activity; that is, when the behavior has a severe lack of coherence. On the other hand, we consider an excessively stereotypic behavior as being equally anomalous, and this we impute to a poor, or at least excessively fixed, mental activity of the subject. Thus a satisfactory description of systems to which we attribute a sophisticated intelligent behavior, like human beings, must be equally far from these two extremes.

However, we meet another kind of problem when in the description of an intelligent system we map two theories that are based on a psychological and a physical approach respectively. We straighforwardly accept that we have fire as a cognitive fact only if we have the related cognitive activity, and that this cognitive activity will occur only when we have someone who can perform it. Nevertheless we equally accept that the fire burns a piece of wood and transforms it into ash with no dependence on someone's thinking of these facts. That is, these transformations can be neither forced, nor forbidden only by the mental activity of someone who thinks that they will or will not occur.

This state of facts may be explained in terms of cognitive activity by observing that, if we thought that the fire was the subject of the burning activity and of the related consequences, then we must ascribe to the fire the burning a piece of wood, and the transformation of the wood into ash. Otherwise we would be contradicting ourselves, and we decide not to contradict ourselves because we want to make inferences and logical deductions[1]. The occurrence of such processes and of their consequences becomes thus independent of the mental activity through which we think of them; and in this way we do not necessarily introduce two different principles: one for the world of physical things, and the other for the world of psychological and mental facts. We thus avoid a dualism, with the well known philosophical difficulties.

We still have to explain the stability of the object's properties, and the stability of the consequences of the interactions among the objects. This stability is in fact a typical characteristic of mental activity in a subject that we consider as being normal.

We can explain this stability by assuming that we use a paradigm when we perform a mental activity. In the paradigm the objects are thought to have certain constitutive features, and so we normally think of the objects with the features that we gave them in the paradigm. We assume again that certain objects have certain roles in certain processes, and that some facts follow the occurrence of a certain process. For instance in the paradigm the fire is the subject of an activity, e.g. the burning, which transforms the wood into ash.

---

[1] Contradiction is a real limit in deductive reasoning, where it sterilizes the function of deduction; because from a contradiction we can deduce both a proposition and its opposite.

The use of a paradigm in performing mental activities also introduces a certain number of constraints to the course of mental activity. The introduction of constraints is equivalent to asserting that we have correlations among the occurrences of mental activities, and so the activities may occur with different probabilities. In this way we can avoid theories where the behavior is too fragmentary and disjointed; and, if the constraints' scheme is sufficiently rich and flexible, then we can also avoid theories where the behavior is too stereotyped.

Furthermore, the correlations give elements to predict which mental activity will occur in certain conditions, or to predict at least the probability distribution of the possible activities. This aim is typical of any scientific theory, and of any attempt to describe and explain the dynamics of the mental activity. Some elementary examples will highlight the underlying problems.

Let us have a color difference that we localize in the surrounding space, and that we think individuates a physical object. Then we usually think that there is a tactile difference too in the same place, and we expect to find it. For instance, we think of our hand reaching the place, and the change in tactile perception. However, we do not expect to find a tactile difference where we do not perceive any visual difference. In fact it is mandatory to mark clearly the presence of a glass door that is made of a single, transparent sheet; otherwise someone will walk into it\footnote[1].

A television screen is a two-dimensional surface, but we see the rooms and the objects that are presented on it as being theedimensional. The camera lens gives a result that is usually very near to a representation of the room and the objects on a plane perpendicular to the optical axis of the lens, and by following the rules of linear perspective.

Since the Renaissance we have been accustomed to seeing the things that are represented in perspective as being threedimensional; and the great spread, in our time, of images that are produced by optical systems has confirmed this habit. Furthermore, in watching television we also became accustomed to assuming the position of the camera as our observation point. This assumption, and the movements of the camera when filming, reinforce the tendency to think of the represented things as being threedimensional, because we experience effects that are similar to stereokinetic ones. This is a good example of a common situation: an acquired habit leads us to perform a mental activity with a higher probability than the other possible ones.

The two examples discussed above also suggest some directions along which the pattern of constraints and expectations grows.

Very early in our life we get accustomed to adding a spatial location to the color differences that we perceive in our visual field, and to thinking that we shall also find a tactile difference in the same places. This pattern of activities is part of the coordination of visual, motor, and tactile activity we need to hold an object. Nevertheless, this simple rule becomes conditioned by other elements when our experience grows. For instance we do not add a tactile difference when the color differences concern something that we thought of as a plane figure, like in book's illustrations.

The images on a television screen immediately give an example of yet increasing sophistication of the constraints' scheme. Here we agree to add a tactile difference where before we located a visual difference, but our paradigm now distinguishes between the constraints that concern the objects represented on the screen, and the constraints that refer to objects, like the television set, which belong to same environment as our body. If, for

[1]·· We can obtain illusive effects by synchronizing visual, hearing, tactile, and smell stimuli according to the patterns that a person expects. This is the leading idea of virtual reality; where the term 'virtual' highlights that the sources of the stimuli are different from those we assumed as a paradigm for these stimulations.

instance, two of the represented objects collide, then we expect to see also the usual consequences of a collision between two physical objects. However, we do not move from our chair when a car is represented on the screen as coming up to us.

Such behavior requires a high level of thought articulation. The world of our experiences and of our knowledge must be applied to the situations that are represented on the screen; but, at the same time, we must expect very different consequences, from the represented actions, both on us and on the represented objects.

Yet the simple cases we discussed above show the main character of the paradigms: they result from learning, and they are historical both with reference to the history of the individual, and to the history of her / his cultural environment.

The assumption of certain facts and of their connections as a paradigm has several consequences, among which:

- when a behavior occurs that follows the paradigm, then we consider it to be a normal behavior;
- we can accept the occurrence of something that does not follow the paradigm, but we must explain it;
- many factors drive the organization of the paradigm, and we also have practical needs among these factors: so the paradigm results from a choice, but we will change our choice only when we have strong reasons;
- the paradigm may be very articulated, and, like current scientific theories, open to additions and reorganizations.

The language reflects the agreement with the paradigm by chiefly using direct designations and the indicative mood of the verb.

As mentioned above, in the paradigm we have objects that act on other objects. Since the occurrence of their actions and of the related consequences is independent of thinking them, observation is necessary in order to build these parts of the paradigm.

In the paradigm there are consequences that concern our body and the objects of our environment, and they often have a strong impact on our actions and behavior. Due to this relevance to our practical life, we still assume that these consequences and expectations really occur. When their occurrence fails, then we have the usual alternatives:

- we can add new conditioning elements to the scheme, which explain the failure, and we use the extended scheme as a new paradigm;
- we can decide not to pursue the mental activity that we have just carried out, and to substitute it with a mental activity from which the occurred consequences follow; for instance, we usually cease to consider something as being nearer to us than another thing, when further tests do not confirm the result of our perception, and we reverse the categorization of the two things;
- we can cease to consider the mental activity as being predictive of another, and the modified scheme still becomes the new paradigm.

The scientific approach does not change the character of the problems discussed above. It only adds the requirement of studying the things in a way that can be repeated. Thus many of the aspects mentioned above can be found when we examine, for instance, the mental activity with which we build a scientific theory.

A theory is a mental construction; so it inherits the freedom, and the importance of the imagination and of inventiveness, which are characteristic of a mental construction. Nevertheless we stipulate that a theory shall include mutual actions among objects, which depend on the object's actions, and relations among objects whose occurrence depends only on the objects involved. Again the occurrence of these actions must be thought of as being independent of the mental activity of the person who develops or uses the theory, otherwise we have a contradiction. We find the same situation in scientific experiments:

that is, we have actions among objects that have an object as agent, or relations among objects that only depend on the objects involved.
In scientific activity we bind to specific technical procedures the use of the mental categories which occur in theories and in the description of experiments. For instance, we must use the techniques of the geodesy and topography to assert that we measured again the distance between the same two points and their difference of height. ‘Same’ and ‘other’ are mental categories, and their occurrence uniquely requires that someone carries out the related mental activity. However, we decide to use ‘same’ in this context only when certain technical procedures were well suited, otherwise the categorization will be incorrect.
We do not always succeed in finding a suitable technical procedure to which a certain categorization can be bound. A good example is the assertion that a certain volume contains the same physical particles that it had at a past instant of time. We construct theories where the datum is only the number of particles that occupy a given volume at a certain instant of time. Then it is a matter of mathematical technology to use this statement directly or to use equivalent mathematical transformations.
In mathematics, where demonstrations take the role of experiments in physics, we use explicit definitions to codify pieces of reasoning and we introduce symbols and their rules of combination. The demonstration thus becomes similar to a sequence of expressions rewriting, which starts from the initial expression and reaches the thesis expression by using the specific hypotheses and the equivalences among expressions that were previously demonstrated.
A compliant use of the mental categories in scientific theories and in the description of scientific experiments, helps us to infer the occurrence of mental categorization, which is a private activity and so it cannot be directly observed. We can infer the occurrence of the mental categorization by the occurrence of the technical procedures to which we bound the categorization.
In the construction of a scientific theory we further agree to explain all the known facts which concern the specific field. A new scientific theory will have to encompass both the old and the new experimental facts; thus we can correctly speak of evolution when we refer to scientific knowledge. This commitment does not hold in other fields of the human activity, for instance in technology, in ethics, or in politics, where the term evolution assumes a different connotation.
We will conclude by recalling the main points of our discussion. A behavior is qualified as being intelligent as a consequence of a mental categorization. Therefore we have to decide which conditions must be satisfied in order to categorize a behavior as being intelligent. These conditions are usually different in different historical periods.
The systems to which we attribute intelligent behavior can be considered as physical systems too. We can thus use the approach of physics or the approach of psychology to describe their behavior. The two descriptions are irreducible as long as the psychological approach admits that the system (the subject in this context) is the cause of its activities and of its changes, where the physical approach assumes that every change of a physical thing is caused by a different physical thing. We can only propose a correspondence between the two descriptions of the same facts, which must be characterized in a way that is compatible with both approaches.
We think of mental activity as potentially having no limits. Nevertheless, we consider as being pathological behaviors where the mental activity is too disconnected, or too stereotyped. Thus we assume that a normal behavior is constrained by the use of paradigms. The paradigms can also be viewed as correlations among the occurrences of the mental activities involved, and so they allow predictions about the occurrence of a mental activity. The paradigms usually concern the constitutive mental activity of things, the relations

that we expect to hold among the things, and the consequences we expect to follow from certain facts.