The brain and intelligence, natural and artificial

Il cervello e l'intelligenza, naturale e artificiale

ATTI DEL SIMPOSIO INTERNAZIONALE. MANIFESTAZIONE UFFICIALE DEL IX CENTENARIO DELL'UNIVERSITÀ DI BOLOGNA SVOLTA A BOLOGNA NEI GIORNI 4-6 LUGLIO 1988 NELL'AULA DEL CENTRO STUDI SAN DOMENICO. PRESIDENTE GIAN CARLO ROTA MASSACHUSETTS INSTITUTE OF TECHNOLOGY CAMBRIDGE, MASS

A CURA DI OTTAVIO BARNABEI, ALESSANDRO BORROMEI, CAMILLO ORLANDO
UNIVERSITÀ DI BOLOGNA

L'inchiostro blu
Indice

GIUSEPPE MINELLI  Sviluppo del cervello: il pensiero di un evoluzionista 9
ERMINIO COSTA  Toward a neurochemistry of memory engrams: role of synaptic strength modulation 19
W.D. HEISS, G. PAWLICK  Measurement of cerebral metabolism and perfusion for the assessment of brain performance 37
ALDO PASOLO, Cues for understanding plasticity and evolution of nervous tissues 59
ANTONIO CONTESTABILE, MARCO VIRGILI  Constraints and flexibility in brain development and function 67
LAURO GALZIGNA  Aspetti biochimici dell'apprendimento e della memoria 83
MARCELLO BRUNELLI, Neurobiological mechanisms of learning 91
SERGIO FERRI  Fattori endogeni che influenzano l'attività cerebrale: sistema peptidico oppioide 103
ALBERTO MUNARI, Lo sviluppo dell'intelligenza: una «equilibrazione maggiorante»? 111
RENZO CANESTRARI  Intelligenza e creatività 121
ALESSANDRO BORROMEI, MASSIMO KOLLETZEK  Considerazioni sull'intelligenza naturale e sull'intelligenza artificiale. Note sulla situazione attuale e prospettive 129
ALESSANDRO BORROMEI, EUCLIDE DI PAOLO, G.C. MODUGNO, IL METODO CLINICO E I SISTEMI ESPERTI DIAGNOSTICI (S.E.D.) 133
ENRICO MARABINI  Intelligenza artificiale e decisionalità diagnostica 141
OLIVIO GALAZZI, Il comportamento come artificialità: per un profilo storico-culturale del concetto di automa 143
ACHILLE ARDIGO, Computer e società. Informatica avanzata, politiche sociali e condizione umana 155
ALBERTO PASQUINELLI  Scienza e linguaggio 169
P. TOMAS M. TYN O.P., La natura dell'intelletto umano. Un tentativo di analisi filosofico-teologica 173
ROBERT SOKOLOWSKY  Knowledge and its representation in writing, computers, and the brain 197
ED GAMBLE, TOMASOS POGGIO, Visual integration and detection of discontinuities: the key role of intensity edges 209
GIAN CARLO ROTA  Closing words at Bologna Congress, 1988 239
Knowledge and its representation in writing, computers, and the brain

by Robert Sokolowski*

Both of the terms used at the beginning of my title, both the term "knowledge" and the term "representation", raise philosophical perplexities. It is hard to say what knowledge is, and it is hard to say what representation is. If each of these terms is so troublesome when taken by itself, do we not move into still greater difficulty, into perplexity compounded, when we join the two of them into the name of one problem? Do we not add one aporia upon another, and thus make philosophical progress even more unlikely? Not necessarily; perhaps precisely by joining these two problems, we may bring clarity to both. Knowledge may become more understandable when we think about how it can be represented, and representation may become clearer for us when we think not about how a landscape or a face or the citizens of a country can be represented, but when we think about how knowledge can be made present to us. In this respect, the fusion of knowledge and representation may be like the combination of hydrogen and oxygen; by themselves, each of these elements is an invisible, intangible gas, but when brought together and catalyzed they become a liquid that can not only be seen and touched but can also be swallowed.

We are to discuss the representation of knowledge. The most controversial kinds of representations of knowledge are those that occur in the brain and in computers. How can knowledge be stored in neural networks, and what do we mean by knowledge-based computer programs? The brain and the computer seem to present the most urgent problems in the representation of knowledge. But I do not want to discuss these two forms of representation simply by themselves. I want instead to begin with a third kind of representation of knowledge, one that has been with us for a long time and is familiar to us, one that can mediate between the brain and computers. I want to introduce the phenomenon of writing as the form of knowledge representation that

* Catholic University of America, Washington D.C., USA
we will discuss first I will therefore work with three forms of representation: with writing, with the brain, and with computers. The representation that occurs in writing will serve as a catalyst to help us understand the representation that occurs in the brain and in the computer. I should mention that when we speak of these three forms of representation, in each case we make use of the part for the whole. We use the trope of metaphor. When we speak of the brain, we mean the entire nervous system and indeed the entire body; when we speak of computers, we mean to include programs, random-access memory, and disks and tapes; when we speak of writing, we mean not just pen and ink, but printing, cut stones, flashing neon lights, and all the other elements that make up the word that can be read. But for the sake of simplicity we will retain our figure of speech and will continue to speak of writing, the brain, and the computer.

We want to examine the representation of knowledge. What is an item of knowledge? It is important not to turn knowledge into something psychological, something mental or internal or private. An item of knowledge is part of the world, not part of our souls. If I know that Enrico Fermi won the Nobel Prize for physics in 1938, what I know is a part of the world articulated in a certain way. Moreover, the item of knowledge is that articulated part of the world as it has been captured or ascertained by someone, as it has been registered by someone and confirmed as true.

What I call a "captured" fact is a state of affairs that has not only been presented to us, but that has been verified in some way or other, so that we know that it is the case. Furthermore, a known or captured fact is one that is appreciated as confirmed. We present or represent it with the characteristic of having been verified. I use the word "captured" to convey the sense contained in the German word "festgestellt"; a captured fact is one that has been pinned down and made secure.

As such, it is an item of knowledge.

We can have knowledge of individual facts, such as the fact that Fermi won the Nobel Prize in 1938, but we can also have knowledge of more general things. We can know general facts, such as the fact that photosynthesis occurs in the leaves of plants, that the circumference of a circle is the diameter multiplied by \( \pi \) and that stories have a beginning, a middle, and an end. But even in such cases, the general fact must have been captured and confirmed if it is to be known. It is, again, not just the general structure that is an item of knowledge, but that structure as having been ascertained.

To say that an item of knowledge is a captured fact means that it must have been captured or known by someone. Sometimes this someone may be just one person: I may say, for example, "The gold has been buried and I alone know where it is." This fact, that the gold is near the tall oak tree, is known only to me, but it must be known to at least one person if it is to be known at all. There are no unappropriated, unowned, items of knowledge. And if a fact is known to one person, it can in principle be known to others. It is knowable or capturable as such. It could be known to many people, and perhaps even to practically everybody: practically everybody knows that the sky looks blue in daytime.

Now to say that facts must have been registered and confirmed in order to be items of knowledge, and that they must have been registered and confirmed by someone, does not imply that these known facts have been transported into someone’s mind or that they have become psychological things. Captured facts are not brought inside anyone; they do not become Lockean ideas. They remain parts of the world. Some of these items of knowledge may be facts from the past, some may be contemporary with those who know them, some may be permanent conditions of things, but they are all articulated parts of the world, not items within the psyche.

Consequently, when we proceed to represent knowledge, we do not represent merely a mental idea that someone has. We represent articulated parts of the world, but we represent them as registered and confirmed by someone.

Let us consider the representation of knowledge that occurs in writing. Suppose I open a book and read the sentence, "Enrico Fermi won the Nobel Prize for physics in 1938." I can adopt two different attitudes toward this sentence. I can either look at the sentence as an object in its own right, or I can let the sentence let me think about a part of the world; I can let it let me think that part of the world in 1938 which was Fermi’s winning the Nobel Prize. If I look at the sentence in itself, I might consider a certain word as a noun and another as a verb. Or I may look at it just as marks on paper, just as patterned thing, and I may ask, for example, whether this line is curved or straight, and whether this dot is a bit of ink or a piece of dirt. But if I think about Enrico Fermi I intend something other than the sentence here on the paper before me. However, this other thing, this fact is represented in the marks before me. This bit of knowledge is embodied in the marks before me. How can this occur? How can the fact be embodied in these marks?
When we try to analyze such embodiment or representation, we must not overlook the reader’s own ability to shift his focus, from taking the sentence as an object to seeing through the sentence to what it represents. The reader’s activity makes a difference; his activity constitutes the sentence either as marks or as a representation. We cannot speak of the representation of knowledge without paying attention to the reader, the dative to whom the knowledge is represented, and to the power of this dative to shift from one attitude to the other.

A similar shift of focus can occur when I look at a picture. Suppose that instead of reading a sentence, I were to see a photograph of Fermi accepting the Nobel Prize from the King of Sweden. Here again I can consider the photograph as an object in its own right; I can notice gray patches here and white patches there, straight lines here and curved lines there. But I can also change my focus and see Fermi receiving the prize from the king. The picture embodies the event, just as the sentence did; the picture here and now represents Fermi’s accepting the Nobel Prize in 1938.

However, it does seem somewhat wrong to say that the photograph simply represents an item of knowledge, the fact that Fermi received the prize in 1938. The picture is much more ambiguous than the sentence. The picture does not as such indicate its own articulation. It does not tell us what we should notice and read in the picture. Just looking at the photograph, I might think that what I should notice is that Fermi, wearing tails, once stood before a tall, thin, old man who wore glasses, behind whom stood two other men dressed in dark suits. We need words to tell us what the picture is supposed to illustrate and verify. Once the words do give us the title of the picture, once the caption states, for example, “Fermi receiving the Nobel Prize for physics in 1938”, then we know what to look for and see in the picture. And then the picture can give us much further information and much more knowledge. Then we can learn from the picture, for example, that two men in dark suits stood behind the King of Sweden when the prize was given. But without this initial caption, the picture is mute. Without the caption, the point or the gist of the picture is unclear and the knowledge available in the picture remains uncaptured. Knowledge seems to be better represented in words than in pictures, even though pictures may serve to add to our knowledge.

The reason pictures are more silent than words is that, as we have seen, knowledge is not merely the presence of things, but the presence of things that have been articulated and registered and confirmed by someone, and when we have only a picture, we do not know what is being confirmed in it. The picture just represent things which need to be thought through or articulated if they are to be registered and confirmed. We have to think through the picture, just as we have to think through the things that we see, in order to constitute an item of knowledge in the picture or in the things.

The words signal to us how we are to articulate the scene before us, whether depicted or real. And it is especially the grammar of the words that signals to us how we are to articulate: a grammatical conjunction signals us to take two or more things together, as one or as a group; a copula signals us to take an object as featured in a certain way; an adverb signals us to modify a feature or a process in a particular fashion, and so on. The written words tell us what to intend and how to intend it. If we or someone like us were not there to let the words carry out their signalling function, if we or someone like us were not there to provide uptake for the words, then the words would not embody knowledge and pictures would not convey information.

And both words and pictures can be taken either to represent something or to be objects in their own right. Our ability to oscillate in our focus, to move from taking the word or picture as an object in its own right to taking the word or picture as representing, permits representation to occur. Only because we can move back and forth in this way do we establish a body in which the knowledge can be embodied. Without our ability to change focus, there would only be marks on paper; there would not be a body for pictures or words, and there would be no representation.

### III

Now let us turn to the representation of knowledge that occurs in the brain. When I intend the fact that Fermi won the Nobel Prize in 1938, there are neural activations in my brain and nervous system. Chemical and electrical processes occur there. Other processes occur when I see and recognize something such as my house or my car. Let us call these neural activations the “brain-word” or the “brain-picture.” More generally, let us speak of the “brain-sign” to cover both brain-words and brain-pictures.

In some sense, brain-signs represent facts and things. Brain-signs represent knowledge. But there is an important difference between the way brain-signs embody knowledge and the way written words and printed pictures do so. First of all, if I am the subject for whom the representation occurs, I do not and I cannot perceive the neural activities through which the fact or the thing is presented to me. I cannot focus on the brain-sign in the way I can focus on the printed word or picture; I cannot focus on the brain-sign as an object in its own right. The brain-event is totally transparent to me. I can intend only the fact that Fermi won the Nobel Prize, I can intend only the automobile that
I now perceive, I cannot oscillate between focusing on what I intend and focusing on the brain-sign through which I intend it. There is no embodiment visible to me; there is no internal word or picture perceptible to me between me and what I intend.

But suppose there is an observer who examines my brain and nervous system while I intend a fact or a thing. He, the neurologist or neurosurgeon, will see a particular pattern of neuronal activity; he will see the brain-event or the brain-sign, but he will see it only as a physical thing. For him it will not function as a word or a picture. He will not see it as the name of Enrico Fermi or as an image of a cat or a house. This observer could not intend a fact or an object through the brain-sign. This restriction would exist even if neuroscience developed to such an extent that it could correlate a particular brain-event with the experience of a particular thing; even then, the observer would see only an effect of the thing being perceived and a neurological activity of the subject perceiving. The brain-event would not be a word or an image of the thing for him than smoke would be a name or a picture of the fire that causes it.

Thus whereas the brain-event is purely transparent to the subject whose brain it is, the brain-event is totally opaque to the observer who examines the brain. Neither the subject nor the observer can be related to the brain-sign in the way one can be related to a public word or picture: neither can oscillate between taking the brain-sign as an object and taking it as representing what it represents. Instead, these two ways of focusing are distributed between two persons: the subject can only see through the brain-event, since for him the brain-event is transparent, and the observer can only look at the brain-event, since for him it is opaque.

How then does the observer know that this neural activity is associated with knowing a certain fact or seeing a certain thing? He knows it because the subject can tell him what he, the subject, intends or perceives, or because the observer himself can see the same thing that the subject is seeing. This brings us to an important characteristic in the observer. Besides being engaged in examining the subject's neural system, the observer is, in principle, also an interlocutor with the subject. And still further, the observer can also perceive the same objects and articulates the same facts that the subject perceives and articulates. The observer can, as a matter of principle, adopt three different attitudes: he can be a simple observer of the subject's body and neural system, he can enter into conversation with the subject, and he can experience, from his own point of view, the same parts of the world that the subject experiences.

This range of possible stances, this plasticity in the observer, is one of the factors that establishes the problem concerning the function of the brain-sign. We are inclined to see the problem of brain-processes as merely a problem in physiology and psychology, but we must not overlook the role of the observer in constituting the various dimensions that let the problem arise. Only because the observer can also see the same world that the subject sees, and only because the observer can talk with the subject and ask him questions and receive answers, can he, the observer, raise the question about how brain-events can function in knowledge. Only because the observer can adopt a non-neurological attitude can he raise a neurological question. These brain-events are interesting to him and they are isolated by him only because he knows the subject is knowing a fact or seeing an object, and because these brain-events seem to be involved in that knowing and seeing.

The complex constitution of the neurological observer raises an interesting moral issue. Because the observer must in principle be able to converse with his subject and quote his subject's perspective on the world, the observer can never treat the subject merely as an organism, merely as an object. The neurological observer can never be totally detached. Even to study the subject's neural system, the observer must in principle be able to speak with the subject and acknowledge the subject's perspective on things. He must treat his subject as an equal, as another person.

How does knowledge enter into this relationship between observer and subject? How can the neurological observer say that these particular neural processes, these brain-signs, represent knowledge of some sort for the subject? The knowledge in question can never be read off the brain-sign itself. The brain-event does not function as a word or a picture for the observer. The observer is informed about the knowledge about the captured and confirmed facts by the subject as a speaker and interlocutor. Only when the observer enters into his conversational mode of being and into his consideration of the parts of the world his interlocutor tells him about does he, the observer himself, grasp the knowledge in question. It is not the brain-event that represents the knowledge to him; it is the subject, as a whole human speaker in public space, as someone who uses words to convey knowledge, who can bring about the representation of knowledge. The brain cannot do it alone.

Thus when we say that the brain 'represents' knowledge, we mean to say that the brain functions as part of a public conversational activity carried out by someone who speaks and thinks. To achieve knowledge is to make a move in a public network; to achieve knowledge, to capture a fact and make this fact available as known to ourselves and to others, is rather like the public activity of saluting a military officer or making a move in a game of chess. To achieve knowledge is not to establish an internal representation, an internal copy, of a
public fact. The brain-sign is a neurological component of a public action; it is like the neurological component in a military salute.

It would therefore be quite wrong to say that items of knowledge are stored as internal representations in the brain just as it would be wrong to say that a military salute is somehow stored as a representation in the brain. The whole body is involved in a salute; the salute is achieved by the soldier as a single, whole, public actor, one placed in a situation in which a salute is appropriate. Likewise, the achievement of an item of knowledge is carried out in a context in which a confirmed fact is called for, and it is carried out by a single, whole speaker or agent or identifier. The brain-process is a part of this bodily public activity of making a display. The brain-event may have a controlling role in the achievement of knowledge, just as another brain-process will have a controlling role in a salute, but the brain-event itself is not the knowledge, nor is it a little copy of the knowledge.

And in trying to work out how the brain-event can have a controlling role in human knowledge, we must not overlook the involvement of the neurological observer, who in one respect is a simple observer examining the physical organism of the brain, but in another respect is also an interlocutor with his subject and a co-perceiver of the world with him. This observer has to bring together two different kinds of things: he must be able to bring together the public activity of capturing knowledge in speech, and the organic process in a brain-event. What distinctions in attitude or focus are required for the observer to carry out this identification? The observer's role in this neurological problem is as strategic and as mysterious as is the role of the observer in the measurements of quantum physics.

To close this discussion of the relationship between the neurological observer and the person whose brain the observer examines, let us draw one more distinction. When the observer and the subject engage in conversation, the subject must manipulate words in order to represent what he wants to say. He must speak or write. His manipulation of words is a responsible activity, an activity that he initiates. But when the observer merely examines the brain of the subject, the subject does not manipulate the neural processes that the observer observes. The subject does not communicate with the observer through these neural activities. The observer is not engaged in reading a message in the brain from the subject about how the world is. Our brain-signs are not something that we say; we do not tell an observer anything through them. The brain-events merely happen in us and to us, we do not shape them in the way we shape our words. Thus the embodiment of knowledge in the neural system must be distinguished from the embodiment of knowledge in writing and speech.

Now let us turn to the third sort of representation of knowledge, to the kind that occurs in the computer. The way knowledge is represented through computers is, in many respects no different from the way it is represented in writing. If a lot of information is stored on a disk or in a computer memory bringing this knowledge to the screen is not very much different from looking it up in a book. In both cases we read words or symbols and understand the captured facts, the knowledge, embodied and signalled in the words and symbols. In this sort of retrieval, furthermore, the computer does not guarantee that the information is indeed knowledge, that it is verified fact. The responsibility behind the knowledge lies with those who registered and confirmed the facts and stored them in the computer, not with the computer itself, just as the warrant for the knowledge represented in a book lies with the author, not with the publishers or with the book as an object made of paper, ink, and glue.

The computer differs from the book in that it may somehow be able to generate new knowledge, either by registering on its own how certain things are, or by deriving new knowledge from what has been stored in it. Let us look at both of these ways in which the computer seems to amplify knowledge. (1) It is possible that computer programs may carry out inferences that would be difficult for human beings to make. A computer could activate a very large database and extremely complex rules of inference, and it might reach conclusions that a man might not be able to reach. The computer may appear to have added to the store of knowledge by doing this; it may appear to represent new knowledge. But the only reason the conclusions can be called knowledge is that they can be derived from premises that have been warranted as knowledge. Inferential reasoning as such does not establish knowledge; it is parasitic on the knowledge from which it begins. If the conclusions are recognized as expressing knowledge, their knowledge is dependent on the semantics and truth-value of the premises. And the premises from which the inferences begin are warranted by those who registered and confirmed and captured the facts expressed in the premises, not by what the computer has done with the premises.

(2) If computers cannot generate knowledge by carrying out inferences, perhaps they can do so by something analogous to perception. Perhaps they can ascertain that certain things are the case and thus establish premises for inferences. For example, a computer might be built to sense the temperature of the air around it and to generate certain expressions that represent the true state of the environment. Would this not be the same as capturing a fact and representing knowledge to someone?
It would not be the same. The computer affected by the temperature is not really different from a thermostat which is set to turn on an electric heater when the temperature falls to a certain point, and it is not really different from a scale that records one's weight on an LCD readout display. Like the thermostat and the scale, the computer undergoes an effect from something in its environment, but something in its environment is not displayed to it. The computer does not articulate an appearance to anyone, neither to itself nor to anyone else.

What is the difference between undergoing an effect and undergoing an experience? When I perceive and state that the door is closed, obviously I undergo an effect from my environment. My nervous system and my body as a whole are affected. But in addition to being affected by this situation, I also display the situation. Within my context, I draw attention to a part of the world, to this door, and I articulate its feature or condition: I display it, and so it is displayed to me, as being closed. What I display is expressed in the words I utter, but my words, "The door is closed," are more than an effect of the door's being closed. They signal my activity of displaying and articulating.

But why can we not say that the computer draws attention to the room and describes it as being cold? Why does the computer not display the room as chilly? What the computer cannot do is to identify a fact as one and the same fact that can be perceived again and again, that can be perceived by different perceivers, that can be referred to again and again in its absence that can be intended in its absence and then confirmed in its presence. The computer cannot identify a fact it cannot recognize as the same again. For this reason, the computer cannot achieve knowledge, because knowledge requires the confirmation of facts as assuredly true, and such confirmation requires that we can go back to the same fact to verify it, and that in this process we recognize the fact as the same. Each time the computer responds to input, it responds purely and simply as an effect of what is going on now, whereas when we perceive and articulate something we appreciate it as having been anticipated in the past, and as capable of being returned to in the future. We identify and reidentify, we do not merely respond. We also appreciate that the same fact we are perceiving can be perceived and articulated by others, and so we can warrant the truth of the fact for others.

We carry out this articulation of facts by manipulating words and other public signs. The words we choose and the way we arrange them enable us to intend certain facts again and again and to convey them to others as knowledge. We can even leave these words behind us in writing as a shell or carapace of our thinking, a shell that can lie dormant for a while and then come to life again, by prompting another thinking of the same facts whether by ourselves or by others.

Thus our representation of knowledge involves a manipulation of words, and it also involves an interweaving of past and future in our changing present, as well as an interweaving of the perspectives of others with our own. In order to represent knowledge, we must be able to live not only in our immediate present environment, but in the past and future of our world and in the mind of others. Only in this expanded context is there space enough, are there dimensions enough, for knowledge to take place and for presentation and representation to occur. In contrast with our rich and expansive motions through so many different dimensions, the computer is impoverished and constricted in the way it can move. As a Turing machine, it can at any instant only respond to the input immediately at hand, and then go on to another part of the program to receive yet another discrete command. The computer is not still in the past and already in the future, and it does not entertain perspectives other than its own; hence the computer can react, but it cannot display or communicate and cannot appreciate knowledge for what it is.

V

The brain sciences, the computer sciences, and the concern with writing are all specialized disciplines. They each mark off a domain for themselves—the brain, the computer, the written word and they develop a terminology and expertise proper to their domain. They put on blinders; they set limits to what they are to examine and they flourish as arts and sciences because they are partial. Their success is a function of their specialization. But sooner or later the partiality of these arts and sciences leads to distress. Questions arise concerning the relationships among the special disciplines, and other questions arise concerning the context from which they have been marked off.

In my paper, I have tried to turn our attention to the wider context within which we must situate the brain, the computer, and writing. Philosophy pulls against partiality; philosophy is by definition the attempt to restore a more comprehensive context to turn back from abstraction and to recover the more concrete whole. Philosophy generates neither a professional terminology nor a special expertise. It tries to work out distinctions between things that might look similar and to bring together things that might appear diverse when they are seen from the viewpoint of the partial disciplines. Only in the light of this wider context, not in any of the special arts and sciences, can we clarify appropriately what it is to be an interlocutor and an observer what it is to perceive and to understand, what it is to write and to read. And only in this wider context, in which no dimensions are to be left out can we clarify what is meant by knowledge and its representation.