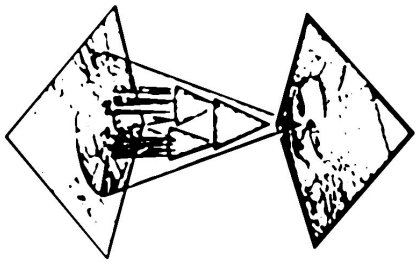
Intentional robots
The design of a goal-seeking, environment-driven, agent

Riccardo Manzotti
LIRA-Lab, DIST, University of Genoa



This work has been carried out by Riccardo Manzotti, during his Ph.D. course in Robotics, under the supervision of Prof. Giulio Sandini at LIRA-Lab, Department of Telecommunication, Computer and System Sciences, University of Genoa, Italy. © 1997-2000 Lira-Lab

The research described in this book has been supported by grants from the Italian Ministry of research and University (MURST), the European Union (projects: VIRGO, SMART, SVAVISCA, NARVAL, ROBVISION) and by the Italian Space Agency (ASI).

Printed in Italy.

Contents

What a surprising number of philosophers of language have said […] is: "If there are deep and difficult problems about representation, then we won't have any representation." And what a surprising large number of philosophers of mind have added is: "if no representation means no belief/desire psychology, then we won't have any of that." Chorus: "We all keep a respectable ontology; troublemakers not allowed."

Jerry Fodor[1]

… as the Eye chases its own gaze through the labyrinth, leaping quantum gaps that are causation, contingency, chance. Electric phantom are flung into being examined, dissected, infinitely iterated.
[…] a thing grows, an auto-catalytic tree, in almost life, feeding through the roots of thought on the rich decay of its own shed images, and ramifying, through myriad lightning branches, up, up, towards hidden lights of vision,
Dying to be born,
The light is strong,
The light is clear,
The Eye at last must see itself,
Myself …
I see:
I
!

William Gibson[2]

It's easier to hide a problem than to solve it.

Ludwig Wittgenstein[3]

[1] (Fodor 1987), p. xi

[2] (Gibson and Sterling 1991)

[3] (Wittgenstein 1995), p. 148.

Foreword

We are conscious beings. In a Cartesian sense there is no empirical fact prior to this one: in order to know something we must be conscious of this something. Here we do not use the term 'knowledge' in the sense commonly prescribed by epistemology (that is a well founded and true belief). We use the term 'knowledge' to denote the act of having something as the object of our consciousness. Being conscious of something entails that, as subjects, we are something and, that being conscious of a particular thing entails being something particular: cogito ergo sum. *Is this argument reasonable? We claim that it is. And we will try to show how it is possible to build a better ontology by using an improved version of the Cartesian cogito.*

There is a different line of reasoning that we believe is capable of leading to the same conclusion. This line of reasoning is less metaphysically biased. Its starting point is the empirical fact that nature, through evolution, has selected organisms capable of being conscious (human beings). Was this a random choice or did it correspond to some necessity in the development of highly complex organisms? There must be some formidable reasons why human beings are conscious and conscious of themselves.

The aim of this thesis is to show that consciousness is central rather than marginal to human development. Consciousness is usually seen as a curious by-product of the brain: something that is produced by the cortical activity and that will eventually be explained by the progress of neuro-physiology. We believe this is a big mistake for two different orders of reasons. A first order relates to the obvious fact that evolution selected human beings so that they would be conscious of their actions and their environment. There must be a number of sound and compelling reasons to make consciousness one of the first points in the agenda of a subjects' designer. There is another order of reasons why consciousness cannot be underestimated: these reasons are related to the fact that consciousness is a more pervasive part reality than is usually admitted. Such reasons point directly to the fact that many concepts, concepts we currently deem to be objective and self-consistent, depend for their very conceivability on the role of consciousness. Consciousness is the point where the world and its representation become one and the same; and for this very reason it provides the most valuable and dependable insight into the structure of reality itself.

This thesis is developed along two separate rationales. On the one hand it lays out a broad framework intended as a criticism of the classic objectivistic framework of science, which also sets the necessary foundations for developing a theory of a conscious subject. On the other hand, this framework is used to implement and develop an artificial

subject. The term 'artificial subject' stands for a subject, which has been brought into being by a voluntary effort, without resorting to biological reproduction. This does not mean that such a subject is different in its subjective nature from a normal subject. It could have different contents but, as long as a subject is real, it is, to all intents and purposes, a subject. It is like talking about mass: as long as something has a mass it must have a real mass. Can we produce artificial water? If we find a way to synthesize water from Oxygen and Hydrogen the product would be as natural as the water that fills rivers and seas. In this case the fact of being artificial would be merely a matter of historical interest: it would not be a substantial or ontological property.

We think that if reality is capable of producing a conscious being through natural selection there are no practical and a priori reasons why the same could not be done in an artificial being. If this attempt were successful, the resulting subject would be a real subject. Once the principle of flight was understood, flight itself could be reproduced; similarly after consciousness has been mastered, a conscious being can be built.

Can we imagine anything more marvellous than the fact that we are conscious? Nothing else in nature is so uniquely baffling. Except for our being conscious, everything that happens in nature in principle, is explicable. It could be difficult, extremely difficult, to find such an explanation but in the end, supposing that enough data is collected, supposing that the appropriate experiments are carried out, it is only a matter of defining what the causes are and what the effects are. In the case of consciousness the problem is that there is no possible cause that can produce it or, alternatively, that consciousness cannot be the effect of any known natural cause. Consciousness is outrageous in so far that it actually rejects the constituted order of science. Nevertheless consciousness is in itself a fact, and trying to deny its reality seems a reckless if not downright useless option. Consciousness appears to pose an impossible problem not because of consciousness itself but in consequence of what we believe nature to be. Over the centuries scientists have developed an objectivistic ontology of nature that is incapable of dealing with the subjective side of it (it follows that it is incapable of dealing with subjects and, therefore, with conscious subjects). Yet, a few empirical facts must be stated: i) human subjects can be defined only in relation with their being conscious subjects; ii) consciousness is seen as the capability of having representations and as the capability of unifying parts of reality.

Why do we perceive an intuitive difference between reality and the representation of reality? Because we describe the former using a reductionistic ontology while we know the latter through non-reductionistic first-person experience. What proof do we have when we affirm that the world is made up of atoms? We have only direct empirical evidence or indirect phenomenal experience gained from objects like instruments, probes and the like. We know that there is an external object made up of atoms, some physical events in the middle, and some chemical activity going on in our brain. We know from empirical first-person experience that, through perception, we somehow can represent the

external world: that is, we get the meaning of an external object. Unfortunately, given the existing extensional and reductionistic ontology, the following problems arise:

- *What we are experiencing cannot be in the external object because such an object is physically different from our brain. Therefore it cannot determine any difference in the quality of what is going on inside the nervous cells of our brain cortex. Besides, it cannot be counted as an external object because such objects do not exist as real unities.*

- *What we are experiencing cannot be in the middle because i) there is no meaning in the physical medium between the brain and the object; ii) if an event is 'in the middle', it is still* outside *of the brain and thus it still an external object.*

- *What we are experiencing cannot be in the brain because i) there are no intrinsic unities in our brain that can correspond to the perceived unities of our experiences; ii) properties of objects in our brains are different from properties of perceived objects (brain matter is a dark, bloody grey matter while the world is luminous, colourful, full of taste and smell and so on); iii) there is no real boundary between the inside and the outside of the brain: there is no strong reason why we should suppose that an event occurring in some space location (inside my skull for example) should become part of someone's experience.*

Given a reductionistic ontology there are no real unities (no surprise at all: that is the very purpose of a reductionistic ontology). Yet we have first-person direct evidence of the existence of unities, so why should we reject such evidence? What is a representation? It is a unity of content. And what is content? Content is a portion of reality. Supposing that it were not, this would entail some kind of dualism. There is no evidence that the content of our conscious states is located anywhere. For example is it conceivable to modify the physical location of my brain without also modifying my conscious state? My conscious states certainly do not depend on the where *and the* when *of my brain activity. What is represented by my conscious states is located in space and time, but from a logical and phenomenal point of view, there is no evidence of the location of my conscious states as such.*

The goal of this thesis is twofold: first to look for a new framework capable of describing subjects and, secondly, to test such a framework by applying its predictions to the construction of an artificial subject. It is conceivable that such a problem posed to our conscious existence cannot be solved using progressive steps. An objective world, abstractly and metaphysically imposed, is ill suited to explain subjectivity. Yet, any and all attempts to understand consciousness must per force be tested empirically.

Finally I must spent a few words to thank Giulio Sandini whose support, both intellectual and practical, has been vital during my PhD; without his help and advice I would have never written this thesis. I must especially thank Penelope Hammond Smith for her help in cleaning my English.

Moreover I wish to dedicate this book to my parents, Iolanda and Bruno, who taught to me to believe in my ideas.

Riccardo Manzotti

Genova, May 2001

1 Robots as Subjects

> *I think that the conscious mind is the most important subject imaginable. We are at the beginning of the neuroscientific revolution. At its end, we shall know how the mind works, what governs our nature, and how we know the world.*
>
> Gerard Edelman[1]

> *Consciousness is the biggest mystery. It may be the largest outstanding obstacle in our quest for a scientific understanding of universe.*
>
> David Chalmers[2]

> *Talking about the mind, for many people, is rather like talking about sex: slightly embarrassing, undignified, maybe even disreputable. "Of course it exists," some might say, "but do we have to talk about it?" Yes, we do.*
>
> Daniel Dennett[3]

There is a traditional distinction between subjects and objects. In our culture the boundary is extremely important. Subjects have features that are not shared by objects. For example, subjects have rights while objects can be treated in any conceivable way; subjects can own objects but no subject can be owned; subjects have their own values while objects have no intrinsic value; subjects have a mind, objects do not. Important as it is, this boundary has not been objectively defined rather floats backwards and forwards. What entities can be seen as subjects? Historically, human beings have been seen as subjects. Besides the idea that human beings had a soul was generally accepted. The relation between the body, the mind, and the soul rapidly became confused: the inability of being identified, as a human being (and therefore as a subject), has been one of the most obnoxious in history[4]. There are three correlated concepts:

---

[1] (Edelman 1992), p. xiii.

[2] (Chalmers 1996), p. 13.

[3] (Dennett 1987), p.1.

[4] As an example we can see the effect of the exclusion of some category of people from humanity: the slaves in the Roman Empire, the Indios before the 1537 Treaty of Pope

being a person, being a human being, and being a subject. The first concept is giuridical and can be defined as such. The second depends on the presence of a particular class of genetic codes. The third is what we are concerned with. It depends on the existence of a real subject of experiences. The practical difficulty in determining its existence has provoked a *de facto* equivalence between the status of human being and the status of subject. This can be questioned for several reasons:

- Being a human being is an anthropomorphic principle without any *a priori* justification. Like the Ptolemaic idea of the earth at the centre of the universe, it might prove itself wrong[5].
- There have been species, different from our own, that showed evidence of being real subjects. For example, specimens of *Homo Neanderthalensis* buried their relatives[6].
- Several species *mutates mutandis* (cats, dogs, dolphins, monkeys) could deserve the status of subjects[7].
- In the future, there could be machines functionally equivalent to human beings. Should they be considered real subjects?
- There are living organisms genetically-human that do not show any evidence of being subjects (clinically dead patients, anencephalic patients).
- There is not any *a priori* connection between the presence of a particular kind of biological material (containing a particular DNA) and the presence of a subject.

There is a natural criterion to distinguish between subjects and objects. The first ones have the capability of having experiences, of being aware of what happens to them and around them. They are «beings in the world», using Martin Heideggers' terminology. In simpler words, they are conscious. On the contrary, objects do not have experiences. They are always unconscious. As a proof of this criterion it is possible to consider that, if a human being is reasonably considered incapable of recovering his/her consciousness, he/she is

Leone III, and the Jewish people during the last war. Related to the difference between subjects and objects is the distinction between persons and objects.

[5] (Khun 1962).

[6] (Trinkaus and P. 1992).

[7] (Allen 1997).

considered clinically dead. The being is no longer a subject but has become an object (the internal organ can be assigned to other humans). Yet this natural criterion is obscured by the fact that, as many have noted[8], to date there is no clear idea of how to deal with the problem of consciousness. Although there have been several recent attempts to face the emergence of consciousness scientifically[9] there is still no consensus about the kind of methods that should be employed. Someone even argued that there is a sort of epistemic gap between the subjective domain and the objective one[10]. And someone even argued that the relation between the two will be forever unknown to men: a modern *ignoramus et ignorabimus*[11]. The problem of the nature of phenomenal consciousness has become so obsessively difficult that has become known as *hard problem*[12].

Nevertheless, it seems that there is some kind of ontological mistake that thwarts any attempt to deal with consciousness explicitly. The aim of this work is to understand why it is so difficult to approach the problem of subjectivity and, then, to propose an alternative framework that could cope with conscious subjects. This proposal of ontological revision must not remain a sterile metaphysical project but must be tested empirically. Two are the scientific fields in which such a proof can be looked for: neuroscience and robotics. The first field, by studying the only objects that correspond to conscious subjects (that is human beings), can be helpful both as a source of evidence and as a test-bed for predictions. Robotics is another natural field in which experiments might be carried out. If there is a theory of mind, which sets the conditions by which an object could let a subject emerge, such conditions could be replicated. Hitherto, there have been only a few sparse attempts to understand and propose an architecture capable of producing a conscious robot[13].

---

[8] Among the others: (Chalmers 1996; Kim 1998; Edelman and Tononi 2000).

[9] As the editor of Nature Neuroscience wrote «Times are changing. [Hard scientists] hope that by combining psychophysics, neuroimaging and electro-physiology, it will eventually be possible to understand the computations that occur between sensory input and motor output, and to pinpoint the differences between cases where a stimulus is consciously perceived and those where it is not», August 2000.

[10] (Levine 1983).

[11] (McGinn 1989).

[12] (Chalmers 1996).

[13] (Aleksander 1994; McCarthy 1995; Aleksander 1996; Martinoli, Holland et al. 2000)

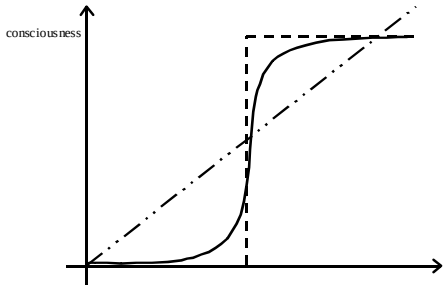


Figure 1-1 Up to now, it is not clear what the essential element that produces phenomenal consciousness is. It is not clear if consciousness arises gradually or abruptly. In the figure there are a few possible curves. For example the dotted line corresponds to *on-off* theories of consciousness. It must be stressed that there is no consensus about what the dimensions along the horizontal axis (genetic code, biological structures, complexity, transcendental soul).

1.1 Intelligence is something that does, consciousness is something that is

The idea of creating an artificial being focused on the capability of replicating human behaviours. At the beginning, when the field of Artificial Intelligence was first being developed, in the '50, Turing proposed his famous test that implicitly stated that being a subject means being able to behave like a human being. This principle flourished in a period in which behaviourism was to be abandoned only to be substituted by several elaborate forms of functionalism. The emphasis was on behaviours, activities, functional relations, information processing. The problem of this approach is that it is based on the existence of human beings that could provide the necessary goals for artificial machines. We will claim that 'intelligence' cannot be defined autonomously from conscious subjects. With a slogan it is possible to claim that

> *intelligence is the capability of finding the procedure to obtain a certain goal given certain practical constraints.*

There could be no intelligence without constraints, goals and the possibility to act. When a behaviour is considered as intelligent, all of these three elements must be present. Imagine a giant whale eating plankton simply by swimming in

the ocean. While this behaviour is perfectly coherent with the main goal of the whale – i.e. eating –, it is difficult to consider the whale intelligent. The presence of constraints is a way of expressing the inherent difficulty of a task. The more it is constrained and the more difficult it is. Imagine a squirrel that must rest hide and then find its food in a forest. This behaviour seems much more intelligent than that of the whale. The squirrel must avoid predators, find suitable resting places, and remember where they are. The number of constraints for each of these tasks is relevant. In other words, it is much easier to make a mistake for the squirrel (hiding food) than for the whale (eating plankton). Yet constraints are not enough: a goal is needed. If we consider someone whose behaviour is aimless, we would not admit any intelligence in it. Intelligence is useful only if a goal is to be pursued. Finally, there must be some activity (mental or physical). Let's think of the classic 'intelligent' games like chess. There is a conventional goal that is defined as a class of configurations of pieces that states one player's victory. There are constraints of all kinds: number of pieces, rules, chessboard size. Finally, there are the possible activities embodied into the players' moves. The more the task is difficult and the more the solution is seen as the testbed for intelligence. Here, the conclusion is that these three elements (a goal, some constraints, and possible actions) are all needed to define intelligence. In the animal kingdom the absence of particular constraints reduces the need to select particularly smart individuals.

The problem that arises at this point is that intelligence by itself does not provide any purpose. It only helps in achieving it. A different candidate – who can explain what goals are – must be proposed. Intelligence by itself is an empty concept that is not self-consistent. It is just a label that we use to denote a series of solutions to problems (given goals and constraints).

If we look at ourselves as conscious subjects, we will find that we *are* before being intelligent. The previous claim about the nature of intelligence can be thus restated as follows.

Intelligence is something that does, while consciousness is something that is.

The distinction between intelligence and consciousness is a distinction between two complete different non homogeneous and partially unrelated concepts, . Of course, this does not mean that what is usually termed 'intelligence' is not important in order to develop a conscious subject. It means that they are two different aspects of subjects that must remain as such. Their division is correlated to the famous distinction between psychological (or cognitive) and

phenomenal mind. This distinction has been epitomized by David Chalmers' *hard problem*. He stated that

> At the root […] lie two different concepts of mind. The first is the phenomenal concept of mind. This is the concept of mind as conscious experience. […] The second is the psychological concept of mind. This is the concept of mind as the causal or explanatory basis for behaviour. […] On the phenomenal concept mind is characterized by the way it *feels*; on the psychological concept mind is characterized by what it *does*[14].

The same duality was addressed by Schelling «Intelligence is productive in a double manner, either blindly and unconsciously or freely and consciously; it is unconsciously productive in Weltanschauung and consciously productive in the creation of an ideal world»[15]. The cognitive side is related to the having a series of skills (behavioural, computational, functional) while the phenomenal side is related to our being something. The very fact that subjects *do* exist is a problem. The understanding of subjects require the understanding of the nature of existence as such.

1.2 What is a subject?

Two properties are suggested as essential to the existence of subjects: being a *unity* and the fact of being capable of *representing*. Both properties are based on the empirical evidence that is provided by human beings as subjects. Everyday human beings experience these two essential qualities of their being conscious (their being in the world). Both are empirical facts and – as such – they must be explained, not hidden. Any theoretical framework that tries to deny them, singly or jointly, is a metaphysical project aiming at superimposing a point of view unsupported by experience over empirical evidence.

Implicit classical metaphysics of the XXth century (orthodox reductionistic objectivist ontology[16]) is badly suited to deal with unity and with representation. As it will become clearer in subsequent chapters, if an objective world is postulated, there are not any suitable candidates for unity and for representations. As a consequence, attempts have been made to eliminate the

[14] (Chalmers 1996), p.11.

[15] This was quoted in (Heidegger 1988), p. 5.

[16] «I declare my starting point to be the objective, materialistic, third-person world of the physical sciences. This is the orthodox choice today in the English-speaking world» (Dennett 1987), p. 5.

source of such embarrassing paradox, that conscious subjects are something real. We claim that a different approach must be followed: to reform a purely objective ontology as something insufficient to explain reality.

If dualism is rejected, subjects must be a part of reality. What kind of part? They have two properties that are not shared by any other known physical object: they are unities and they represent the world. No physical object seems to possess such capacities. In fact both the existence of unities and the existence of representations in the physical world seems to be based on the existence of subjects. If these two terms really correspond to what distinguishes subjects from objects, we can claim that it must be understood what is unity and what is a representation in order to understand what is a subject. Building an artificial being means building a structure capable of letting unities and representations occur. As a concise statement, we can claim that

a subject is a unified set of representations.

Of course, this statement shifts the burden of the definition of what a subject is on the definition of what unities and representations are. If it will be possible to find suitable candidates for these two aspects, the statement could be taken as a rough sketch of a constitutive formula for subjects.

1.2.1 What is unity?

There is nothing more intuitive and simple than the simple fact of unity. In any conscious experience (let's think of perception for example) anything experienced always constitutes a unity of something. I am looking at a landscape: I see a tree, a mountain, a group of hills, and a cluster of clouds, a pedestrian walking along a path. Even when I am looking at something, which is a set of other entities, it is because I perceive it as a unity made of other unities. The multiplicity is derived from the elementary unities, which constitutes it as a whole. There is no multiplicity without the perceptions of smaller unities.

Furthermore, the sphere of my experiences is unified under the bigger unities of my experience. I can say, like Descartes, that the perception of my being a subject is the perception of the unity that coalesces my perceptions[17].

[17] Two caveats must be made at this point. First there is more, inside a subject's conscious sphere, than just perceptions: for example concepts, thoughts, beliefs, and feelings. Secondly, there has been a widely used rationale, started probably with Hume, against the supposed direct perception of the subject as a whole. In this chapter, it is not

Even if the subject could not be perceived as a whole, its contents are perceived under some kind of unity. We can imagine the several contents of my experience occurring separately in different conscious subjects. I am looking at a tree and at a table. I am conscious of both of them at the same time. If I was looking only at the tree and *you* were looking only at the table, the same conscious contents would occur in two separate conscious unities: it would not be the same. The fact that, at once, I am conscious of many things is the expression of the existence of some kind of unity that requires a proper explanation.

More importantly, unity cannot be further decomposed. It is an original fact of experience and there is nothing mysterious about it. Simply, the world seems to be composed by entities that have intrinsic unity. Denying such fundamental and empirical fact entails the unreality of the experienced unities. The negation of experienced unities entails that empirical evidence should be rejected on the basis of some *a priori* abstract principle which is considered to be more important than experience itself, which is an example of bad metaphysics.

This framework incapable of dealing with unities derives from the reductionistic attitude of Democritus' atomism. It states that there can be no real unities apart from the fundamental atoms that constitute all reality. It further maintains that when a whole is composed by a group of atoms, such a whole is nothing more than the mereological combination of these atoms. Taken as an *a priori* metaphysical principles it states something that is against common experience. From an experiential point of view, unity is before the intuition that the perceived unities are combinations of atoms. In real life, most of human activity aims at getting unities with properties that are different from their parts. Nevertheless Democritus' thesis has become a fundamental pillar of present day science and even of common thought. Although it has a certain degree of truth and usefulness, it must not be taken as an *a priori* ontological principle if it cannot cope with *all aspects* of empirical evidence.

If we accept Democritus' mereological ontological principle, there is no practical way of obtaining a subject (or a real unity as well) starting from the material world. It doesn't matter how smart the engineering effort: anything that is obtained from some objective materials won't be, in the end, anything more than the parts it is composed of. Any effort will be doomed by the presupposed incapability of reality to produce real unities. Yet, we are subjects and our very existence denies Democritus' principle.

---

supposed that the subject itself could be perceived, but that its objects are perceived as a unity.

It follows that building an artificial subject, entails the ability to locate unities in the world and the ability to bring them into existence.

1.2.2 What is a representation?

> *[…] our representation of things as they are given to us, does not conform to these things as they are in themselves*
> Immanuel Kant[18]

If we look at the world around us, we perceive it. We perceive it consciously. In a sense, speaking of unconscious perception is even misleading: it is an oxymoron. To say that an unconscious physical process corresponds to perception is a mere metaphor. Why should we define a physical process as a perceptive process if it is not correlated to a conscious subject? Could we say that a video camera perceives the real world? The camera is just a physical structure that is causally related to events that are normally the content of visual perceptions that we have as conscious human beings[19].

Yet we perceive the world and we represent it. If we accept the physicalistic framework, we should conclude that our biological brains have the property of being capable of representing the external world. This is an astonishing fact since there is not any other example of physical entity representing autonomously another physical entity.

For example, if we consider a rose that we perceive consciously (let's say that we can see and smell it), we must conclude that our brain is representing it. Yet, our brain does not own any of the properties of the rose, neither the colour nor the smell. As we will see in Chapter 0, there are no physical objects or properties that can sustain such a baffling capability. In the physical world, nothing points autonomously at anything different from itself. If something is a representation of something different from itself, it is because some conscious observer is attributing this role to it. If a road sign indicates I must stop, it is just because there is an agreement between conscious human beings to view the road sign as a symbol (or a representation) of the need to halt. The same is true for every symbol that man has created. As William Lyons put it «A particular

[18] (Kant 1958).

[19] The impossibility of defining perception without conscious subjects is only one case of a more general principle. Conscious subject are inherently constitutive of many known aspects of reality. The mind is always the conscious mind. This position has been advocated, among the others, by Franz Brentano and John Searle (Brentano 1874; Searle 1983, 1992).

process in my calculator only represents the number 5 in so far as it is linked electronically with the LCD display of a simple line drawing which conventionally is taken to represent the number 5 in Arabic notation for simple whole numbers»[20].

Representations are born with man. Every representation is the result of an *interpretation*: a process by which a subject *chooses* something different from the symbol as its meaning. «An interpretation expresses the will that an event had a precise meaning. It is the will that something means *something more* than itself[21]». For us, this is easy to do because we, as conscious beings, live so literally inside representations that we cannot even imagine a world devoid of them. Nevertheless, in a purely extensional world, representation is a paradox. How can an extension (that is an object) mean *something more* than itself? To represent means to possess an arrow pointing at something outside.

In this sense, the term 'representation' is practically a synonym of 'intentionality', or *aboutness*, that is the capability *to refer to*. It is difficult to grasp the extent to which the problem of representation is fundamental to our comprehension of the world: whenever we think of the world we make use of representation; whenever we experience the world we do it by means of representations. We can even say that our existence is practically unconceivable without representations. Also the external world is unconceivable if not by means of representations.

Representations are fundamental in many other fields. For example, without them it is impossible to define computations in a physical apparatus; information is a void concept. «Computations are a modification among representations»[22] and «there can be no computations without representations»[23]. There is not any way to recognise a physical apparatus devoted to information processing and computations without recurring to the fact that conscious human users can use it to get conscious representations. A microwave transforms material into, yet nobody considers it as an information-processing machine. A mechanical calculator is seen as an information-processing machine because human beings look at the positions of its gear as representations of numbers. The same can be said for PCs in general. They are just physical systems. Since their output is so heavily linked to *our* representations, we promote them to the rank of information-processing

[20] (Lyons, 1998), p. 60.

[21] (Severino 1990), p. 59.

[22] (Clark 1997).

[23] (Fodor, 1976), p. 34.

machine. Yet the real representations are in the mind of their users. They possess only a second-order capability of representations[24].

Up to now, intentionality seems to be primarily, originally, a real feature of human brains[25]. It is more correct to say that it is a real feature of conscious subjects. For there are unconscious living human brains that apparently do not possess any kind of intentionality. The link between human brains and intentionality is valid only in so far a human brain allows a conscious subject to emerge. There is no *a priori* reason why a subject could not be endorsed by a different physical structure. The fact that only human brains have been associated with conscious subjects is just evidence of inductive nature, which is an unsound basis for generalization.

The capability of having representations, real first-order intrinsic representations, is something that is apparently a distinctive characteristic of subjects. Such capability is something only they have that distinguishes them from objects. Hitherto no human artefact has been capable of having genuine representations. Yet reality seems to have this capability and the proof is that we, as conscious subjects, do possess intentionality and do represent the world.

1.3 Consciousness and science

> *Look at the neurons for as long as you like, and you still will not find phenomenal consciousness*
>
> Michael Tye[26]

If human brains are the only things capable of referring intentionally to the external world – albeit when associated with the existence of conscious subjects -, what physical structure is necessary for a conscious event to happen? Are there any scientific theories that can explain how consciousness arises from matter? The explanation of the emergence of consciousness has suffered from the same problems as the explanation of the existence of intentionality and representation. Science seems unable to explain these features of reality not because of insufficient data but because of metaphysical or categorial mistakes. As David Chalmers wrote:

---

[24] (Searle 1980; Searle 1983; Smith 1996).

[25] We accept here the thesis that the intentionality of language is derived from the intentionality of conscious subjects.

[26] (Tye 1996), p. xi.

> The impressive progress of the physical and cognitive sciences has not shed significant light on the question of how and why cognitive functioning is accompanied by conscious experience. The progress in the understanding of the mind has almost recently centred on the explanation of behaviour. This progress leaves the question of the conscious experience untouched.[27]

Formal arguments state that a subjective experience, as it is not a physical object, does not need to share the properties of physical objects, among which the property of occupying one spatio-temporal point. However, not all philosophers and scientists are ready to give up the physicality of subjective experience. Given the fact that there are no accepted laws connecting the realm of subjective conscious experience with that of objective physical events, many different and incoherent approaches have been adopted. The solution proposed to bridge the gap between the physical and the phenomenal domains range from their total identity to their anomalous relations, from various degrees of dependence or supervenience to their total independence[28].

The only kind of evidence we have of the existence of mental objects is subjective in nature. We would not know anything about the existence of mental objects, if we could not access them in the private perspective of our first-person subjective experience. In a pure extensional and physical world, there would be no reason to suppose that there should be strange objects like pain, phenomenological colours, moods, and so on[29]. During most of the XXth century the widespread was to try to eliminate consciousness as well as any kind of phenomenal entity. Eliminativism, identity theory, behaviourism and some kinds of functionalism aimed at the same goal: the complete elimination of consciousness from science. Their failure prepared the ground for an upsurge of interest towards scientific methods applied to the study of consciousness. As a result there was an explosion of theories trying to explain consciousness. These theories can be divided into a broad categorization based on their attitude towards the representation problem. Three groups can be outlined.

The first is the attempt to reduce everything to physical entities inside the skull. In other words, representations do not really represent anything in so far as they never really refer to anything outside the brain. The perceived

---

[27] (Chalmers, 1996b), p. 25.

[28] (Davidson 1980; Churchland 1989; McGinn 1989; Kim 1993).

[29] An infinite literature is concerned with the status that must be given to phenomenal entities (Galilei 1623; Descartes 1641; Locke 1690; Leibneiz 1714; Eddington 1929; Nagel 1974; Kripke 1980; McGinn 1989; Shoemaker 1990; Shoemaker 1994; Strawson 1994; Russell 1995; Chalmers 1996; Stubenberg 1998; Block 1999).

properties are due only to particular phenomena inside the brain. In a sense this is a Kantian position. Perceived objects are neural phenomena occurring internally while represented objects are external events noumenically unknowable. This approach is what scientists like Francis Crick and Christopher Koch are following looking for particular kind of oscillations in the brain. In short, they and others look into the brain to see if there is anything that can be the correlates of the brain owner's states of consciousness[30]. For example Francis Crick wrote, «It is difficult for many people to accept that what they see is a symbolic interpretation of the world […] in fact we do not have a direct knowledge of objects in the world. […] Our Astonishing Hypothesis says […] that it's all done by nerve cells[31]». Apart from the technical details, the general framework of this approach is similar to Paul and Patricia Churchland's neurophilosophy. There is no real access to the outside world. Everything we experience is just a neural feature. Yet, there is a logical problem. If what we experience is internal to our brain how do we know that there are brains? Not from our direct experience. For nobody sees his/her own brain directly. We perceive it as an external object. Therefore there is the risk of an infinite logical regress. Experimental results are far from being complete or generally accepted. For instance, there is still no consensus on what the real correlates of a consciousness state are. How many neurons should be activated in order to produce a conscious feeling[32]? How can they refer to something that is in the external environment? A related approach is given by the so-called representational theory that presupposes some innate representational medium in the brain[33]. According to it our brain states represent something because they have had this property from the very beginning. Yet representations are «really in the head». Theories belonging to this group are usually sophisticated versions of the identity theory[34]. However, they are internalist regarding where to locate the physical medium for representation.

---

[30] (Churchland 1985; Churchland 1989; Churchland 1990; Crick and Koch 1990).

[31] (Crick 1994), p 33.

[32] The idea that a large number of neurons is needed to have a conscious representation of something (an image for instance) has been recently challenged by experimental results that support the old idea of the grandmother cell (Kreiman, Koch et al. 2000). A limited number of neurons firing could be sufficient to activate a conscious state. Their number could be much smaller than the number prescribed by the traditional information theory.

[33] (Fodor 1987; Pulverlmuller 1999).

[34] (Armstrong 1968).

The second approach is focused on what the content of such a state is. Given the fact that the brain does not seem to be influent on the various properties of these states, external objects seem a logical alternative. According to this view, the content «isn't in the head». The most famous thought experiment was Putnam's Twin Earth case[35]. Imagine two people (John and twin-John), biologically identical, who live on two planets (our Earth and Twin Earth), which are identical in all respects – except one. On Twin Earth water is substituted by XYZ. XYZ is phenomenically identical with water but it is made of a different physical substance. As a result, where John has a belief about water, twin-John has a belief about XYZ. Even if John and twin-John are identical, their beliefs refer to different entities. Although Putnam has subsequently modified his view, this position has been represented by several exponents of the externalist mainstream, among which Drestke and Tye[36]. They try to define abstract conditions according to which the external information can be represented in the brain. They are often but not necessarily externalist. Another problem is that these theories do not say anything precise about what the appropriate brain structure should be in order to produce consciousness and they need to explain how meaning, that they locate outside the brain, can be part of the brain structure given a physicalistic ontology.

A third alternative is the functionalist point of view[37]. Functionalists look neither to the internal medium nor to the external target of a mental action, but are interested in the functional structure that deals with both of them. Here, the problem is that the typical functionalist structure is a pure abstract relational structure with no place for the qualitative meaning usually associated with experience. Besides, it has the so-called property of independence from real implementations. This property is the strength and the weakness of this position because it frees functionalism from the burden of materialism but lacks a proper (and physically acceptable) ontological domain. Furthermore, there are the problem of phenomenological quality and the problem of first-person perspective.

Of course, other taxonomies can be devised to divide the theories on the nature of the mental[38]. The one sketched above wants to stress the importance of locating the content whether inside or outside the brain. All these onslaughts on the citadel of consciousness show an absence of a clear understanding of the structure of an elementary act of consciousness, such as representation, and

[35] (Putnam 1975).

[36] (Dretske 1993; Dretske 1995; Tye 1996).

[37] (Putnam 1975; Dennett 1996).

[38] For example, an interesting survey is provided by (Block 1999) or by (Tye 1991).

therefore of the correlated properties of the part of reality involved in it. For instance, why should we suppose that a billion neurons should be better than just one in producing a conscious experience? Until psychophysical laws are not be set down, as far as we know, one single neuron could be sufficient to instantiate a conscious event. To date, there is not one single line in literature that constrains what should be the physical properties of a physical correlates of a conscious event. How many neurons are necessary for my feeling of redness to be produced? Science does not seem to provide an answer to two fundamental question that we will define as the *nature question* and the *representation question*. The first is «what is the nature of a physical event in order to be able to produce consciousness?» This question will later lead to the *second question* that is «how can an event refer to other events and carry their meanings?» Given the right perspective, we will try to show how these two questions have a common answer.

1.3.1 Two caveats: content and mental

A first caveat. We will use the word content under the following suppositions. Usually there are several alternatives to what is considered to be the content of mental states: intentional content, conceptual content, referential content, representational content, and phenomenal content[39]. Not all authors would agree on this taxonomy. Besides, if an intentional or a representational theory of content is accepted, the reference of a certain mental state can be seen as something different from its content. Here a different approach will be followed. Given the fact that the way in which the mind achieves all previous kinds of content is still largely unknown, the problem of content will be addressed in a rougher but more general way. In other words, content will be everything that constitutes the object of a mental conscious state. That is, if a mental state differs from another mental state in some respect, two are the possible *explanandum*. First, the difference can derive from a difference in the object (viz. the content) of the two mental states. Secondarily, the difference can originate from the modality or the way of accessing to the same object. In principle both options could be pursued. A first example is given by the dichotomy between Hume's ideas and Kant's categories[40], where the object approach is preferred and any difference between mental states will always imply a difference in their content. Besides, there will be no difference between

[39] (Kim 1998).

[40] For a more recent survey and a comparison between vehicle and process theory of representation see (O'Brien and Opie 1999).

representational and phenomenal content, or between representational and intentional content. This does not mean that a different way of accessing an object would not determine a difference in the approached object (for example hearing or seeing a barking dog is surely a different mental state because in the first case the content refers to the barking and in the second to the image of the animal). Following this point of view, *the sensory modality is given by the nature of the perceived object*.

Another caveat. Another caveat is how the word 'mental' and 'conscious' will be used in this thesis. As a general rule, 'mental' will mean 'conscious'. The Cartesian principle that everything that is present to mind must be present to consciousness is held true here. The notion of an unconscious mental state is a contradiction in terms. We are aware that this choice might be considered controversial but, after all, why should any process or event be called mental if it isn't followed, at a certain point, by a conscious event? For instance, unconscious processes are considered part of the mental domain because, in some way and some time, they will influence some conscious state. They are mental, not because of their intrinsic nature, but because they will modify true conscious mental states. It follows that, if a mental state or process never provokes any effect on the corresponding conscious subject, do we call it a 'mental state or process'? A brain cancer is not considered a mental process[41].

Let's consider the unconscious mind again. If a thought or a mental state were to have no effect on the conscious experiences of a subject, why should we call it a thought? Or why should we call that state a mental state? If there were no subjective conscious experiences, any event would remain a simple physical event devoid of meaning. Imagine a glass full of water. It can be seen as a simple physical event or it can be seen as an incredibly powerful computational device calculating the position and the speed of billions of H_2O molecules. Where is the difference? The same rationale can be used with brains. If we look at them from the point of view of physics, they are just an incredibly complex bunch of interacting neurons. Nevertheless their activity is usually defined as mental even if it is not directly linked to consciousness. Even in a brain, there are plenty of events and processes that nobody would call mental events or processes: the rising and falling of blood pressure, or the growing activity of several kinds of supporting cells. Why are such activities not considered as mental? The answer is that they do not have anything to do with consciousness. A possible drawback of this choice is that it goes against a venerable and long established tradition started at the beginning of this century with Freud's work

[41] A similar position was maintained by Franz Brentano and, more recently, by Joh Searle. According to Brentano, there are no unconscious psychical phenomena.

on the unconscious and indirectly sustained by the behaviourist mainstream. If consciousness is an epiphenomenon, it is clear that the mind must be defined in consciousness-independent terms. Notwithstanding the authority of this tradition, we claim that it is not possible to define the mark of mental without any reference to consciousness. We claim that the burden of the proof is on the shoulder of those who deny the identity between mind and consciousness.

Having explicitly stressed these two caveats we can now examine a series of problematic cases that arise from the application of traditional scientific theories on the phenomenal events of conscious subjects.

1.3.2 The brain is not made of chocolate

Let's imagine looking at the brain of someone who is looking at a green meadow, or licking the brain of someone who is eating a delicious chocolate ice cream[42]. In the first case we will not see anything green and in the second case we will not taste anything like chocolate. The example may seem trivial but it is not. If it were, why are neurologists so happy about having found out that there are some kinds of retinotopia going on in several parts of the brain? The fact that some contorted way of shape preservations exists, does not entail that there is any possible way to preserve an enormous list of properties, which do not have any conceivable way of being reproduced. If there is retinotopia it is only a consequence of some practical and contingent constraints on the location of nerve cells. Even if there is retinotopia, smell-topia or taste-topia are not plausible. Spatial relations can be easily reproduced by another physical object, phenomenal properties cannot. The relation between the conscious event and its content must be of a complete different nature. The brain is a physical object that does not have the properties that it represents. The brain is different from its represented content that must be somewhere else. The reason why subjects are conscious of the external objects and not of their brain activities is straightforward: because the properties of the external objects are different from the properties of brain objects (as long as science is right in telling us what a brain is). Therefore

brain events do not possess the same properties as external events.

[42] (Russell 1995).

1.3.3 Problem of distance and delay

Every conscious event must have content. Without such a content the conscious event would be void. It does not represent anything: therefore it could not be a conscious event. Although some authors maintain that certain mental states lack any content, we claim that their content is precisely what it is like to be in the particular mental state. Their content is precisely what makes a mental state recognisable to its owner[43].

Surprisingly, this content cannot exist at the same time in which we are having the experience of it. This is a trivial consequence of the speed limit of information transmission (and inasmuch of causation transmission). Let's think what happens when we see something. Somewhere in space there is an object that is reflecting light. After a while (a small time but nevertheless a finite one) our eyes receive the light and start complex chemical reactions inside their cones and their rods. These reactions trigger a series of causal effects that are relatively fast but slower than the almost instantaneous time requested by the light to reach our retina. Something like 400 ms later our brain finishes processing visual information in the visual cortex[44]. Whenever the conscious event occurs, it occurs later than the visual phenomena in front of the subject. Besides, it happens in a physically distinct spatial location. It is a distinct physical event. An extreme case of this is illustrated when we look at the stars, at the sun or at the moon. We look at the sun but an explosion could have destroyed it 8 minutes previously, yet we could still be conscious of its existence. A critique to this rationale is that we are conscious of the image on the retina and not of our star several billions of kilometres far. The answer is that, near as it might be, even the event on the retina is not the same as the conscious event and it is not, therefore, coincident either spatially or temporally. It is well known that the brain can be elicited directly and that it can produce visual conscious events without any stimuli on the retina. The conclusion is that the retina is neither sufficient nor necessary to provoke conscious events. From the point of view of consciousness, the retina is as distant as the farthest star. If we remove the constraint regarding the distance in space and time between a conscious event and its content, there are no more objections to the fact that when we are looking at something, we are conscious of that object and not of our retinal activity. When we look at a dog, for

[43] For example, John Searke claim that certain mental states – like depression – lack any content (Searle, 1983).

[44] (Kandel, Schwartz et al. 1991; Milner and Goodale 1995).

instance, we are conscious of that dog and not of the doggish shaped chemical activity on our retina. The conclusion is that

> *perceived physical events and corresponding brain events are temporally and spatially separated.*

1.3.4 Displaced brain

We might imagine connecting a brain to its body by means of a very complicated set of radio transmitters that are able to substitute the normal links between the brain and the body. There are no objections to this operation. As long as the connections are working there is no reason to suppose that this subject will feel any difference in her conscious experience. Her brain is working exactly in the same way as before, her body reactions are occurring in the same way and is peripheral sensorial organs are sending her all the necessary information. All the causal connections between her and the external world are preserved. At this point, imagine removing her brain from her skull and taking it somewhere else. In Daniel Dennett's version of this thought experiment, the brain was removed and located in a place different far that in which the sensing and acting body[45] was living. The conscious events happen somewhere, in a place that might be unknown to by the owner of the brain. She has no way, based on introspection, to know where her brain is. Where are the correlates of his conscious states physically located? The conclusion is that

> *conscious states do not tell anything about where they are physically located*

1.3.5 The brain is not the world

There is nothing in the brain that can be seen as the equivalent of what we experience everyday as the conscious content of our mind. Given the fact that neither any phenomenal property nor any objective knowledge exists in our brain, it seems mandatory to suppose the existence of a separate domain for the mental private entities (dualism) or a separate domain for the objective entities (Frege's third reign). The example of the swamp man comes to mind. Let's imagine an accidental replica, molecule by molecule, of a normal brain. This replica would be an object identical to the brain of a normal man, Let's say Smith. There will be a Smith and a swamp Smith. The two would be identical

[45] (Dennett 1978).

by definition so that everything that comes as a result of the former should exist also as a result of the latter. If the meaning (viz. the content of conscious states) supervenes locally on Smith's brain, it is not possible to avoid a paradox. That is, if the swamp brain is identical to the normal brain it must contain the meaning (the content) of everything that has been Smith's past experiences. If this is true, we must accept the fact that an object like our brain can contain the meaning of things, which it has never been exposed to. It seems like magic the idea that a piece of matter should contain the meaning of other physical objects without ever having been in relation with. If this conclusion is rejected the other horn of the dilemma must be chosen. We ought to accept that the swamp brain has no conscious experience even if it is a molecule-by-molecule replica of a normal brain. This is impossible because, given a physicalistic ontology, everything that is identical in physical terms must be identical in every respect (strong supervenience on the physical). Of course, there is a third option that entails rejecting physicalism. This option requires equating consciousness with its content that is to its representation (see Chapter 5 and followings). The conclusion is that

> *the content of mental states can neither be inside the head nor outside it.*

1.3.6 Breaking the wall

One of the most frequent yet vaguely defined concepts in the cognitive field is the distinction between internal events and external events. For example, is the chemical activity in the retina internal to the structure of the brain or not? The neural activity in the cortex? The memory of my computer? The light that is being emitted by an electrical bulb and that eventually becomes the content of my conscious visual perception? There is no physical distinction between events that are traditionally conceived as internal events and events conceived as external. The boundary represented by the skull is nothing more than an aesthetically-appealing container. There is no mental field to enter into, nor mental physical substance to cross. There is no 'pineal' threshold to pass. From a physical point of view, if we refute the existence of conscious states as something different from normal physical events, there is no reason to consider anything as internal or as external. In short

> *there is no objective threshold between an internal mental domain and an external one.*

Summary

The world is intuitively divided between subjects and objects. This partition lacks a clear understanding of what a subject is. We propose that a subject is a *unified set of representations*. This proposal highlights two unresolved problems: the problem of unity and the problem of representation. While we have empirical evidence of the existence of both, we usually accept an objective extensional ontology in which there is no space for either.

Being a subject is being a conscious subject. The link between consciousness and the capability of representing has been a long established. The problem of consciousness is beginning to be a serious scientific problem and its difficulties seem to arise not only from practical obstacles but also from theoretical barriers. A series of thought experiments are analysed to show the paradoxes that arise when our classic extensional categories are applied to our mental existence. Cognitive mind and phenomenal mind seem to be two distinct entities and what can explain the first cannot explain the second. Intelligence and consciousness belong to different conceptual domains. The first *does*, the second *is*.

It seems that there is no place for a conscious mind in an extensional objective third-person world. Yet our phenomenal subjective first-world experience is undeniable evidence and we must find its proper place.

2 The Aladdin lamp

> *How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue is just as unaccountable as the appearance of Djin when Aladdin rubbed his lamp.*
>
> Thomas H. Huxley[1]

During the XXth century, scientists and philosophers tried to explain consciousness starting from other entities. *We claim, on the contrary, that most of the entities, which are believed to exist independent of subjects, are meaningless without consciousness.*

As Thomas Huxley observed, more than one century ago, the existence of consciousness is «unaccountable». This conception derives from the idea that objectivist science should be capable of explaining all empirical facts except consciousness. As a result, most of the research on the mind has been conducted as if consciousness could be left in the sidelines until the last minute. If this were true, the world would be composed of entities that are independent of the existence of conscious subjects. Information, objects, meaning, observers, communications, and representations would exist autonomously. In other words, they would be pure objective entities. Yet, as we will show in this chapter, this is not the case. All the supposed-objective entities are dependent on the existence of conscious subjects. They cannot exist independently of their relations with conscious subjects. They have been proposed separately in the attempt to remove all subjective elements from reality. This effort has encountered serious difficulties in two separate but fundamental fields: the understanding of consciousness and the quantum mechanics description of reality. In both cases there are phenomena that cannot be suitably explained without referring to the existence of subjects. As Werner Heisenberg wrote «no description is possible between two separate observations[2]» and 'observation' entails some kind of 'conscious observation'. The traditionally accepted framework of objectivistic science appears insufficient to describe the complete spectrum of empirical facts that range from objective to subjective facts.

---

[1] (Huxley 1866).

[2] (Heisenberg 1958).

2.1 Reductionism

Why do the final goals of objective epistemology have to be intrinsically reductionistic[3] in its final goals? Science looks for an explanation of phenomena or better: to be able to reduce empirical descriptions of facts to objective explanations that require a smaller number of unexplained empirical facts. This method has led to the famous hierarchy of sciences that sees physical facts at the bottom and social or psychological studies at the top passing through chemistry, biology, anatomy and psychophysics. The objective method compels us to be reductionistic because it based on so-called objective propositions. An objective proposition is composed of two kinds of terms: names and predicates. They ought to correspond, more or less, to objective entities, which have been described elsewhere. This correspondence has to be certified by the scientific community through the commonly accepted scientific protocols. The whole process is of epistemic nature and it is therefore highly controversial to know if these objective entities are real or are just useful concept. Realism and strumentalism are at loggerhead. The correspondence would not exist without the existence of the community of scientists. In order to explain an empirical fact objectively, the only possible move is to substitute that event with other entities accepted by the scientific community. To give an explanation of something entails being capable of substituting an entity with one or more other objective entities in such a way that this substitution will be *salva veritate* (we can term this operation the *reductionistic move*). This method is epistemically proficient only if it allows us to use an increasingly smaller number of objective entities. The final goal is to substitute every objective entity (like *gold*, *table* or *Drusus*) with the same objective entity (something like *mass* or *energy*). Ideally, we could rewrite every possible sentence about an objective fact by using a sentence (it does not matter how complicated or long) made up of only a combination of basic objective entities. Even if this were possible, it inescapably follows that at the end of the process there will remain at least one objective entity that would be impossible to explain. Science is driven by its own structure to a progressive reduction of the epistemically needed objects (a kind of Ockam's razor), so there is an unjustified (but universally accepted) belief that this epistemic criterion is also an ontological

[3] Here reductionism must be seen as ontological reductionism. That is, the attempt to reduce the entities of the world to the same stuff. It is different from epistemological reductionism that is the attempt to reduce all our assertions about the world to the same kind of assertions (objective, protocols, by acquaintance, etc.).

criterion. In other words, the more an objective entity can be used to substitute other objective entities, the more that entity will be real. For example, you can substitute the names of *Atos*, *Portos* and *Aramis* for the *three musketeers* in every possible objective sentence[4] and, for a natural feeling of epistemic gratitude; we feel that they are more real than their trio. After all, they exist by themselves while the trio, as a separate objective entity needs them to exist. How can we affirm this? Because there are more objective sentences that contain *Atos*, *Portos* and *Aramis* than sentences that contain *three musketeers*. Every possible sentence in which the name of the trio occurs can be substituted by the same sentence with their three singular names *salva veritate*. There are, of course, sentences like «The three musketters is the name of a famous trio» in which the substitution cannot occur. Yet they must not be taken in consideration since they use the name 'three musketeers' to refer to the name as such and not to what it represents in normal usage.

In the same way, the terms of objective science can be used to substitute other empirical facts in an enormous quantity of empirical sentences. Instead of speaking of *rivers*, *rocks*, *planet*s, and all empirical entities, it is possible to utter sentences that speak only of *mass*, *space*, *time*, and *energy*. Chemical processes collapsed to a purely physical description thanks to quantum mechanics. Even biological beings seem to be reducible to chemical reactions among extremely complex organic substances. Because every description of the physical world can be substituted by a description that speak only of *mass*, *space*, *time*, and *energy*, we feel that these are the only physically acceptable real things. Is it true? Is the *reductionist move* a compelling step?

Apart from the logical confusion between ontology and epistemology implicit in the reductionist move, we shall argue that there two reasons to reject it as an absolute principle.

The first reason is that there are many sentences that do not seem easily reducible to objective sentences. We can divide them in two groups: intentional sentences and phenomenal sentences. Intentional sentences are those that contain intentional term like beliefs. It is not clear at all if a sentence like «Heatcliff believes that Catherine is his own life-blood» will ever be reducible to objective entities. Phenomenal sentences are those that refer to phenomenal

[4] By "objective sentence" here we mean extensional sentences in which terms with the same meaning can be substituted *salva significatione* as well as *salva veritate*. We are explicitly avoiding all problems related to phenomena of semantical opacity and *oratio obliqua*. This avoidance is not casual because it is clear that these phenomena are closely related with the mental existence of the owners of beliefs. It is difficult to explain belief without using some kind of conscious beings.

entities like pain, smells, perceived colours, thoughts. Although many authors denied the existence of private mental entities, many others claimed that they are real. How could they be explained in objective and physical terms?

The second reason stems from a *reductio ad absurdum* of physicalistic reductionism itself. Physicalistic reductionism claims that everything can be reduced to a few physical fundamental entities. Consciousness and mental entities are at the top of a hierarchy that must collapse on the fundamental level. Consciousness can be seen as an empirical phenomenon for which it should be possible to give an objective explanation. To do this, we should reduce consciousness to other objective entities and avoid using terms that require consciousness. Our claim is that

> *all possible objective entities, which can be used to reduce consciousness, require the notion of consciousness to be explained.*

In other words, the existence of conscious subjects lies at the bottom of the fundamental objective entities. According to this rationale there would be no pure objective entities. All events would be a mixture of subjective and objective and consciousness would be an irreducible aspect of reality. In this chapter we take into consideration four traditional candidates for an objective reduction of consciousness: material objects, information, objective meaning (a kind of Frege's meaning) and dynamic systems like the brain. We will try to show that each one of these entities requires consciousness in order to be meaningful.

2.2 The problem of objects

> *Myth of physical objects, […] posits comparable, epistemologically, to the gods of Homer*
>
> W. V. O. Quine[5]

What is more self-evident than the existence of objects? Than the everyday objects of our lives: chairs, tables, computers screens, and keyboards? Given the sceptical point of view, it is true that we could doubt of everything but, leaving aside such extreme position, we ought to concede that objects exist. Let's analyse their autonomy in relation with consciousness. Being autonomous from

[5] (Quine 1951).

consciousness entails being autonomous from conscious observers[6]. If this last sentence were not true it would mean that 'being an object' is what is normally called a secondary property of matter. Traditionally, all properties of entities are divided into two broad classes: primary properties and secondary properties. With these two terms, we simply refer to the following kind of properties: the first are the properties that we can assume entities would have without the existence of conscious observers (like mass or electric charge) while the second are dependent on the existence and characteristics of conscious observers (like being warm or fresh, or being ugly or beautiful, or being Riccardo's favourite colour). A hidden assumption, which is here denied, is that the existence of an object must be considered as belonging to the class of primary properties. As an «intuition pump» we propose the following example. For those interested in a more precise argument we refer to the Appendix 11.1.

Let's think of three crosses (Figure 2-1). Consider the first one: a normal grey cross on a white sheet. That cross can be seen as a real physical object. Would it exist without your observation? Now consider the second cross: it is a matrix of numbers. After the first look, you should be able to see that there are a row and a column of ones against a background of zeros. It is unquestionably a cross, albeit a cross made of numbers like the previous one was made of grey patches. Now look at the third cross. Where is the cross? Look carefully. If you are not fond of mathematics you may not notice that there is a row and a column of prime numbers. However, if you were a skilled mathematician, you would see it immediately and without effort. The existence of these crosses depends on your existence and on your ability to observe them. If you were not a mathematician, you could see only two of them; if you were not able to read you could see just one; and, finally, if you were blind there would not be any cross at all. As a separate part of reality, each cross comes into being when becomes the object of our experience. It would even be possible to have a cross emerging from a purely random matrix of casual numbers. It would be enough to imagine the existence of an observer that would give a particular meaning to the numbers located on the internal row and on the internal column. Being an object is something that depends not only on the physical properties of the thing in itself but also on the properties of its conscious observers. We might

[6] From now on, we will use the term 'observer' as a synonym for conscious observer, because we think that it is not possible to define an unconscious observer meaningfully. An unconscious observer is a contradiction. An unconscious observer is simply a physical phenomenon that is causally connected with another physical phenomenon. This is too broad a definition to be useful. If we were to follow such a definition, we could define a puddle as an observer of the weather.

ask what would remain if we could remove all secondary properties? A sceptic would object that nothing would remain, and that this would be a manifestly absurd conclusion. We answer this objection by saying that this conclusion is not mandatory, because the fact that objects are dependent on consciousness is not the same as the fact that it is possible to reduce objects to conscious subjects. Although conscious observers are not a sufficient condition for the existence of objects, they are nevertheless a necessary condition.

In other words, a physical object is not only a set of physical particles but also an interpretation[7]. If we had photoreceptors, we would select certain classes of objects; if we had ultrasound sensors we would select a different class of objects from the physical continuum. Even with the same kind of perceptive capabilities, we could always make different semantic choices. Think of constellations of stars. We can choose to connect one star to a constellation or to remove it. There are no fixed rules based on their magnitude or position and the historical choices are what they are: we are fond only of conventional choices, for sentimental reasons (Figure 2-2).

Every object is like a constellation of stars. Nelson Goodman wrote: «Has a constellation been there as long as the stars that compose it, or did it come into being only when selected and designated? In the latter case, the constellation was created by [us] … a constellation becomes such only through being chosen from among all configurations … As we thus make constellations by picking out and putting together certain stars rather than others, so we make stars by drawing certain boundaries rather than others. Nothing dictates whether the skies shall be marked off into constellations or other objects»[8]. Similarly William James stated that «'Wholes' are not realities, parts only are realities. […] The 'whole', be it a bird or constellation, is nothing but our vision, nothing but an effect on our sensorium when a lot of things act on it together. It is not realized by any organ or any star, or experienced apart from the consciousness of an onlooker»[9].

It follows that, without any observers, macrophysical objects do not exist as wholes. Observers require consciousness. We go onto argue that in order to

[7] It is a widespread idea that physical objects can be seen as a simplifying conjecture that is produced to cope with the multitude of sensory information. But what is a conjecture or an interpretation then? Objects result from a cut that the subject executes on the physical continuum using its own internal meanings.

[8] (Goodman 1978), p. 36.

[9](James 1908), p. 194. A similar consideration can be found in (James and Kuklick 1981). A more recent development of the same strem of thought is represented by (Smith 1996; Smith 1998).

have an world made of the familiar objects, we need a world of subjective conscious observers. Objects are a mixture of objective and subjective ontology.

> *The existence of an object as a whole depends on the physical properties of observers as well as on their semantic choices.*

0	*0*	*1*	*0*	*0*
0	*0*	*1*	*0*	*0*
1	*1*	*1*	*1*	*1*
0	*0*	*1*	*0*	*0*
0	*0*	*1*	*0*	*0*

30	*8*	*17*	*25*	*28*
18	*15*	*11*	*9*	*14*
31	*3*	*5*	*23*	*29*
10	*22*	*19*	*4*	*20*
16	*8*	*7*	*6*	*12*

Figure 2-1 Three crosses: do they exist in the same way?

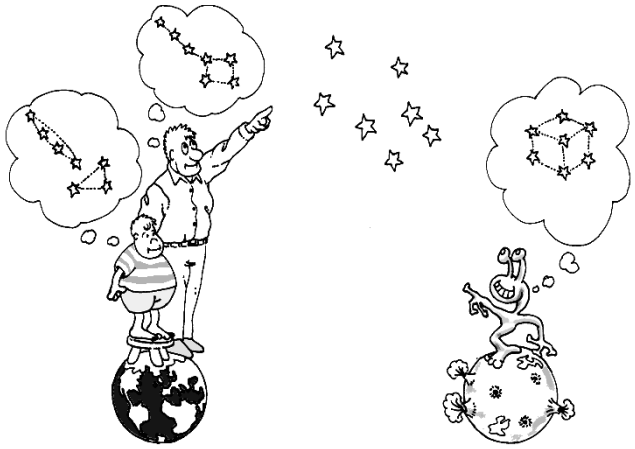


Figure 2-2 Three different conscious observers look at the same seven stars (*Dubhe*, *Merak*, *Phecda*, *Megrez*, *Alioth*, *Mizar*, *Alkaid that compose the Ursa Major*). Due to the different physical properties, one observer sees a different constellation. Yet the two observers on the earth, which are physically identical, make different semantic choices and see different groups of stars.

2.3 The problem of meaning

Ceci n'est pas une pipe

René Magritte

If objects also depend on an observer's semantic choices, what is that determine them? A possible answer is that they are driven by their internal meanings. Here, it is impossible to draw a detailed history of the term 'meaning'. Yet it is important to underline that meaning is a complex function of two necessary aspects: a subjective aspect and an objective aspect. We cannot split reality in any one of them. Reality requires both of them. The conundrum we have to deal with is: «how is it possible that meaning is developed by an unconscious material system?» Let's imagine that the world is completely devoid of conscious beings. Is it still possible to conceive that there anything similar to meaning would remain? We claim that in a purely extensional world meaning would be unconceivable. The very existence of meaning depends on the power of something to refer to something else. If this power, usually a burden for subjects, is lost, no meaning is possible. Yet, there have been various attempts to define meaning as something autonomous in relation to subjects. For example, meaning has been seen i) as something that is related to external objects in themselves or ii) as something that is capable of producing relationships with external objects by itself. Let's analyse these two points of view briefly.

Firstly, meaning has been defined as a class of objects[10]. In the absence of conscious subjects the meaning of an external object cannot be anything else than the object itself. For example, the meaning of a name or a description is derived from the object and it can be intentionally effective only as a result of a subject's interpretation. Unfortunately, as we have seen in the previous paragraph, if objects are selected on the basis of the semantic choices of subjects, objects cannot be used as an autonomous foundation for meaning. In other words, if objects are secondary properties of matter, they depends on subjects and they cannot be used as a foundation for meaning. For example, consider a group of six atoms on a circumference at regular distance (Figure 2-3). They represent the vertexes of a hexagon, or the vertexes of a King David star, or two triangles. It is the observer who is grouping the physical continuum in order to produce particular classes of objects. If meaning is the object in itself

[10] In this case meaning is to be intended like Frege's *bedeutung*, i.e. the real reference of a thought in one's own mind.

why should we accept one dot in place of another one? It seems extremely difficult to come by a definition of meaning that will survive the elimination of consciousness without being extended to the whole physical continuum.

Secondly, meaning has been defined as something that is capable of indicating its own object in the world but that is different from such object[11]. Concepts, names, natural kinds and intension are all examples of this point of view. They do not belong to the extensional world. They are not physical objects. For different reasons, all these entities have been considered capable of carrying the burden of meaning and capable of referring to the external world. Traditionally an intension is defined as a function $f:D \rightarrow R$[12]. That is a function mapping from possible worlds to physical referents. For example, the intension of water would pick up H_2O in this world and XYZ in XYZ worlds. It is not completely clear if an intension is a mental object (and as such a consciousness related object) or a public, inter-subjective, practical relation. It has been assumed that intensions live in a kind of platonic objective world and do not depend on actual minds to be real. Examples of this point of view are Popper's world 3 or Frege's *sinn* (Figure 2-4). There is no need to emphasise the difficulties of such a position: the existence of the proposed domain is ontologically expensive and empirically groundless; besides it entails all sort of problems of interaction similar to the paradoxes of dualism[13]. Alternatively, intension should be something that is part of our mental structure and that is capable of picking out an object in the external world. Yet this entails that the mind must have a real existence of some kind. If this is denied intension and extension become pure logical entities as in the case of neo-positivism (Figure 2-5). In other words, we can assume that there are structures that carry the meaning of the external worlds into our mind. In order to define such entities we must previously define what the mind is. The absence of a theory of mind

[11] For example, we can think of Frege's meaning or *sinn*. For him, *sinn* was something completely objective and independent of mental representations. Of course his platonic idealism was more a faith than a logical conclusion. What ontological support can such a meaning have?

[12] There are of course more recent and better working definitions for intension. In particular we refer to Kaplan's division between content and character, Block's division between narrow and wide content and Chalmers' division between primary and secondary intension (Chalmers 1996). However, for the scope of this thesis it is enough to mention the general concept of intension.

[13] It is not by chance that Frege defined the having of a thought or the grasping of a meaning as the «most mysterious thing of all». If reality is split in more than one domain all kind of problems derive from the interactions among the various domains.

has surprisingly split the structure of meaning and the structure of what the meaning is referring to (Figure 2-6). Since the meaning has to refer to something external while at the same time should occupy a representational internal role, many researchers proposed the introduction of a twofold structure. A sort of parallel evolution modified what was originally thought as the external reference. This evolution, while logically appealing, must still explain what the ontological references to the various levels are.

Another example should help to clarify these various positions. let's imagine a familiar object in everyday life: a thermometer. We recognize it since we own the related meaning. As in the case of the hexagon or the King David Star, we are capable of selecting the objects corresponding to the meaning of 'thermometer' since we own the corresponding meaning. Does this mean that we own the intension of 'thermometer'? We will show that there is no autonomous meaning corresponding to 'thermometer' and that this concept depends on the properties of conscious subjects. Why do we not consider other objects (knives, oranges) as thermometers? A naïve answer would be that the real thermometer is measuring the physical phenomenon we call 'the room temperature of the room' while knives are not. This is only a half-truth because also knives modify their physical properties (for example their length) according to the temperature of the room, too. There is a causal relation between the state of the knife and that of the room as well as between a thermometer and the same room. The real difference is that, in the first case, human beings can easily read the mercuric level while, in the second case, they cannot perceive the length variation of a knife. If there were no human beings, that familiar object which we have agreed to call thermometer, would not be a thermometer any more than any other object or atom in the same room. Applying the same kind of rationale to other meaning, it seems that not only is the existence of objects a secondary property, but also their meaning.

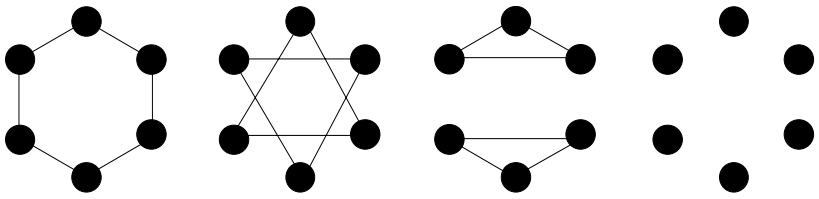


Figure 2-3 Six black dots. They have been repeated four times. Each time a different possible meaning has been shown: an hexagon, King David's Star, two triangles, and six isolated dots.

We have now rejected the idea that meaning is something directly related to objects. Let's go back to meaning as a function capable of picking out the appropriate objects in different worlds – that is meaning as a relational property. If we assume that meaning is something that is independent of minds we are trapped in a series of paradoxes. The main problem is that there is nothing, in an extensional world, that can be used as a candidate for the relation in itself. Saying that 'intensions do not require actual conscious minds' is the same as saying that 'one extension can be the intension of another extension'. If there were nothing more than physical facts, it should mean that there is nothing more than the physical particles that constitute objects. So, if my mind (which after all should not be anything more than a large collection of extensions, or objects or atoms) was intensionally picking out something from the external world (Let's say the thermometer), it would mean that one extension (that constitutes my mind) is picking out another extension (the object called thermometer). The problem is that this relation is neither a physical object nor an objective entity. So it does not exist. We can say that the object A represents the object B but the relation is a pure mental object. No one could objectively measure anything going on between those two objects. We could summarize this long argument by using the simple slogan:

no intension from extensions alone

This last sentence can be seen as the ontological puzzle of meaning. There is also an epistemological puzzle of meaning. Our mind states are manifestly able to refer to something external. Why was it possible to deny for a long time the existence of meaning as something different from external objects? Why is it so difficult for us to distinguish between external objects and their conscious meaning? The reason is that we cannot perceive anything without perceiving its meaning. We cannot perceive the world except by making the external extensional entities part of our conscious experience. How can we bring the meaning of the external objects to the subjective representation of it? Apparently, we are confronted with two different classes of objects: the internal carriers of meaning (or qualia); the external sources of meaning (objects); and, in the middle, the information carriers that pass through sensations. How can 'meaning' pass through information channels? Externalists claim that the tracking (this is their word) is due to an evolutionary, or intentional, or teleological link between the external object and the internal representation of it. Functionalists claim that meaning lies in the functional relation causally linking behaviour to states of facts. Here, we make a different, more radical claim that will be developed later. If every object were uniquely, constantly and

perfectly representing itself, how would be it possible that other objects (our own mental states) have anything to do with it?

Let's suppose that the mass of an object has to be measured by a group of scientists. They would have no problem in finding out the value of its mass if they could physically come in contact with that object. The same would be true for its total charge or dimensions, or its number of protons: all physical properties. Now, let's imagine the same group of scientists receiving a strange plastic badge from an unknown alien civilization in the outer space. They do not know anything of that civilization and yet they try to understand the amount of information stored on that piece of plastic. Neither do they know what storage devices were used nor what were the physical features of their alien owners. As a result, they cannot make any useful hypothesis. Can they determine the meaning of that piece of matter? Can they determine the total amount of information on the piece of plastic? They cannot because it depends on the properties of its users as well as on the properties of the devices they used. While the mass or the charge supervene exclusively on the piece of plastic the same does not hold for the information it carries – *a fortiori* for the involved meaning.

Another example is as follows. A group of scouts is in the forest. They find a broken branch of a tree. Is that a sign of something? Can they determine whether someone has purposively broken that branch in order to leave a sign? The branch could have been broken in such a way as to be physically identical to a naturally broken one. There could have been a meaning in that too. Or, as far as they know, the branch could not have been broken and *that* could have been a sign. Two scouts, the previous day, could have made an oath whose fulfilment would be marked by the rupture of that branch. That branch is still intact and thus the oath has been broken. Could the incoming group of scouts examine the branch and understand if there was any meaning linked to it? They cannot because meaning, quite obviously, does not physically change the objects with which it is associated to.

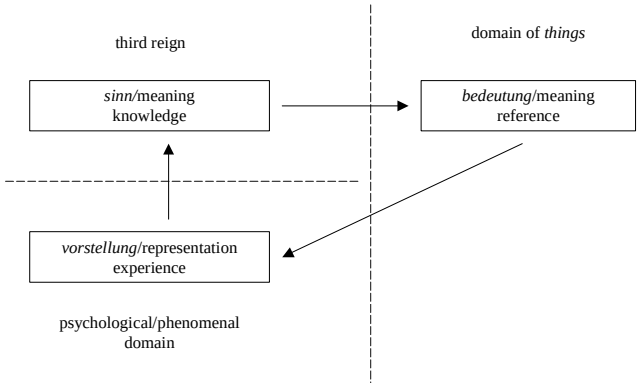


Figure 2-4 Frege suggested to split reality into three separate domains: objective epistemic entities (*sinn*), phenomenal of psychological entities (*vorstellung*), extensional entities (*bedeutung*).

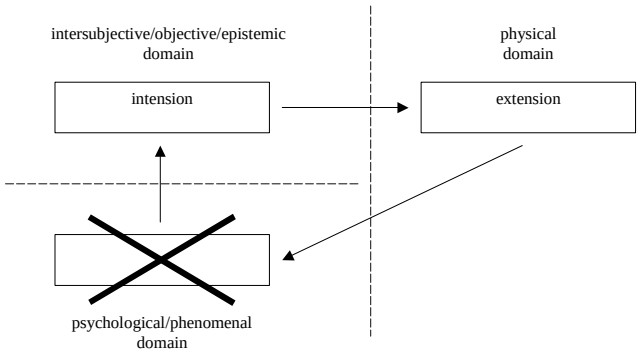


Figure 2-5 Neo-positivism and neo-empirism simplified the Frege's view. Intension lost any relation with the psychological domain; the ontological burden was carried exclusively by the physical world.

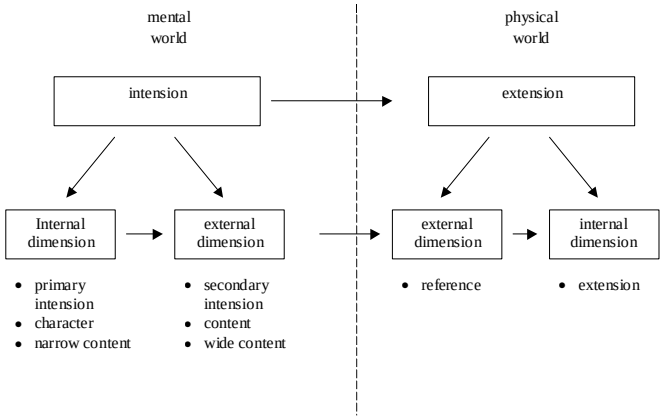


Figure 2-6 At the end of XXth century the framework of intension/extension entered into a crisis. As a result, there has been various proposal of doubling both the intensional side and the extensional side.

2.4 The problem of information

At the end of the previous paragraph we mentioned the difference between meaning and information. Information is apparently related to the causal interactions among events while meaning is completely dependent on the existence of objects. In other words, information seems a better candidate for a completely objective description of reality. There is also the widespread conviction that human beings are conscious because they are information-processing beings. The main reason for this broadly accepted idea is the abused software-hardware metaphor for mind-brain and the ubiquity of computers[14]. This point of view entails that being an information-processing device has to be independent of conscious beings and, consequently, that information exists by itself[15]. Yet, several authors have challenged this rationale[16]. We make the same claim. *Information cannot be objectively defined, therefore it cannot be a fundamental element of reality*. An example will help on clarifying our point of view.

[14] (Block 1995).

[15] A standpoint according to which information can be defined autonomously has been developed by Fred Drestke (Dretske 1981).

[16] Most notably John Searle carried on a series of attacks to the idea that information is an autonomous and objective entity (Searle 1980; Searle 1992).

In front of ourselves, we can see a 3½-inch floppy disk. What is the amount of information it contains? The first answer we can imagine is the usual capacity reported on its cover: 1.44 Mbytes (equivalent to approximately one million and five hundred thousand characters). It is an objective, acceptable, useful indication of its information capacity, so where is the problem? The truth is that those 1.44 Mbytes do not represent the disk capacity in itself. On the contrary they represent a complex causal chain among objects (composed by human beings, computers, drives and disks) that is capable of exchanging that quantity of information by using the disk. A practical demonstration of this is the following. At the beginning of the '80s, there were 720 Kbytes disks. At that time, if somebody had asked how much information could be stored on them, the answer would have been: «720 Kbytes each!» Nevertheless, towards the end of '80s, manufacturers of floppy disk drives succeeded in building smarter drives that were able to make a better use of the existing disks. In this way they brought their capacity to 1.44 Mbytes. Was there any physical difference in disks? No, they were exactly the same and the proof is that it was possible to use the old disks with the new drives to store an increased capacity. Without any direct physical action on them, they changed their capacity in the blink of an eye! The information contained in each disk was not dependent on the physical properties of the disks but also on that of its users.

Another example is the following. Let's imagine someone who does not like computers and that, because of his/her antipathy, uses the drives only to write on them short messages of twenty characters at most like: «see you at 5 p.m.» or «coming soon». What would the capacity of floppy disks be in a world where all people were like this person? The answer is just a few bytes corresponding to the number of characters that is possible to write on the disks.

The information capacity is not an autonomous property of an object (like mass, electric charge or size) but it is something dependent on a complex causal chain. One single object is not enough to sustain the burden of information. Yet, it could be claimed that a complex system of interacting objects could be enough. We will argue that even in that case it is impossible to define information independently from conscious subjects.

The standard theory of information is based on the work of Claude Shannon and Warren Weaver[17]. According to them, information is defined as a sequence of different states. Information is not, of course, a physical object, but it can be represented and transmitted by a physical support (communication channels and computer storage memory are a valid example). If we are able to distinguish two separate states of a particular physical system, we can use them to represent

[17] We refer to (Shannon 1948; Shannon and Weaver 1949).

the simplest unit of information conceivable (0,1): the famous *bit*. Starting from these elementary premises, among the others, Shannon and Weaver developed their theory of communication in which they distinguished between meaning and information. They were clear in refusing to be concerned with the semantics aspects of communication. Weaver in his introduction pointed out that meaning and information were two distinct aspects. He claimed that «two messages, one of which is heavily loaded with meaning and the other of which is pure nonsense, can be exactly equivalent, from the present viewpoint, as regards information»[18]. In other words there is no precise connection between meaning and information. As we have mentioned before, there was a widespread tendency to split reality in two halves, one related to the subjective domain and the other related to the objective domain. Meaning and information is another example.

Perhaps Shannon and Weaver have been over-criticized for not having been greatly concerned with the notion of meaning, since their theory was mainly focused on the process of communication. Probably the mistake was to take their working model as an ontological model. In other words, they were implicitly supposing that there was an observer associating meaning to some state of matter somewhere in the chain of communication. Without conscious beings, it would be more precise not to speak of communication (which requires information which requires consciousness) but of *interaction*. For example, let's imagine a set of gears whose state is mutually linked. We are not saying that they communicate between themselves but that they *interact*. We do not say that the gear communicates its state to the engine: we say that they somehow *interact*. We say that there is a communication only when at both ends of the process there are (implicitly or explicitly) two conscious beings. The main reason is that communication requires information and information, in turn, requires meaning.

Shannon and Weaver were interested in the physical properties necessary to improve the establishment of a causal chain between two subjects. *The only way to distinguish between a physical event carrying information and a physical event that does not carry anything is to know if any conscious subjects is attributing some meaning to that event*. The consequence is that

meaning and information cannot be divided.

If we rejected this principle, there would be no useful limit to what information can be. Every physical event could carry information. A glass of

[18] (Shannon and Weaver 1949).

water would contain billions of terabytes of information about the state of its molecules, their position, their speed, their spin and so on. Yet all those physical events are not considered 'information' since no one is able to associate them any meanings. Physical events become informational events when they are the targets of conscious subjects' semantic choices. Computers are information-processing devices only because in front of their screens there are conscious beings that are associating meanings with their states.

2.5 The brain

> *It was an intriguing and exciting idea that mental events could just* be *brain processes.*
>
> Jaegwon Kim[19]

Another classical supporter of consciousness is the brain. Many scientists and many philosophers believe that biological human brains possess some kind of special features that enable them to produce consciousness[20]. Here we do not take in consideration the possibility that consciousness is the result of a completely unknown physical phenomenon: something like a fifth fundamental natural force. In that case, the actual attempts to find an explanation of consciousness should be radically modified. Hitherto, no special phenomena have been discovered inside the brain. Chemical reactions and physical laws seem to go on exactly in the same way inside our skulls as in the rest of the universe. So we make the explicit hypothesis that the brain is made of the same stuff as the rest of the physical world. We believe that, at this point, there is no evidence of any special feature of the brain. It must be considered as a system, which is unique only because of its internal organization and not because of some strange new phenomenon – let's say a 'consciousness field'.

What is a brain? Can it be defined autonomously and independently because of its capability of producing consciousness? From a physical point of view it is an object, transparent to causal chains and processing information just as every other set of atoms. In this sense, it could not escape from the same problems we have mentioned for objects. As a whole, it does not exist by itself. Brains exist because conscious subjects isolate them from the physical continuum. They

[19] (Kim 1998), p. 2.

[20] Among the proposals made by scientists, the two most famous hypotheses are those of Roger Penrose's microtubula (Penrose 1994) and Francis Crick's oscillation (Crick 1994); among the philosophers John Searle (Searle 1992).

give the meaning of 'brain' to the set of atoms that constitutes it but – as happens with all other objects – the brain, as a whole, would not exist without a conscious being's active selection.

Because of the previous rationale, a series of related concepts loose their autonomy. An important example is the concept of 'state of the brain'. If the brain does not exist by itself as a whole, its state cannot be more real. It would be like considering the state of a constellation of stars. The state of such a constellation would be dependent on the semantic choices of conscious subjects. Traditionally it has been accepted that, given a particular brain, we can observe a particular set of states that are relevant for the existence of conscious experience. Let's imagine being able to define such a state. Here we challenge that such a state is autonomous. What events should be included in its description? In the brain, not all physical events are relevant to our states of consciousness. For example, the pressure of the rachidian fluids, within a reasonable range, can vary without effecting what we feel, what we perceive, and what we think. This ambiguity is fatal to many attempts to use the brain and its states as a support for consciousness. For example, Gerard Edelmann tries to define a phenomenal state as a vector in the space defined by all possible states of neural units belonging to a brain[21]. The problem is that such space is an abstract structure built upon an abstract object as it is possible to see from the following.

We could define S as the class of all the physical events that describe our brain as a physical object (conscious events and unconscious ones)[22]. Inside S we could define $S^\star$ as the class of all the physical events relevant for consciousness: of course $S^\star \subset S$. For example, we could imagine that $S^\star$ is constituted by all the neuronal axon-spike frequencies plus their phases, or maybe by the quantum state of microtubules, or whatever. All physical events, which belong to S but not to $S^\star$, would have only an indirect relation with consciousness. An example of physical event belonging to $S^\star$ is the pressure of the rachidian fluids. It is an event internal to the brain but unrelated to consciousness. $S^\star$ is the locus of consciousness.

A first paradox is that if we were able to reproduce the state $S^\star$ of a human being in a particular period of time $[t_0, t_1]$ we might suppose that the particular conscious state associated with that human being during that period could also be reproducible, again and again, for a virtually unlimited number of times. If we could slow down the speed of the sequence of state $S^\star$ we would be able to

[21] (Edelman and Tononi 2000).

[22] Due to the Indeterminacy Principle it is true that we would have problems measuring S, but for us it would be sufficient that S exists.

slow down the time of the conscious experience (not for itself but for a normal observer). This seems a little bit counterintuitive but we could live with it. It may be more difficult to accept the fact that we could imagine slowing down the repetition until the brain virtually stops and still we could say that that brain is still in the eternal experience of the same instant. Objections to this conclusion are possible. For example that the events defined in S* are of such a nature that they require a particular time scale.

There are even more radical problems. S* must be defined independently from the subjective experience of consciousness. Yet, how can we decide if a physical event is part of S* unless a reference to some subjective report of being conscious is used? We should have an objective criterion to distinguish between conscious events and unconscious events. To date, such an event has not been proposed by anyone. It is extremely difficult to define what is the particular nature of the events contained in S* that makes them suitable for consciousness. Let's now suppose that it is possible to record such events and that we can reproduce them using a different kind of physical phenomena, which possess, as a whole, the same structure. Would that structure possess consciousness? It seems really a tough bullet to bite.

Let's suppose that we could reproduce a version of a brain that, with a given interpretation, is isomorphic to the structure of the brain[23]. For example, the internal structure of a computer program simulating a calculator is isomorphic to the structure of a mechanical calculator. The problem is that we need an interpretation in order to identify the correct level where look for the isomorphism between the two structures. In turn an interpretation requires meaning so it is subject to the same problems outlined in § 0: if we looked at the level of pistons, wheels and nails of the mechanical version of a calculator,we will not find any kind of isomorphism within the program. However, if we look at the level of registry, addendum and operator signs we will find an almost complete isomorphism. The problem is how to find the correct interpretation to associate with the physical system. Every computer user knows where the correct level is. The problem is that there is no brain user. The mind is its own user so it cannot produce itself. Given the right interpretation, we can claim that almost any system can be seen as an information processing system. After all, each system is processing its next state at each state: a river – or a cloud – is constantly computing its shape.

[23] When are two structures isomorphic? When the correspondence between the relations of their internal parts is complete and when a causal effect in one of the two structures provokes the same kind of modifications in the other.

Let's return to the first problem. We imagine having a physical system $S^{\star} \subset S$ composed of all the physical events necessary to consciousness. Who can give us the appropriate interpretation in order to select $S^{\star}$ from S? Why should neurons be so special as to be the recipient of consciousness?[24] Let's imagine that another system that is identical to the first one expect for just one neuron. This system, if real, should surely be a conscious system almost identical to the original one. We can call that system $S^{\star}(x)$ where x represents the cardinal number of the removed neuron. Each neuron can be identified with a cardinal number going from 1 to 10^{11}. For every conscious system $S^{\star}$, we could imagine at least 10^{11} conscious subsystems $S^{\star}(x)$. We would conclude that, if my conscious states are not constrained by some different principle, there should exist 10^{11} possible versions of me at the same time. It seems a really difficult conclusion to accept[25].

Besides, there is no scientific evidence that attributes particular causal properties to the brain so that it may gather the meaning from somewhere (from the physical world where the meaning is not lying around by itself or from the objective world that does not belong to reality in the physical sense).

Finally, let's consider some ambiguous aspects of the traditional concept of the brain. We are accustomed to the idea that we use our brain because our brain is an information-processing machine of elevated complexity. The brain is capable of processing an enormous quantity of information. It is able to receive it through the sensory system, process it and, eventually, provide a useful output. Now let's think about our digestive apparatus (stomach, liver, intestines and some other minor organs) in nobler terms than as usual. It could be viewed as an information-processing device. Its goal is to compute the state of each molecule of material we insert in it by applying a particular set of transformations. Given this interpretation, the digestive apparatus looks like a powerful computing machine. Given the appropriate interpretation, the quantity of information of the whole system is, more or less, even greater than

[24] Among neuroscientists there is a great deal of disagreement about the correct scale at which conscious relevant phenomena are produced. Edelmann has proposed large neuronal groups, Crick believes in patterns of firing neurons while Penrose suggests smaller structures inside microtubules, (Edelman 1992), (Crick and Koch 1990), (Penrose 1994; Hameroff 1998).

[25] The example is a version of the famous case of Eubulide's paradox of the bald man. Such a paradox aimed at showing the fact that many properties supervene on wholes as such and that are not reducible to mereologic collections of parts. It could be argued that the wholes, which Eubulide was referring to, are dependent on the existence of conscious subjects.

that processed by the brain. Nevertheless, it is difficult to believe that a stomach is conscious (or a liver or a digestive apparatus for that matter). Why is the brain conscious while the digestive apparatus is not? What is the magical property of organization of events in our brain that provides them with meaning? Why are our stomachs not conscious?[26] Information is never information by itself. Information becomes what it is when it is linked to meaning.

There is one last epistemic rationale that can be raised against the identity between brain states and conscious states. If conscious content is nothing but neural activity how can we know neural activity itself? In other words, if we have a phenomenal experience of the world, and derive from this that the world is made up of atoms (that are different from phenomenal experiences), it does not make sense to eliminate the epistemic starting point. At least we should need a different access to objective reality. The problem is that we do not have any direct access to what we call 'neural activity': we build it using our phenomenal experiences. We cannot start from something that is built upon something else.

All in all, physical objects, meaning, information and systems like the brain[27] cannot be defined without referring to consciousness. It follows that they cannot be used to define it. In other words, consciousness seems to be an irreducible aspect of reality. Its understanding will entail a complete redefinition of our basic assumptions about what exists and about what we are. Consciousness puts our whole ontology being an unavoidable *weltknot* at stake. The first belief that must be relinquished in order to admit consciousness inside our ontology is the conviction that reality is constituted only by physical facts that can be explained by objective science. The failure of objective science is not the same as the failure of the epistemic enterprise. The identification of objective science with human epistemic powers is perhaos only a temporary flash in the pan mainly due to the technological euphoria of this century. Consciousness exists because it is an empirical fact, and as such it cannot be denied. While the brain is surely part of the phenomenon that produces consciousness, it cannot be seen as the only site of consciousness. It must also be clear that the brain cannot produce consciousness by virtue of its static and dynamic organization (information processing) alone, given the fact that that organization is a by-product of our selection of the world implemented by

[26] Of course we can consider seriously the possibility that the liver could be conscious but then we will be fighting against the problem of the existence of objects. How is it possible to find a boundary for one object (*problem of boundaries*)?

[27] And their by-products as well.

choosing meanings. In order to be functionally linked to the external world information, in the brain, must be related causally with it. What the relevant properties of this link are will be analysed in the next chapter.

Summary

Traditionally, scientists and philosophers have tried to explain consciousness by starting from other entities. On the contrary, we claim that most, if not all of the entities, which are believed to exist independently of subjects, without consciousness are void of meaning. As example we put forward information, objects, meaning and physical objects or a system like the brain. We believe that these four broad categories are not autonomous from consciousness; it follows that they cannot be used to explain it.

Apparently safe concepts like 'information' or 'physical object' are indebted to the meaningful activity of the subjective self. Furthermore, they would not exist without the existence of conscious beings whose consciousness remains beyond their capabilities.

All efforts to reduce consciousness to something purely objective derive from the acceptance of some kind of reductionism. This attitude, albeit useful in many cases, can be deleterious if taken as an ontological principle instead of an epistemic tool.

3 Representation, perception and subjects

It appears increasingly that the main joint business of the philosophy of language and the philosophy of mind is the problem of representation itself: the metaphysical question of the place of meaning in the world order. How can anything be about anything?

Jerry Fodor[1]

Sphinx or 'living image'[2]

We have previously defined a subject as something essential to complete our picture of reality –a subject means a *conscious* subject. The main feature, which distinguishes a subject from objective entities, is its capability of having experiences in the form of representations of the external world. If the essential property of a subject is its capacity of having experiences and being capable of having representations, it is imperative to understand what a *representation* is. Besides, perception can be seen as a special case of representation in which the represented object is an external object. We claim here that *the essential property of consciousness is being able to represent other entities (for example external objects or events) to conscious mental states.* What really is a representation, then?

It is easy to speak of representation when there is a code of some kind. For example names like 'Peter' or 'John' correspond to individuals since there is a social convention to relate those names to the correct individuals. Graphical symbols correspond to stars, musical notes, letters, and whatever due to the existence of a code that puts those kinds of physical events (graphical symbols) in relation to the appropriate kind of events. The problem is that the nature of that relation is derived (as shown previously) from the existence of subjects. A graphical symbol has no physical relation with what it represents apart from the

[1] (Fodor 1987), p. xi.

[2] (Becker-Colonna 1966).

conscious subject's semantic choice. If the nature of semantics seems to depend on the existence of subjects, the following temporary conclusion can then be reached:

> *there is no semantics without subjects but in the same way there are no subjects without semantics.*

Below, the relation between semantics, representation and perception is analysed. A taxonomy of representations is proposed. We distinguish between *autonomous* and *derived* representations. An autonomous representation is a piece of reality that is referring to another piece of reality without the help of other agents. A derived representation is something that is referring to something else owing to a conscious observer. In a purely extensional and objective world, there are *only* derived representations: it is a paradox.

For example, a road sign means something by virtue of the agreement among human beings. Another example is given by the levels of electronic activity in a transistor inside a computer. Those levels mean something because external users assign certain meanings to them. Given any transistor-based machine, it is possible to imagine swapping all the electronic levels of its logical gates (from low to high and vice versa) and to have a functionally equivalent machine. Up to now several different definitions of representation have been given[3]. Another issue relevant to the present discussion is the attribution of a separate phenomenal domain to the experienced quality. To this point, Ned Block's survey is relevant[4]. He maintains that four different kinds of content can be distinguished in literature: representational, intentional, phenomenal, and functional. In this thesis, instead of thinking that there are different kinds of content, a different approach, which consists in equating experience with its representational content, is pursued. We outline a possible taxonomy of representations, based on an autonomous/derived dichotomy. We will claim that mental representations are autonomous representations and how they achieve this status is something that cannot be explained by a purely extensional language. The two rationales stand against the use of not autonomous representations in understanding consciousness. First, if derived representations supervene on conscious observers, their use is affected by circularity. Secondarily, there is the problem of the nature of the relation of representation. In derived representations, relations between representing

[3] Two well known cases are the definition proposed by Michale Tye (Tye 1990; Tye 1996), and, with differences, by Fred Dretske (Dretske 1995).

[4] (Block 1999).

events and represented events are abstract objects. Abstract objects supervene (if they exist at all) on minds. Therefore, relations cannot be used to build a conscious being without going again into a classical bootstrap problem. Conversely, to understand consciousness an *elementary*, *autonomous*, *intrinsic* representational unit is needed.

A few words must be spent on the hieroglyphs representing the Egyptian name of the sphinx at the beginning of this chapter. We chose such symbols for three reasons. First, looking at such strange signs, it is easy to recognize the arbitrariness of the relation between them and their concepts. If the cultural link between their Egyptian creators and us were completely cut (this risk has been run several times and the link has been partially severed), we would not know anything about what they are referring to. There is nothing physical connecting them to their references, nothing extensional. Secondly, according to an interpretation, etymologically sphinx means 'living image' – that is an image, which lives autonomously. In other words, it is an image that is the represented object and the representing entity at once. It is a 'living' image, in the sense that, to be what it is, it does not require an external observer. It is the ideal autonomous representation[5]. Finally, the sphinx is the traditional symbol for the eternal puzzles that challenges human comprehension. Of course, consciousness is the best contemporary match for the sphinx.

3.1 The link between the mind and the world: perception

> *Even the most brilliant scientist could not tell how electrical signals in the brain become perceptions*
>
> Bruce E. Goldstein[6]

Perception is the link between the external world and the conscious subject. It is the point where the external world becomes a representation in someone's

[5] The Holy Host, in the interpretation of the Roman Catholic Church, is another example of an object that has been reputed being an ideal autonomous representation. According to the orthodox theological roman dogma, the Holy Host is the body of Christ after the Eucharist. It is not just a symbol but it refers really and autonomously to the body of Christ.

[6] (Goldstein 1996)

experience. We claim that there isn't any way of defining perception without referring to conscious subjects. A camera, a thermometer, a measurement tool are just physical objects that are capable of letting events occur following a causal relation with other events. In conscious subjects, something different happens during perception: subjects have an experience and its content is related to the external world.

Similarly there is no substantial difference between sensation and perception. While the former has been historically separated from the latter, there are no compelling reasons to maintain such a division. Every physical process can be considered to be outside the mind as long as it is not part of the personal subjective experience. Every physical process can be considered part of a sensory process if it can produce conscious representations of the external world.

The link between consciousness and perception can be criticised since it is possible to speak of perception or of sensory processing in lower animals and robots. Of course, its nature is more metaphorical than the result of a precise theory about the nature of perception. If having a motor activity following the occurrence a certain event in the surroundings of the event were enough, why do not we say that a TV perceives the signal of its remote control, or that a computer perceives my typing on its keyboard? It is as contradictory as saying that a gear is communicating with an engine. Yet it seems much more acceptable to say that a cat is perceiving or that a robot has a sensory apparatus. The main reason, behind such a difference in the usage of the words 'perception' and 'sensation', comes from an implicit anthropomorphic prejudice. The more an animal – or a machine – is similar to a human being the more the use of the word 'perception' seems acceptable. Why are human beings so special in this respect? Because, implicitly, they are recognized as conscious beings. Consciousness and perception are deeply related. For example, no human artefacts perceive the world. It will be impossible to claim that anything perceives from *sensa* while it is impossible to distinguish between subjects and objects. In short, unconscious perception is a contradiction. A few caveats must be made.

Perception is not interpretation. The act of interpretation is frequently confused with that of perception. They are two different activities. As cognitive beings, we are able to interpret the external world by giving it different meanings. When we look at a picture, we can select consciously or unconsciously a different meaning for it (Figure 3-1). This is possible because we possess the different meanings we are going to attribute to different physical events. For example, in the case of the female face/sax player figure (Figure 3-1 in the centre), we can switch from one meaning to another because we have,

previously, perceived 'female faces' and 'sax players'. Someone, who had never seen a sax player, would not be able to see that figure as ambiguous. It would be only a shape resembling a female face. For example, a subject affected by prosopoagnosia is unable to recognize faces. He would look at the same figure and would be unable to see anything more than the sax player. The ability to interpret requires the possession of previously acquired meanings or representations of the external world. *Perception is the process by which conscious subjects acquire these meanings for the first time and not the process by which such meanings are subsequently assigned to other physical events (interpretation)*. If we look at a completely random pattern (Figure 3-2), we cannot give any interpretation to it. Yet we perceive it. If that pattern were presented to us several times in critical contexts, we would end by being able to recognize it and by being able to perceive it as such.

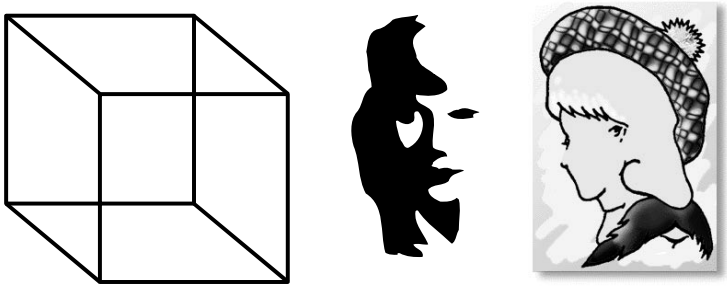


Figure 3-1 Three famous cases of ambiguous figures: Necker's cube (left), Sara Nader (centre), and Mother/Father/Daughter (right). While the first figure flips independently of our conscious will, the other two figures depend on our interpretation of them.

Figure 3-2 A random generated figure. This figures does not show any ambiguity because we do not possess any visual archetype apt to give a meaning to it. Therefore we cannot interpret it.

Blind-sight is not perception. Typically blind-sight patients have lesions of their primary visual cortex and cannot see light consciously, in the parts of the visual field functionally linked to the damaged cortex[7]. However, there has been clinical evidence suggesting that patients can discriminate between various properties of visual stimuli presented to blind areas of the field. For example, they have been shown to be able to discriminate stimulus location, movement direction and velocity, size, orientation and, sometimes, even colour. One neuropsychological explanation, for blind-sight, is that patients lack the primary retino-geniculo-striate pathways required to conscious perception. Their ability to discriminate is not part of what makes them conscious subjects. In other words, they still possess some neurological pathway that provides information to the higher cortical areas. This information is stripped of almost all phenomenal content and is lacking any meaning. Patients do not perceive the information consciously; what they are able to do is a consequence of some neural reflexes and not the result of their cognitive activity.

Information is not perception. When we perceive some information, we do have a kind of phenomenal experience; and having such phenomenal experience (a *quale*) informs us that something is occurring. Yet information, as it is usually defined, is missing all explicit reference to its meaning. Let's consider visual perception. When we look at something, the brain processes, which produce our conscious experience of something bright and colourful, does not possess these properties. The visual nerve is not transparent to light! What is passing along the nerve fibres is different from what we assume should be at their ends. The same is true for every sensory channel. Information needs to be interpreted by associating the appropriate meaning to each pattern of stimuli, but the meanings are not contained in information itself. The same pattern sequence can be used to mean very different things. Similarly, the same neuron firings can be correlated to the perception of very different meanings.

Perception is associated to having representations of events that we suppose belong to external reality. Yet this definition labours under the problematic distinction made between the external world and a baffling internal mental domain. Two possible options might be followed to explain conscious perception. The first is to accept the classic extensional framework, the second is to analyse the nature of representations and to propose a different framework, empirically verifiable, in which representations are possible. The first option can be pursued by proposing some version of the causal theory of perception. The second option stems from the observation that, if we refuse perception as a link with reality as such, we are doomed to fall into the sceptical prison of

[7] (Holt 1999; Kentridge and Heywood 1999; Marzi 1999).

radical doubt. So if perception is really representing something, suitable candidates for representation must be proposed.

Box 3-1 The paradox of perception

The New Bonnet, 1858 by Francis W. Edmonds. Oil on canvas, Metropolitan Museum, New York.

In the painting above a young lady is admiring her light blue hat. If we accept the physical ontology, we must accept the fact that she 'is' her brain. Yet her brain is different from the object of her perception. That object is blue, her brain isn't. The blue, which is the phenomenal property that is passing from the object to her experience, is not passing through her senses. Her optical nerves are not transparent to light only to causal effects and electrical waves. She cannot be experiencing the blue of her hat since that blue has never entered into her skull and into her brain. Apparently, according to the objective extensional ontology there is a insurmountable gap between the external object and the internal activity going on in her brain. No physical objects could ever perceive any other physical external object. Does she just perceive her brain? She doesn't since how does she know that she has a brain? From perception. But if we deny that the object of perception can be directly perceived then the brain cannot be perceived either. What she – and we – believe that is a brain is something which undergoes the same epistemic gap as the blue hat. *If the hat is not what it seems, neither is the brain*. The epistemic opacity, which should eliminate the so-called external objects, eliminates internal structures as well. In the end if the true representational power of perception is denied the epistemic gap destroys the ontological basis of the physical world.

3.2 On the causal theory of perception

> *Whoever accepts the causal theory of perception is compelled to conclude that percepts are in our heads, for they come at the end of a causal chain of physical events leading, spatially, from the object to the brain of the percipient. We cannot suppose that, at the end of this process, the last effect suddenly jumps back to the starting-point, like a stretched rope when it snaps*
>
> Bertrand Russell[8]

In looking for a physical candidate for the activity of representation, a classical option is the use of causation. Functionalist theories – as well as externalist and intentionalist ones – so as to try to make use of causation like a path between external objects and the internal mental processes (the representation). As said above, in order to accept these points of view, we must assume causation is a concept independent of consciousness. However, as a type-type relation, causation cannot be defined independently of the semantic choices of subjects. In other words, the selection of the appropriate chain of events is not logically different from the selection of the appropriate set of entities. Choosing and selecting causes is not different from choosing objects. In this respect, a causal chain is similar to a constellation of related events. After a general analysis of causation here we will examine the paradoxes that arise from a consciousness-independent vision of consciousness.

Let's imagine a girl's brain (Petra's brain) and let's see if it is possible to use causation to define the relation between internal and external events. There are three events A, B, C causally linked in Petra's brain (Figure 3-3). We suppose that between A, B and C there is a classic causal relation. This hypothesis entails that whenever event A happens in the external world, neural (internal) event B is produced as an effect of A. Eventually, internal event C is produced as well. We do not have any reason to suppose any kind of constrains for B and C. They do not have to be unique events (like the attraction between a proton and an electron in a Hydrogen atom, for example). They can be a sparse collection of micro events loosely distributed both in time and in space. They can be a collection of neural activities sparsely distributed in the brain or they can be just a unique single 'grandmother' event. The only requirement is that they be linked causally to external event A. Each time, A happens, B has to

[8] (Russell 1927).

follow. The same is true for B and C and, conversely, for A and C[9]. The first problem is a direct consequence of what we have said in § 2.2. If A is a macroscopic event it will suffer from the same problems of a macroscopic object. It will not be possible to define it without introducing a conscious observer to our definition. Imagine a very simple case of macroscopic causation. If I look out of my window of my lab, I can see the lighthouse of Genova. I could say that the impressive bulk of the lighthouse is the macroscopic object that is causing the internal event I am referring to when I say that I am looking at lighthouse of Genova. I could also divide the stream of that single image into an enormous number of smaller images, even single photons, and follow a different causal pathway for each of them. In this case, we will have a large collection of causation pathways that at a higher level is a macroscopic causation but that, from a physical point of view, does not exist as a whole[10].

As a consequence of physicalism and of reductionism, the higher level exists only as a (conscious) interpretation of the lower level.

Even if the problem of macroscopical causation could be solved, other difficulties lurk ahead. Let's suppose to have another set of external and internal events X, Y, Z with the same kind of causal relations. C and Z are the conscious mental states, while B and Y are unconscious brain states. When C and Z occur, Petra has an experience of A or X. Somehow, following the functionalist paradigm, the causal connection between the external event and the internal ones is carrying the content and meaning of A and X. Each time C happens Petra has conscious qualia of A[11]. First it is very difficult to imagine how that the meaning of A happens to be preserved through a series of

[9] For the sake of the argument we do not enter into the details of how we know that B will always follow A. Let's simply say that we are reasonably sure that it will follow.

[10] What makes a set to be a unit, or a set of causal pathways, is the fact that a conscious being insists on taking it to be a unit. It is not possible to obtain a unified object unless a *principle of unity* is recognized. We think that this principle of unity has to be embodied into a conscious being (Newman 1988).

[11] Of course, we could also assume that Petra is having the qualia of C. In other words, there is no meaning transmission from the external world to the internal world. Experience content is just the meaning of our intrinsic brain structures. Apart from the ugliness of such a standpoint there are two conundrums to be solved. First, if we exclude causation as a meaning carrier why are we conscious only of neural states that are intentionally projected towards their objects? Secondly why should certain states carry a conscious meaning and others should not? Why should all matter, as a whole, not be conscious? It would seem very difficult to stop the panpsychistic wave from absorbing the whole universe.

causation jumps. Secondarily, what exactly constitutes the causation link? What do we mean when we say that B is causally related to A? We can imagine several causal pathways connecting the two events. The basic form is to say that whenever A happens sooner or later B happens too. Let's think about what could bind the two events. We must distinguish between a microscopic causation and macroscopic one. The first could be an intrinsic property of matter that we do not want to investigate any further here because it is on a different level from that which is relevant to the functional structure. The second is much more elusive and difficult to define (not least because we think it does not exist). For example it can be constituted by complicated machinery of pipes and valves, or by the integration cell of a neuron or by a computer able to detect the situation A and to connect it with the reaction B. We have to distinguish two different kinds of causation:

i) causation is confirmed by induction, observing that in our experience the event B follows invariably the event A;

ii) causation is determined by studying and analysing the inner structure of event A and by finding that the structure of A is such that B has to follow whenever A occurs.

The general idea is that the first kind of causation cannot be considered real causation. The reason is that if we accept that «if B follows A then A causes B» then we will have a huge class of causal relations that we are not able to deal with. For example, we should conclude that night is caused by day, winter by spring and so on. The reason for this refusal is that, in these examples, the locality of autonomy of the causation agent is lacking. The causation agent should be the only thing responsible for its effect. The day is not the only thing responsible for the night because there are several other effects that have to be taken into account. Nevertheless, it is true that whenever the sun goes below the horizon line, night falls. The first kind of causation seems to be a product of an interpretation of events more than the expression of an internal objective structure. On the contrary, the second kind of causation requires the ability to detect an inner structure between the two phenomena so that they are, at least, nomologically related. Of course, there are still all kinds of epistemological problems associated with the method we use to assert the existence of a casual relationship. It is also highly probable that the contempt of science for causation will get rid of macrophysical causation too. Even if we could not be sure of it, and even if science could eliminate causation, up to now causation has been used widely to explain consciousness. Thus, we will assume that such a thing exists and we will examine whether it fails in the following six examples.

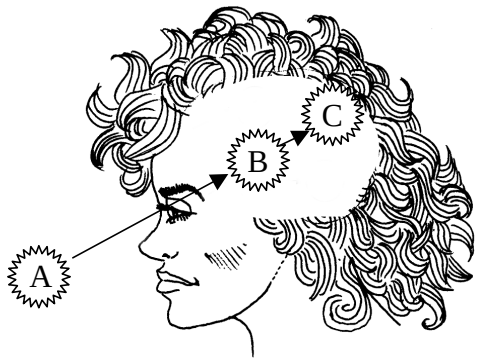


Figure 3-3 Petra and her brain. If the causal theory of perception were true, the meaning of an external object should pass from the external objects to the cortical areas responsible for conscious experience. From a physicalistic point of view, there aren't any theories to explain how meaning transmission occurs. 'Meaning' hasn't any place. In figure, the event C, within the brain, should assume the meaning of the external event A. The event B, between A and C, should not have any effect on the quality of the perceived conscious event. B could be inside or outside the brain.

3.2.1 Meaning transmission

Why should we relate causal relation to meaning transmission? What happens when we perceive something? We perceive the meaning of the external object together with information. From what we have said above, it seems difficult to link causation to meaning transmission. If we observe a causal pathway, we can see that it is transmitting only a causal wave. At every jump, the new effect has nothing to do, by itself, with the originating phenomenon. There is no observable transmission of meaning or physical properties as such. Let's imagine a red flower held in Petra's hands. Its surface reflects only red light waves. The photons hit the inside of Petra's retina. As an effect of this collision, a series of chemical reactions is partially modified in its dynamics. This is no longer a visual event in the strict sense. This is only a chemical event with different colours than the original flower. Along the axons of the several layers of neurons connected to Petra's retina, sodium-potassium chemical reactions rapidly follow each other. If you examine them, you may have no idea that they are a consequence of a remote visual event. Eventually, after hundreds of causal jumps, Petra is having the conscious experience of red. What carried

the meaning from the surface of the flower to Petra's internal brain processes along all the causal jumps?

3.2.2 The little man in the mid of the causation pathway

Let's suppose that there is a little man inside Petra's brain or, more realistically that we have surgically removed a large portion of Petra's brain. The aim is to put the occurrence of the intermediate event (B) under the control of an external agent. Let's also suppose that all functionality in the remaining portion of Petra's brain has been preserved, and that an electronic device capable of reproducing all correct stimuli substitutes the missing part. In other words, we have inserted a silicon substitute – functionally equivalent – to her missing biological part. The only difference is that there is an operator able to interfere with it. Along the causal chain, a deterministic link has been substituted by an autonomous agent. The nomological necessity of the causal path is so disrupted. He (the operator is the little man) can see the neurological input of Petra's brain on a monitor of his own. Further, he can decide to let it pass through the causal chain or he can decide to stop it. He is an honest little man, so he has always pressed the «let it go» button and, since he is so upright, Petra has never missed anything. The causation chain is now different from the one before our intervention. Has this any effect on Petra? The necessity of the internal event in Petra's brain is no longer guaranteed. There is no nomologic reason of any kind to link the conscious event to the external one. The causation that is carrying meaning to Petra is of the kind that we have previously dismissed as insufficient. Nevertheless, Petra's conscious brain activity is identical to what it would have been without this implant. In Petra's brain, the activity is the same and her mind should be the same but … the causal relationship has disappeared! What are we to conclude in this case?

3.2.3 Fingers in the eyes

It is well known that in humans the occipital part of the two hemispheres of the brain is usually dedicated to visual activity. It is also widely accepted that what's going on in these areas is correlated to the conscious experience of visual events. The purpose of this paragraph is to challenge the link between these parts of the brain and the normal quality (visual) associated with its conscious activity. Let's imagine that we cover Petra's eyes at birth and that she is literally kept in the dark. For twenty years the only stimuli she receives through her

optic nerve will be caused not by light but by occasional touching from time to time[12].

There's a familiar case of synesthesia that can be experienced by each of us just by pressing our eyes. If we press our eyes, we can see light just as an effect of a pressure phenomenon. A mechanical event (pressure) provokes a visual experience. If we could carry out the aforementioned experiment, it seems reasonable that the internal phenomenon should remain connected to it, after a relatively long period of causal connection between an external phenomenon (the light) and an internal phenomenon (visual cortex activation). As far as we know, there is no logical reason why this should happen, but let's consider the power of habit. We know that our eyeballs are sensitive to pressure because of the simple experiment we just described. The eyeballs can be used as tactile sensors albeit of a very primitive kind. Let's return to poor Petra, still with her eyes firmly closed, and let's analyse her situation. What in other humans is called visual cortex, in her case, it has never been linked causally to light but only to pressure. We should expect that i) her visual cortex never developed properly and that ii) she experiences something of tactile nature when her eyes are pressed. What will happen if we uncover her eyes? Our intuition is that for the same reason why we can see light when our eyes are pressed, she should feel a pressure when her eyes perceive visual stimuli[13].

3.2.4 Objects are transparent to causal chains

Traditionally the brain is seen as the centre of our cognitive capabilities. It seems obvious that there is a natural boundary for what happens inside the skull and outside it. Unfortunately, if we look at it from the point of view of causal chains, these boundaries seem almost to disappear. It is true that the skull is opaque and so, when we look at it, we can see it as an object; nevertheless from the point of view of causal chains, it is completely transparent. Causal reactions are going in and out of the brain to the outside world (fortunately). If we remove all secondary qualities (brought in by conscious subjects) we are left with no macroscopic objects to be the external carriers of meaning. Furthermore, we are left with no meaningful distinction between a causal effect inside of a brain and a causal effect outside of a brain. A

[12] It is clear that a thought-experiment (*gedanken experiment*) has no reason to be ethical and the thought-experiment described here is highly unethical.

[13] Although it may seem unrealistic, the described case took place several times. It has been reported of congenital blind children claiming to feel tactile subjective sensations when they are exposed to light after surgery (Senden 1932).

causal effect does not know if it has to produce consciousness or just be a 'dull' casual effect.

3.2.5 A brain in a vat has no causes

Once upon a time, there was a brain in a vat. It is usually assumed that the vat-brain should be conscious insofar as we are able to provide all the necessary inputs and outputs (for example by using a supercomputer). It is widely accepted that the brain should believe that it is living in a world, which corresponds to the information we are providing through the supercomputer. From the point of view of ontological reductionism, we ought to conclude that two identical groups of atoms with the same structure should develop the same properties. It is an application of the supervenience of the mental on the physical. Let's suppose that our supercomputer is capable of building a finely reconstructed model of the external world and provide the brain in a vat with a completely consistent environment. For the sake of the argument, we can admit that the vat-brain is passing through exactly the same states it would have passed if it were located in Petra's skull while she was walking through Piazza S. Marco in Venice. If the vat-brain is passing through the same states as Petra's brain, it should correspond to a conscious subject having the same conscious experiences of Petra. We cannot avoid this conclusion because the two brains are exactly the same on an ontological reductionistic basis. Yet, the external causes of the vat-brain are completely different from the external causes of Petra's brain. While Petra's brain is linked causally with the object known as the Tetrarchs' bass-relief, the vat's brain is causally linked with a bunch of transistors in a supercomputer. Even if the states of the two brains are the same, if we think that meaning is carried by causation, we have to conclude that they are having experiences of a very different sort. For a physicalistic it will be very difficult to admit that two physically identical objects are producing completely different effects because of a relation not of physical nature (see § 3.2.2). Ontological reductionism, local supervenience and causal theory of perception are mutually contradictory.

3.2.6 Stopping the causal chain reaction!

All events occurring in Petra's brain have been caused by other physical events. According to the causal theory of perception meaning is somehow carried by causal relations. As a result, the content of Petra's conscious states is different from the neural event occurring in her brain. Yet this event is not

identical to its proximate cause. For instance, the activity in her visual cortex has the activity in her eyes as its proximate causes. Yet Petra is conscious of the external objects and not of the chemical reactions going on in her retinas. She is not conscious of the activity in her optical nerve either. In turn, before the perceived objects there have been other more distant causes. Each object is where it is because a set of events has determined its story; but Petra does not perceive these remote causes consciously.

It is as if, along the causal chain, there is a point in which the content arise. Alternatively, it is as if Petra – by tracing back to the causes of the states of her brain – had come up against a barrier. The barrier is the point in which she becomes conscious of something. Among the infinite series of causes of her neural events there is an event that is the conscious content. Why a particular event? All events occurring before that event are invisible to Petra: she cannot see through the objects normally perceived. All events occurring after that event are invisible to Petra: they are transparent to her. As a result she is somehow coincident with the objects of her perception. There is nothing before and nothing after. Yet there's no reason to choose one event or another. From what physicalism tells us, there is no reason to prefer one cause to another. From a logical point of view it could be equally possible for Petra to be conscious of the first cause of everything (a sort of causal *primum movens*) or for Petra to be conscious of only the last neural event (in this case the causal theory of perception would collapse on identity theory).

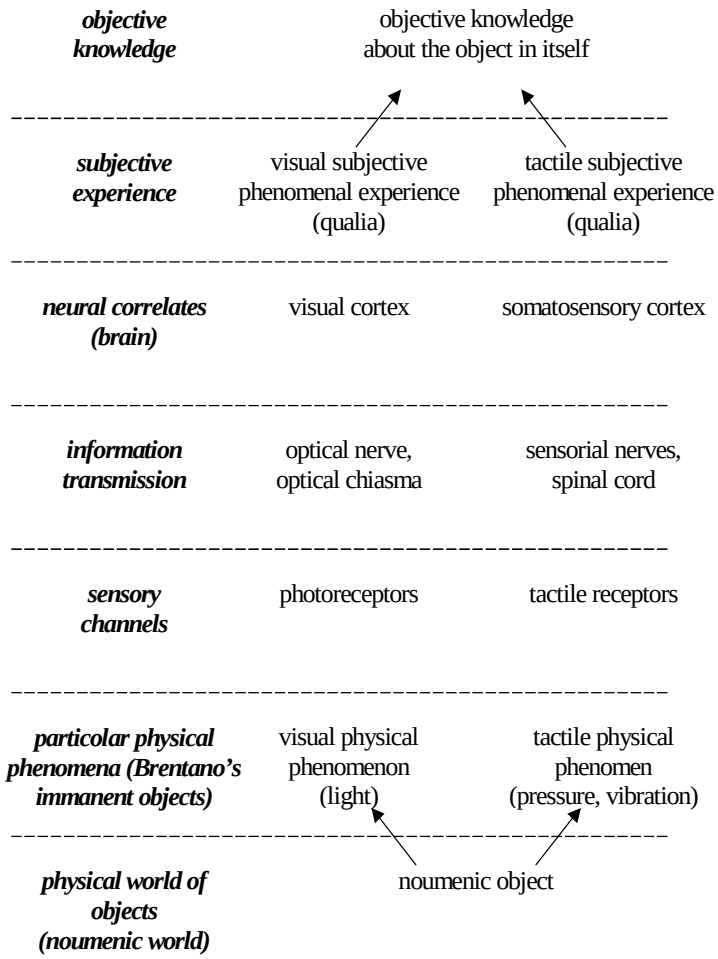


Figure 3-4 A possible diagram of the phases of perception. The physical world is at the bottom while objective knowledge is at the top. Although this series of stages is more or less accepted there are several dishomogeneities lurking between the levels. If there were an efficient theory of mind, such dishomogeneities ought to disappear.

3.3 Taxonomy of representations

> *More generally, how can any state in nature represent anything at all?*
>
> Michael Tye[14]

As outlined above, the most baffling property of mental states is their intrinsic capability of representing the external world. For instance, if we look at a written word on a sheet of paper, let's say the word 'FOX', we see a set of dots on a physical piece of paper. These dots would not refer to an animal if it was not for the presence of a conscious reader who is able to link those signs to a mental meaning and, afterwards, connect it with the appropriate kind of animal in the external world. The same is true for all artefacts. We are so accustomed to associating such mental meanings to things, that sometimes we forget that they have only a derived capacity of representation[15]. Here, a *caveat* is that such a theory of meaning is going against the whole mainstream of analytical philosophy. Locating the source of meaning in the obscure mental entities is dangerously similar to Locke's theory of ideas and meaning. Yet, if a theory of consciousness were to prove itself successful, it could be able to give a more robust ontology to those awkward entities, generally called mental entities (ideas, *vorstellung*, or whatever). Representation would find a foundation. Yet, up to now, as Searle once remarked «representation is the most abused term in the history of philosophy»[16].

What is the generally accepted concept of 'representation'? In order to have an up-to-date definition, we will refer to the definition given by MIT Encyclopædia of Cognitive Science:

> […] We can say that any representation has four essential aspects: (1) it is realized by a representation bearer; (2) it has content or represents one or more objects; (3) its representation relations are somehow "grounded"; and (4) it can be interpreted by (will function as a representation for) some interpreter[17].

According to this definition, in order to have a representation, it seems that we need i) to find suitable representation bearers, ii) to know what content is, iii) to be able to 'ground' such representations and, finally, iv) to have an interpreter. The problem is that these points depend on some pre-theoretical intuitions of

[14] (Tye 1996), p. 99.

[15] 'Mental meaning' denotes some kind of conscious content. § 0.

[16] (Searle 1983), p. 21.

[17] (Wilson and Keil 1999).

what a representation is. Besides, they are dependent on the existence of conscious subjects. In giving a definition of representation, the biggest problem is to avoid circularity.

> *If consciousness must be founded on representation a foundation for representation must be located without making use of conscious observers.*

It is not clear, from a definition like this one, what the relation between the different topics is. Point i) and point iv) are directly related to an external interpreter and, as such, cannot be used to understand what a representation is. Point ii) and iii) are little more than a tautology. For example, point ii) states that a representation has, as its essential aspect, the capability of representing. Point i) states that there must be a physical medium for each representation.. Not very enlightening. Below, these topics will be referred to, collectively, as MITR.

Here we propose a different taxonomy as a working hypothesis. Every representation, we claim, can be classified on the basis of two separate dimensions: the autonomous-derived axis and the similarity-relation axis. The first is related to the degree of autonomy of a representation while the second refers to the way the representation is made. The following two couples of definitions can outline the two dimensions:

> Autonomous representation*: an object that autonomously represents another object. It stands for the other object without the need of any external observer (AR).*

> Derived representation*: an object that stands for another object following an external observer's choice. It would not stand for anything by itself. (DR).*

and

> Representation by similarity*: the representing entity has something in common with the represented object such as logic form, shape, colour, or whatever other property (SR).*

> Representation by relation*: the representing entity has nothing in common with the represented object but there is some kind of relation that links the two (RR).*

A few words will clarify the scope of each of these definitions and will show how they can be combined together. It is possible to suppose that they are mutually incompatible in the sense that for any representation it must be true that (¬RR&SR)&(¬AR&DR). This is not strictly true in all cases but it can be held for the purpose of this discussion. Relaxing this constraint will not bring substantial differences.

Let's examine each of these categories briefly. ARs are what is needed to avoid the aforementioned circularity when defining consciousness. Unfortunately, there are no practical examples apart from mental states. For instance, when the external world is perceived, our mental states represent something. In other words, a mental state is an entity that should be able to refer to content and that is capable of doing so without having to resort to an external agent: this is the foundation of semantics. However, our mental states are known from a first-person (subjective) point of view – a point of view, by definition, not objective. They *have content autonomously*. In principle, this does not entail any commitment as to what content is. It could be representing the phenomenal, or the cognitive, or the intentional or the functional character of conscious experience. In practice, if an extensional ontology is assumed, there is nothing that differs more than the physical things, at our disposal. In a purely physicalistic ontology, the representational bearer of MITR must be a physical object like the one, which it refers to. *If our mental states have to be reduced to something physical, we must have a good explanation about how it is possible for a physical thing to have the meaning of a different physical thing*. Unfortunately, in nature, we don't have one single example of something that represents something else autonomously[18].

Regarding DRs, they derive their representational capacity from conscious beings. Any object can be used as a representation of every other object, given the existence of a conscious observer stipulating the appropriate kind of association: a kind of De Saussure's arbitrariness. The group of letters or the graphical shape of a symbol associated with a concept is arbitrary. A sign is a sign by means of the meaning that someone gives to it. The property of having some meaning is entirely *conscious observer relative*. At the same time, the identical physical object can have several distinct meanings for different observers, while remaining the very same physical object. Of course, the fact that the word 'Franz Brentano' denotes an Austrian philosopher is not a property of that group of letters but rather of the subjects that chose them. The

[18] As already mentioned, there are a few examples of Autonomous Representations in human history. One is the sphinx and the other is the Holy Host in the interpretation of the Roman Catholic Church.

same kind of argument can be extended to other two concepts: information and physical objects. It is possible to argue that being the carrier of a specific amount of information – as well as being a particular object – is not an intrinsic property of a physical set of particles. It is a complex function of that set of particles and of the properties (both physical and semantic) of a conscious observer[19]. The point is that being a DR seems to be a property that requires the existence and the participation of a conscious being. In short, meanings supervene not only on physical objects. All syntactical representations in computer science can be seen as cases of DR. In this sense, a computer is not capable of having mental states because its states have a meaning only by virtue of its users. For instance: «whatever type of internal representations a functionalist system may employ, a procedure is needed to establish the meanings of the individual units […] in those representations»[20]. John Searle has carried on the same concept repeatedly and extensively: «Computation is not only disembodied; it cannot by itself provide a meaningful relation between symbols and world entities»[21].

As far the mechanism underlying the act of the representation are concerned, there are the two following categories. In classical philosophy to represent meant to be capable of reproducing the properties of something. «Mental images, according to Aristotle, must resemble or copy what they represent. The thought lying behind this claim is presumably that real pictures must resemble what is pictured and not just represent it by playing a conventional symbolic role»[22]. Correspondingly, St. Thomas' *imago vicaria* represented something because it was the image, the reproduction of that something. In both cases they were SRs. With Descartes there was no particular problem in explaining representation because a thinking substance (*res cogitans*) is capable of copying the properties of the external world. In this sense the Cartesian ideas are replicas of the external objects even if they lack their existence[23]. The most

[19] (Goodman 1979; James and Kuklick 1981).

[20] (Edelman 1992), p. 226.

[21] (Searle 1992), p. 114.

[22] (Tye 1991), p. 2.

[23] Apparently most of the authors agree that Cartesian mental images are replicas of external objects. For example for (Tye 1991) «Descartes, like Aristotle, holds that percepts (and mental images) copy objects in the external world» p. 4. This opinion is supported by Cartesian passages like this one: «quas tanquam a rebus extra me existentibus desumptas considero, quaenam me moveat ratio ut illas istis rebus *similes* esse existimem (with reference to those that appear to come from certain objects out of me, what grounds there are for thinking them similar to these objects.).» and «Nihilque

obvious drawback of such a notion is the infinite regress of Hume's theatre. The concept, which is the commonsense option, started to enter into a critical phase with Locke who ambiguously argued «that the ideas of primary qualities of bodies are resemblances of them, and their patterns do really exist in the bodies themselves, but the ideas produced in us by these secondary qualities have no resemblance of them at all[24]». In other words, the similarity between the mental objects and their external counterparts was becoming less plausible. In his famous introduction Kant observed that «our representation of things as they are given to us, does not conform to these things as they are in themselves[25]». But puzzlingly in a previous writing, he had observed that what he called «referring to» derived from some kind of conformation with the external objects[26]. The commonsense notion of resemblance was showing increasing difficulties when applied to mental states. Nevertheless it has survived in several areas. At the beginning of XXth century, gestalt psychologists believed that the reaction of the brain to the experience of a circle should correspond to an electric field of circular shape in the brain. It should have the same shape (the same property) of the represented object[27]. The immediate problem is that

magis obvium est, quàm ut luc judicem istam rem suam *similitudinem* potius quàm aliud quid in me immittere» (And it is very reasonable to suppose that this object impresses me with its own similarity rather than any other thing (Descartes 1641), III, 8. This interpretation is in contrast with the nature of mental entities that are completely different from their references. For example, (Hacking 1975) claims that Cartesian images are completely different from what they represent. It is possible to find passages in which Descartes states this opposed point of view quite explicitly: «& quamvis ad ignem accedens sentio calorem, ut etiam ad eundem nimis prope accedens sentio dolorem, nulla profecto ratio est quae suadeat in igne aliquid esse simile isti calori, ut neque etiam isti dolori, sed tantummodo in eo aliquid esse, quodcunque demum sit, quod istos in nobis sensus caloris vel doloris efficiat; & quamvis etiam in aliquo spatio nihil sit quod moveat sensum, non ideo sequitur in eo nullum esse corpus» (And, though on approaching the fire I feel heat, and even pain when approaching it too closely, I have, however, from this no ground for holding that something resembling the heat I feel is in the fire, any more than that there is something similar to the pain) (Descartes 1641), VI, 15. In short, there is substantial evidence to note a tension in Descartes' writings about the real nature of images, a tension that is largely anticipating Locke's ambiguities.

[24] (Locke 1690).

[25] (Kant 1958).

[26] (Kant 1783).

[27] (D'Agostini 1997)..

it is very difficult to suppose that when someone experiences a delicious taste of cheese, the brain produces an electric field with the delicious smell of cheese (see § 1.3.2). While spatial properties are reproducible, albeit with both theoretical and practical difficulties, other kinds of properties aren't reproducible at all. Notwithstanding such counterexamples, there were more sophisticated attempts to find some way of reproducing the external world inside the experiencing subject's brain. The model of Marr and other visionists represents one of the most up-to-date examples[28]. Usually the attempts dealt more with visual representations (easier to reproduce) than other sensorial modalities. Recently it has been observed that «we are required to think of representational content as a special kind of *correspondence* between intrinsic properties of neural activation pattern and aspects of the world», and that «representation exploits a *structural isomorphism* between its physical substrate and its physical domain»[29]. While the notion of structural isomorphism is a synonym of SR, the idea of a correspondence introduces the RR.

An RR delegates the problem of representation to some kind of relation: and the burden of the correspondence depends on the nature of the relation. Several kinds of relation have been proposed to fill this gap. For example, given the existence of conscious observers in a dualistic style, RR can collapse on DR. Unfortunately, if consciousness is to be based on representation DRs are not the right kind. The most famous example is given by the notion of intentionality in the sense of *aboutness*[30]. If there was such a thing as representation, intentionality could be the basis for ARs as well as RRs.

In short Table 3-1 summarizes the different kinds of representations. As it is possible to notice, in the left column there are no valid extensional candidates, while in the right one there are no valid candidates to sustain consciousness. The conclusion of this paragraph is that a purely extensional world cannot sustain autonomous representations and, consequently, neither derivative representations. To explain representation and correspondingly consciousness, a new hypothesis must be advanced whose first goal should be to propose a convincing candidate for ARs.

[28] (Marr 1991).

[29] (O'Brien and Opie 1999), p.180.

[30] (Brentano 1973; Searle 1983).

	Autonomous Representation (AR)	**Derived Representation (DR)**
Relational Representation (RR)	Husserl's and Brentano's intentional objects, Locke's ideas, Kant's phenomena, Holy Host	A road sign, written words, electronic levels in computers
Similarity Representation (SR)	Aristotle's and S.Thomas' imago vicaria, Cartesio's impressions, Locke, gestalt electromagnetic fields, Structural isomorphism of neural patterns	A picture, a portrait, a statue

Table 3-1 A taxonomy for representations. In the column of AR there are no natural of physical carrier of representation. How can representation be naturalized if there are no natural autonomous carriers of representation?

3.4 Maps

In biological systems as well as in robotics, when dealing with maps and representations, it is easy to forget that their true nature lies in being a collection of semantic relations with physical entities, which represent their content. Here the difference between a notational system and a real map, between an autonomous and a derived representation, between an extrinsic criterion of correspondence and an intrinsic semantic relation is emphasised. A taxonomy for maps is proposed. What really is a map or a representation? Working in the field of robotics or of neurophysiology, it is common to use these concepts to define several features of functional parts of the biological cognitive sensory systems and of robotics architecture. Both maps and representations deal with semantics. That is, their meaning refers to something that could be physically outside of the system involved. Both maps and representations deal with syntax because they are part of the internal computations of the system, which they belong to. The interactions between the twofold nature of both concepts cannot be underestimated.

Basically, two classic kinds of maps can be proposed. In everyday life, a map is usually a graphical representation of a spatial area where several locations and objects belonging to that area together with their mutual relations are reported[31]. These two classes of entities (objects to be located and their relations) are both important in order to produce a map. However, if the second class (the relations) were eliminated we would no longer have a map but simply a list of names (something very similar to the ancient Roman compilation of names of towns). Conversely, if the first class (the referred objects) were eliminated we would no longer have a map but a mere abstract structure. In the first case only the semantic structure would survived, in the second case only the syntactic one. The capability i) of referring to some kinds of entities in the external world (or in some logical space) and ii) of representing some kinds of relations among them are essential for a map to be a map. In short, semantics versus syntax.

Neuro-scientists and robotic engineers frequently use the syntactical kind of map. It represents a set of entities (usually geometrical points, or pixel, or force fields, or whatever) whose only relevant meaning is given by the relations they have with the other entities. For example, let's imagine a squared array of N×N pixels. It is just a brief notation to denote a set of N×N entities each of which has as its intrinsic meaning the property of being located in a particular place of that squared array. The pixel $P_{i,j}$ is a convenient way to denote the pixel that is just one unit *to the left of* the pixel $P_{i-1,j}$ and just one unit *lower than* the pixel $P_{i,j-1}$. Other possible meanings are available. For example, being the pixel that is i units to the left of the pixel $P_{0,j}$. A problem arises. Two different relations have to be introduced in order to give sense to this structure: the *being to the left of* and the *being lower than* relations. Are they innocent entities or are they imposing ontological commitments on the peaceful neutrality of our original conception of the second kind of maps? In other words, a map is a map, even if some kind of denotation, at least to the relations defined among the entities of the map is acknowledged? The answer seems to be no, at least if we are willing to define useful objective maps referring to properties belonging to the physical world. Let's imagine that the map of the previous example had been defined as a N×N set of P elements and that each element $P_{i,j}$ had the property of being one unit more *blurp* than the element $P_{i-1,j}$ and one unit *grund* than the element $P_{i,j-1}$ (where *blurp* and *grund* are just two meaningless labels). Would we have defined a map or just an abstract structure? We advocate the latter thesis. A map is essentially a semantic structure whose nature is to denote objects and

[31] (Kosslyn, Thompson et al. 1995).

their relations in some real or logical space referring to physical entities. Without this semantic attitude a map is simply not a map.

It follows that the simple geometrical example of a squared array of pixels is just a short way to denote a set of physical entities (pixels of an image or just space locations). It means that the geometrical structure denoted simply by using the notation $P_{i,j}$ must be reduced to its semantic content in order to be meaningful. Besides, it means that the notation $P_{i,j}$ is just a shortcut to avoid drawing a real map of that squared array. $P_{i,j}$ is short for '*the pixel whose space location is i units to the left and j units lower than the element* $P_{0,0}$'. A meaningful definition of this is necessary in order to have the semantic relation with the following entities: the origin point $P_{i,j}$ and the geometrical relation (*to the left of* and *lower than*) or, in alternative, to the semantic relation with each of those entities. For example, for a very experienced chess player, each position on the chessboard is not a mere logical location but rather a precise entity with a definite meaning (whatever this meaning might be). The player can give a different name to each position on the chessboard. He/she does not need a Cartesian notation to identify each position. In other words, it is like passing from a naming system like avenues and streets in New York to a system based on historical names (although usually less organized and homogeneous). The fact that in engineering or scientific maps the first naming system is practically always preferred because of its practical and direct of notation must not hide the fact that a map is always a semantic map. Independently from the notation used, a map is a semantic structure.

What is the essential difference between the two systems? It is the existence of a handy notational system that is nothing but a typographical convention. Without it, a dictionary would be needed: a long list of semantic arrows to allow its readers to point at the semantic targets of words. The difference between a dictionary and a map is the fact that a geometrical map owns a typographical generator of names. Hybrid examples between a simple dictionary and a Cartesian map are notational generator of codes such as the system used in big companies to denote one particular mechanical part among the several thousands used daily. In this case the map is a set of semantic relations, not the handy and practical rules we are using to refer to such semantic relations.

An example, familiar to neuroscientists as well as roboticists, is given here. It regards geometrical mapping of images and the meaning of perceived objects in space. Let's consider a normal space in front of an observer, a geometrical three-dimensional space containing objects of various shape and size, at certain spatial locations. The classic idea is that points in space can be mapped using a Cartesian system. Points $P_{ES} \in$ *External-Space* should be represented by logical points $P_{CM} \in$ *Cartesian-Map*. The second set of points P_{CM} should preserve the

original and relevant relations existing between points P_{ES}. Let's now suppose that the scene is seen by a biological system, more precisely the human vision system. Several layers of non linear mapping should be added. For simplicity let's say $P_{RM} \in$ *Retinal-Map*, $P_{GM} \in$ *Geniculate-Map*, and $P_{VM} \in$ *Visual-Cortex-Map*. These steps ought to introduce a series of modifications on the relations between those points. The geometry of lenses, the passage from a three dimensional space to a two dimensional projection and perhaps the not uniform distribution of photoreceptors on the retina itself, are all causes of modifications of the original mapping[32]. The final result in the visual cortex would be a very different mapping of points from the original one. Some points might have even disappeared! Applying one handy, practical system of notation to the final product would give completely different results with respect to the original disposition of points. For example, two points on the final image can be relatively near while the same two points in the original space might be relatively far away.

The real problem is that the notational system we are using is simply not semantically correct regarding the meaning of each point. Such a system is imposed from the outside to relate semantic values and there is no reason why such imposed system should be correct simply because it is arbitrary. Nevertheless, the problem does not belong to the map but to the notational system that has erroneously been confused with the map itself. When a map is built, the problem is that its intrinsic and natural semantic system should be used but what is such a system?

Representing the previous series of transformation such as

$$P_{VCM} = T_{VCM}\,(T_{GM}\,(T_{RM}(P_{ES}))) = T_{total}\,(P_{ES})$$

However, a part from exceptions and ambiguities, in order to perceive the original world P_{ES} an inverse operator $T^{-1}()$ such that $P_{ES} = P_{PES} = T^{-1}(T_{total}(P_{ES}))$ is needed. Sadly, there is absolutely no evidence of the existence of such a function inside any biological system and, besides, in normal subjects, reported perceptions of the external world seem to be entirely Cartesian even following all the non linear transformations introduced by the peculiarities of sensorial systems[33]. How can this apparent paradox be possible? The answer lies in the nature of the relation between mapped points and their related content. In a biological system there are no notational systems that from an external and

[32] (Schwartz 1977; Wilson 1983; Carpenter 1991).

[33] (Kandel, Schwartz et al. 1991; Eric 1994). About this point, Kevin O'Regan's work is particularly interesting (O'Regan 1992).

unnatural point of view expect to assign their content to each mapped point. In natural systems there is only one possible natural principle that might connect a representation with its natural content. This principle is the causal chain leading to the physical cause that is responsible for the mapped points. The only possible conclusion is that each point on the Cortical Map must have precisely the content of the corresponding point in the external three-dimensional space. In other words, it is as if each point had its own private semantic chain providing it with its content. Conceptually, this corresponds to changing from a content given to each mapping point of the cortex from an external and arbitrary notational convention that cannot be built into that physical mapping to an intrinsic and semantic theory of representation. *A map is nothing but a bundle of semantic relations, which are representations*.

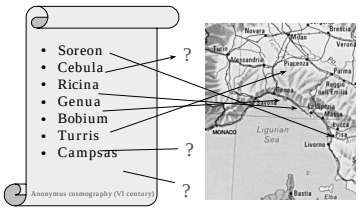


Figure 3-5 A list of names is a structure with a poor syntax and a very strong semantics. If the link is lost there is no way of connecting the symbols on the right with the corresponding places in physical world.

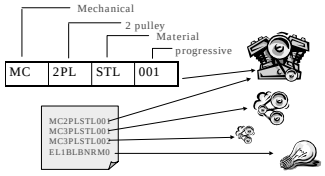
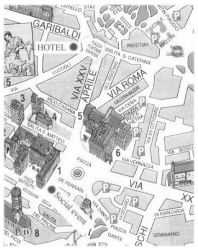


Figure 3-6 Code generator. The syntax could be more or less strong while the semantics still plays an important role.



Genova like
(poor

Names don't say anything about the position

Manhattan like
(stronger

Names say something but not always

Figure 3-7 City maps. In old town there is usually a poor syntax (street names do not say anything about their location). In recent towns there is a stronger syntax (the name of a street in Manhattan shows its location).

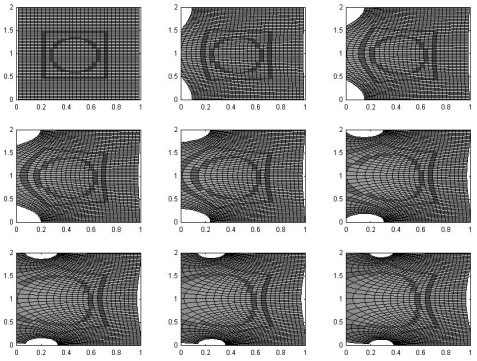


Figure 3-8 Abstract logical maps. There are no explicit indications to what they refer to. They are relevant because they owns a precise syntactical structure[34].

[34] The map shows a disparity transformation on a log polar plane (Manzotti, Sandini et al. 2001).

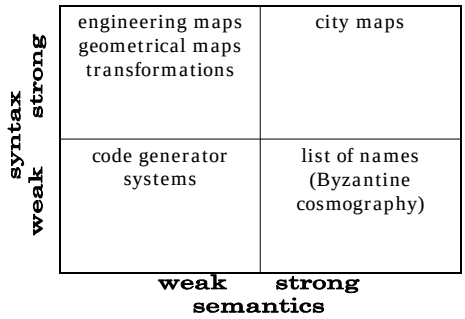


Figure 3-9 A comparison between maps with various degrees of semantics and syntax in their structure.

Summary

The heart of our being subjects lies in our having representations. Perception itself is a form of representation. When subjects perceive the world, they represent it. Yet, there is still no accepted scientific theory of how our brains represent the external world. A series of paradoxes arise when an extensional object like a brain assumes the meaning of different extensional objects. Perception seems to be a paradox in itself. The analysis of the causal theory of perception produces a series of conceptual problems that are conceptually analysed.

The essence of perception is its supposed capability of referring to external objects or events. This capability is usually identified with intentionality. Semantics, perception and representations seem to possess a deep link of some kind. A series of candidates for representation are examined. The result is that there are no acceptable candidates for representation in an extensional world. Yet we live saturated in a world made up of representations. Everything we come in contact with is brought into our experiential world by a representation.

The nature of maps is then examined. A map double structure syntactical and semantic helps distinguish between these different aspects of representation.

4 Requirements for a theory of intentional subjects

> *The role played by simplicity cannot be overstated. […] It may be, for example, that we will find overarching laws that subsume the phenomena of both physics and consciousness into a grander theory.*
>
> David Chalmers[1]

> *A major danger attending any revolutionary proposal in the sciences is that too much of the 'old view' may be discarded – that healthy babies may be carried away by floods of bathwater.*
>
> Andy Clark[2]

Before entering into the details of the proposed Theory of Mind, a few words must be said about the criteria to be followed for its formulation. Some problems must be highlighted immediately.

Let's suppose we want to upheld a theory according stating that proprieties x and the proprieties y are derived from a more fundamental set of properties z. Any attempt to use x or y to explain z would be manifestly circular and therefore a failure. How is it possible to avoid this mistake when trying to delve into the more fundamental aspects of language and reality? There are no easy answers only a few suggestions. All terms should be used at their face value and all hidden connotations ought to be ignored. We must examine each concept searching for obscurities or faults. Anything that is not based on sound foundations must be avoided. Of course a similar way of proceeding might be suspect because of its manifest appeal to intuition. Nevertheless, since a theory of the mind cannot be anything but a theory of ourselves, and since such a theory cannot but touch our capacity of judgement, intuition must play a central role.

In the following paragraphs, a series of criteria are proposed in order to compare the different theories that wish to explain consciousness. Some of

---

[1] (Chalmers 1996), p. 216

[2] (Clark 1997), p. 22

these criteria can be used generally to evaluate different theories. How can the criteria themselves be evaluated? *Quis custodiet custodes*? We believe that the fundamental principle lies in considering experience (conscious experience) as the ultimate source of knowledge about reality and the ultimate judge about our theory about the constitution of the world.

4.1 Ontological economy (Ockam's razor)

Between two theories – both capable of explaining the same phenomena – there is always a difference in the number of entities used. For instance, it is possible to explain the nature of gravitational attraction providing different principles for the movement of heavenly bodies and for the movement of earthily ones. In heavenly spheres a body follows perfect circles along concentric spheres called epicycles, while on the earth a body moves along straight trajectories towards its natural place that is the centre of our planet: two principles for the same phenomenon. However, if we accept the universal principle of Gravity proposed by Newton it is possible to explain both classes of phenomena by using just one principle: the gravitational force. Newton's theory is more economical from the point of view of abstract entities that have to be used. The same can be said of the great theoretical unification of the XXth century: electromagnetic and nuclear force.

In the same context there were two different principles for explaining two apparently different physical phenomena. Yet thanks to Newton's theory of gravitation a unique principle for both phenomena can be found: gravitational attraction. Newton's theory is better than previous theories since it is cheaper from an ontological point of view. Nowadays, the faith in this progressive reduction of explanatory principles drives physics towards a great unification of physical forces[3].

Given two explanations of the same group of phenomena, one that uses less ontological entities is invariably preferred. Yet, Ockam's principle is founded on anything but our preference of simplicity and the evidence of an extremely long list of successes. The former motivation is nothing but a hope, while the latter cannot constitute a proof. It is possible that, given a set of phenomena and two competing theories (both capable of explaining the phenomena), researchers will choose the simpler one. After a few years, new empirical facts

[3] This goal is not always successful to the same degree. For example when Einstein searched for unification between electromagnetism and gravitation, his attempt was unsuccessful.

not compatible with the simpler of the two theories are discovered but they can fit in the framework of the more complex one. There are historical examples: given the limited astronomical knowledge the Middle Ages, the hypothesis that the earth lay motionless at the centre of the universe was simpler than the hypothesis that we were on a globe rotating at enormous speed and rocketing in an immense void space.

Yet, as soon as further astronomical facts were recorded, the theory of the earth at the centre of the universe became insufficient. The winning theory, following Ockam's principle, became inadequate. In that case Ockam's principle was wrong. Generally, Ockam's principle holds (albeit with reservation[4]) if *all relevant* facts are known. If this were not true, Ockam's principle would not allow any valid inference. «What are the relevant facts?» and «When is it possible to be sure to have collected them *all*?» are questions doomed to remain without answers. Thomas Nagel wrote:

> Any reductionistic program has to be based on an analysis of what is to be reduced. If the analysis leaves something out, the problem will be falsely posed. It is useless to base the defence of materialism on any analysis of mental phenomena that fails to deal explicitly with their subjective character[5].

Yet, Ockam's razor has often been a precious tool and it was – and is – the only universal criterion which allows us to compare theories capable of explaining the same facts. Once we have collected all empirical facts we have no other way of choosing among equivalent theories. Maybe the most important thing we can derive from this principle is that no empirical fact can be rejected. *All* empirical facts must be explained and no accepted theory, independently of importance or past successes may reject even *one single* empirical fact for which has still to be found a suitable explanation.

4.2 Direct experience

What is the final demonstration of a theory? When a subject recognizes a theory as the true description of reality? Direct experience is a universally

[4] The fundamental critique is that Ockam's principle is an epistemic criterion that claims ontological validity. This principle works using descriptions of reality and not on reality itself. Entities that are not to be multiplied are epistemic entities (concepts). Yet the undemonstrated assumption is that this economy reigns also at the ontological level of reality.

[5] (Nagel 1974), p. 437

accepted example (let's think of Galileo's telescope). What is the difference between direct experience and a classical scientific experiment? In direct experience there must be a conscious experience of at least one conscious subject; in a scientific experiment this link must remain in an objective domain with no links to the subject. For this very reason, direct experience might appear suspect because it openly makes use of conscious experience[6].

A practical example of direct experience is the following. Let's suppose that I want to show that pain has an extremely unpleasant phenomenal quality for someone who, due to a genetic anomaly, does not have any direct subjective conscious experience of it. Could he understand what the quality of my pain-experience is[7]? If there is no direct conscious experience of something, it is impossible to have any knowledge of the associated phenomenal state. Such quality cannot be described objectively. *The only way of communicating the subjective content of experience to other people seems to be trying to provoke the same experience in them*. If I want someone to know what I feel when I get pinched I can pinch that person. If we cause pain in a normal subject, the person would immediately know what pain is (at least its subjective pain). The problem of proceeding this way is that it depends on the physical and mental structure of individual subjects and on the acquired knowledge of these structures.

Let's now imagine building a device that can modify conscious states. This device is capable of modifying only the phenomenal qualities of experiences without affecting any objective elements like behaviour. Could it be possible to show the efficacy of this device objectively without resorting to direct experience? Is there proof of what it is doing without having a direct experience of it? No. Yet if a subject tried out the device on himself/herself, he/she would immediately be convinced of its efficacy. Can we accept this direct experience as a proof? We think so, since, if this possibility is ruled out, all empirical facts that are known only through a subjective experience must be excluded from reality. It is not impossible that, in the end, all facts (both subjective and objective) will turn out to be based on phenomenal experience. Following the previous rationale this would entail the cancellation of reality as well.

---

[6] A classical example of this kind of subjective judgement is given by the paradox of phenomenal judgement (Chalmers 1996), p. 150. Our 'objective' judgements are based on our phenomenal subjective experiences: on our ability to compare different subjective experiences whose content is intrinsically subjective. Morris Schick raised the same problem many years before (Schlick 1938).

[7] It is an obvious application of the knowledge argument raised by (Jackson 1986) or by (Nagel 1974).

4.3 Explicative power and predicting capability

A proof of the robustness of a theory is its capability of predicting events that have not happened yet: events that no other theory is capable of foreseeing. The astronomer who predicted a solar eclipse for the first time at court of a Chinese emperor had a well-deserved triumph. The ability to predict the future is the aspect that, more than any other, shows the relationship between a theory and nature. Yet all our positivistic faith must still have its roots in a supposed principle of uniformity that reassures us against Hume's scepticism.

The authority of modern science is, for the most part, based on its capability of predicting events before their actual observation. Thanks to the empirical confirmation of such predictions, science has rightly come to stand for its impartiality and objectivity. Yet, nowadays, objective science must face an apparently insuperable obstacle. No scientific theory predicts the arising of consciousness from matter but consciousness is an empirical fact (the *first* empirical fact). No scientific theory is capable of making any suggestion about how to deal with phenomenal experience. Jaegwon Kim said

> We are not capable of designing, through theoretical reasoning, a wholly new kind of structure that we can predict will be conscious; I don't think we even know how to begin; or indeed how to measure our success.[8]

Current scientific theories, being objectivistic in their structural framework, do not even know how to accept empirical subjective facts among the reputable objective facts. Of course, a theory capable of predicting phenomenal experience might run for the role of a global theory (mind and matter might be defined conjunctly).

Predicting the properties of phenomenal experience (its existence and specific qualities) is a crucial point. The experiment might require a redefinition of the experimental protocols in such a way as to address subjective facts without carrying out their impossible translation into objective reports[9].

The optimum would be to find a crucial experiment, as has happened for most of the scientific theories that effectively have revolutionised in previous categories. Something like Foucault's pendulum for the rotation of the earth, the precession of perihelia of Mercury for general relativity, the falling of a feather and a piece of lead for the inertial movement. It should be possible to propose some circumstances in which every theory gives a different prediction

[8] (Kim 1998), p.102.

[9] An example in this direction is represented by the work of (Varela and Shear 1999; Varela 2000).

(for example dealing with the when and the how of conscious phenomenal experience) and in which only one theory succeeds in predicting it.

4.4 Experiential adequacy

Each statement dealing with a theory of mind must find a direct correspondence with empirical facts –both objective and subjective empirical facts are suitable. No fact can be rejected because of any abstract restrictions, or any abstract framework.

From this point of view the optimal theory of mind is a theory super-empirical. Nothing that is part of an experience can be *a priori* discarded in order to facilitate or simplify the structure of a theory. As an example, let's think of neo-positivism (or positivism) that accepted only so-called objective facts as real. Although neo-positivists were willing to use empirical facts only, they ended up using only a subset of the total empirical domain (objective facts or even reports about objective facts). They pretended to derive all knowledge about reality from an *a priori* narrowed window. An ideal theory of mind should not restrict experience as such in any way, but ought to accept both objective and subjective facts.

Each and every entity belonging to experience must find a place in the description of reality: this is real empiricism. Every attempt to reduce any portion of experience to mere appearance must be regarded as metaphysics of the worst species. Moreover, every proposed entity, if real, must entail a difference in empirical experience. This is a way of bridging the gap between the ontological problem and the epistemological one. Besides, to say that a fact entails a difference in the empirical domain entails that the fact entails a difference in the experience of real subject – i.e. the difference in the conscious experience of a real subject.

In practice, what does complete adherence to empirical experience mean? It means that the Cartesian list of properties of mental entities must be used as a compelling starting point (Table 4-1). A framework capable of dealing with this must be searched for. Yet, if the handy objective entities used by science up to now turn out to be inadequate to do this job, what must be done? Should a portion of empirical experience be denied or should the abstract framework of science be radically reformed? We opt for the second[10]. Objective physicalistic metaphysics has failed to achieve its ambitious aim: so much the worse.

[10] Clearly other authors have preferred the first option. Daniel Dennett provides a pure example. See footnote 16 of § 1.3.

Mental entities	*Extensional entities*
They have qualities	They do not have qualities
They have content	They do not have content
They have subject's dependent properties	They do not have subject's dependent properties
They are unities	They are just what their parts are
They are privates	They are public
They represent	They do not intrinsically represent anything

Table 4-1 Comparison between the properties of mental entities and material (extensional) properties. They look rather different.

The idea that subjective facts are real has gained wider and wider acceptance[11]. Leopold Stubenberg makes a straightforward statement about this concept in what he calls *principle to phenomenological adequacy*.

> I will reject everything that does not square with what I take to be the phenomenological data. […] 'So much the worse for phenomenology' is not a viable option for one who adheres to the principle to phenomenological adequacy. The phenomenology is that which the theory of consciousness is supposed to illuminate. If a theory requires us to disregard the deliverances of phenomenology then it is not the theory I seek[12].

In practice, Stubenberg and others refuse the dogma of the exclusive acceptability of objective third-person facts. Not only, doesn't this entail any return to introspection, but also that it makes it possible to argue that, from an epistemic point of view, objective facts are derived from subjective ones and that the former cannot be more real than the latter[13].

4.5 The compatibility of empirical science

A further criterion is the applicability and compatibility with empirical sciences. Frequently a theory of consciousness has been viewed as the last

[11] Among the others Chalmers, Block, Searle, Shoemaker, and Stubenberg.

[12] (Stubenberg 1998), p. 36.

[13] Many researchers are looking for a way of mixing subjective reports with objective ones (Shoemaker 1994; Varela and Shear 1999; Varela 2000).

chapter in the last volume of a neurosciences encyclopædia[14]. Reality might be different. It is further possible that a complete theory of conscious mind might reveal a wider horizon for normal science. To have an explanation of the mind it might necessary to build new foundations both for the mind and for the material world as such. Of course this theory must still be loose compatible with what is known of the physical world. In this anticipated theory of mind, empirical sciences would acquire that meaning it has never acquired in its own right. It is also conceivable that a theory of mind might shape itself around psychophysical laws like those proposed by David Chalmers[15].

> These fundamental (or *basic)* laws will be cast at a level connecting basic properties of experience with simple features of the physical world. The laws should be precise, and should together leave no room for under-determination. When combined with the physical facts about a system, they should enable us to perfectly predict the phenomenal facts about the system.[16]

Even this kind of bridging principles might be incapable of spanning the real nature of mind since it belongs to the old dualistic framework. A more radical revolution might be needed.

The importance of merging together subjective domains with objective science must not be underestimated. Hopefully, empirical sciences should extend their traditional scope to a new domain of facts thanks to the bridging principles deriving from a great unification (something similar to the just mentioned Chalmers' psychophysical laws). Empirical science would maintain its control over objective facts. There would be no exception to the causal closure of the physical and objective realms. For example, a theory of consciousness that had to to suppose a direct action on matter by some kind of spiritual substance not belonging to the objective world would not be a theory compatible with the present scientific framework. The physical world must maintain its supremacy within its proper boundaries and, from the point of view of objective facts, its closure.

The framework into which objective facts have been placed in the last four centuries of scientific advancement must be perceived as an advantage rather

---

[14] For example, Antonio R. Damasio claims that «solving the mystery of consciousness is not the same as solving all the mysteries of the mind. Consciousness is an indispensable ingredient of the creative human mind, but it is not all of human mind, and, as I see it, it is not the summit of mental complexity, either. » (Damasio 1999), p.44.

[15] (Chalmers 1996). Not casually, Chalmers defines his own position as a sort of dualism of property.

[16] (Chalmers 1996), p. 277

than an obstacle. A correct theory of mind cannot be independent of our knowledge of the physical processes underlying our mental activity, and it cannot fail to address the characteristics of out mental world directly: subjectivity, first-person perspective, unity, representation, and having content. Too many theories of mind – of purely theoretical nature –had no link with all the empirical data collected by scientists. When dealing with the brain too often scientists did forget the fact that there is always a conscious subject behind those grey cells.

It is conceivable that a convincing theory of mind might change the meaning of many present-day scientific theories and the rightful domain of such theories (objective facts). In scientific research this is something that can always happen. General relativity did not change the equation of the gravitational attraction but it gave a new meaning to the known concept of space, time and speed. «Philosophy never reverts to its old position after the shock of a great philosopher »[17]. As far as we know, it is improbable that the study of consciousness could reveal unknown physical phenomena. Like Newtonian laws keep their validity in most of circumstances, so traditional mental concepts continue to be applicable. The very emergence of consciousness, as supporters of emergentism have sometimes stated, is a void concept: or the emergent phenomenal property is a physical fact (and therefore it is not ontologically emergent) or it is not a physical fact (and thus it is not emergentism but dualism). Besides to date there has been no convincing proofs, up to now, of any special kind of physical phenomenon going on in our brain.

Nevertheless the real challenge that a theory of mind must accept is the apparent diversity between physical objective facts and subjective phenomenal facts, along with the definition of a wider framework that could accommodate both of them without necessarily reducing one to the other.

4.6 Everyday experience compatibility

The proper domain of a science of mind should include everyday life and should explain how commonsense psychological theories arise: this is the commonsense framework of beliefs/desires, which we usually adopt to understand other people's behaviour. As Jerry Fodor wrote, a theory of mind that does not respect the efficacy of such concepts should not be taken seriously into consideration.

[17] (Whitehead 1927), p. 56. About the effect of a change in normal categories and the way it affects the activities of researchers see also (Popper 1959; Kuhn 1962).

> The main moral is supposed to be that we have, as things now stand, no decisive reason to doubt that very many commonsense belief/desire explanations are – literally – true. Which is just as well, because if commonsense intentional psychology really were to collapse, that would be, beyond comparison, the greatest intellectual catastrophe in the history of our species; if we are that wrong about the mind, than that's the worst we've ever been about anything. […] Nothing except our commonsense physics – our intuitive commitment to a world of observer-independent, middle-sized objects – comes as near our cognitive core as intentional explanation does[18].

After all, Newton's theory of gravitation explained both the orbit of the moon and the falling of common objects. A theory of mind is also a theory of the subject: common everyday individual subjects. These subjects should recognize themselves in the description proposed by this theory. A mental framework must be able to explain those everyday subjective facts that have been traditionally neglected by science. In the long run such a theory should come up with a convincing explanation of its dynamics[19].

Everyday experiences should be an explained without resorting to their advocated dissolution into the objective reports of the hard sciences.

4.7 Possible candidates

The above criteria can now be applied to several theories of mind that were proposed in the past. Clearly a convincing theory of mind should score positively with all of them. This is not a historical work so the candidates considered here have been chosen as pure examples of points of view. Besides, the theories in question neatly represent a precise range of conceptual possibilities. The list of candidates includes: Descartes' dualism of substance, Berkeley's idealism, Armstrong's pure physicalism, not reductionistic functionalism (Locke e Fodor), functional reductionism (Hume e Dennett)[20]. This list at could have been lengthened at will. Yet these five points of view are enough for a first overview. Theories not directly represented in this paragraph can be compared to the one we have chosen. For example, most of cognitive

[18] (Fodor 1987), p. xii

[19] (Di Francesco 1996), p. 18.

[20] Clearly, each of these authors is used in these paragraphs in a highly stereotyped way. This use is useful only to achieve a quick overview of the general panorama of failures related with the comprehension of the mind.

scientists could be classified as reductionistic functionalists; supporters of eliminativism would match well with pure physicalism. Behaviourists would oscillate between these two positions and so on.

The results of this comparison are shown in Table 4-2 that illustrates how no mentioned theory succeeds in solving all problems. Dualism, for example, is not compatible with empirical sciences; it would be very expensive from an ontological point of view. Pure physicalism fails on many points: the existence of subject, freedom, the ontology of subjective states, and the quality of phenomenal states. Idealism drains the physical world of value and it does not offer any plausible link between the physical structure and the spiritual life of a subject. No version of functionalism deals adequately with the quality of phenomenal states, with the very existence of subjects (reductionistic version), with the nature of phenomenal objects, with representation and freedom, and suffer of an excessive ontological prodigality (not reductionistic version).

	Substance dualism (Descartes)	Idealistic eliminativism (Berkeley)	Materialistic eliminativism (Armstrong)	not reductionistic functionalism (Locke-Fodor)	Reductionistic functionalism (Hume-Dennett)
Subjective entities	yes	yes	no	no	no
Aboutness	yes	yes	no	yes	no
Unity	yes	yes	no	no	no
Mental causation	yes	yes/no	yes	yes	no
Freedom	yes	yes	no	no	no
Objective entities	yes	yes/no	yes	yes	yes
Physical entities	yes	no	yes	yes	yes
Representation	yes	yes	no	no	no
Physical coherency	no	yes/no	yes	yes	yes
Empirical science compatibility	no	no	yes	yes	yes
Phenomenal quality	yes	yes	no	no	no

Table 4-2 A comparison among different theories of the subject.

Summary

What requirements must a theory of consciousness have in order to be able to distinguish between subjects and objects? Does such a theory have particular needs or can it be treated like any other scientific theory? Due to the *weltknot* posed by consciousness and due to the limitations of extensional ontology, we claim that a different criterion must be used. A theory regarding consciousness cannot limit itself to a narrow scope. The conscious mind is the point (both conceptual and real) in which reality knows and experiences itself. It is the point where what exists is identical to what is represented. That is why it is so difficult to solve it.

Here are proposed six criteria: ontological economy (Ockam's razor), direct experience, explicative power and predicting capability, experiential adequacy, empirical science compatibility, everyday experience compatibility.

We propose the following approach: first a revision of the fundamental ontological framework of reality, and an *a priori* conceptual reshaping of the fundamental categories; then a series of empirical experiments with the aim to verify the predictions made by such a theory.

The basic idea is that the objective framework metaphysically denied the existence of the subjective domain and that this was a consequence of the Cartesian scissor. It appears that reality has been arbitrarily split in two halves: it must regain its unity.

5 Intentionalizing nature

Almost all really new ideas have a certain aspect of foolishness when they are first produced
Alfred N. Whitehead[1]

In the last century, two approaches dominated the field of philosophy. On the one hand, there was the problem of «*what is there?*»: the old ontological problem. On the other there was the problem of «*how do we know that there is something?*»: the old gnoseological problem[2]. In this chapter we claim that it is possible to propose a unique principle capable of satisfying ontological and epistemological requirements. This principle will be termed i*ntentional relation* or *onphene*. This term refers to the elementary constituent of the fundamental domain. We have introduced this term to emphasise the connection between it and Brentano's intentionality. However, the two are not the same so we will try to highlight what the commonalties and the differences are. The onphene will be used to deal with the well-known mind-body problem in order to reach a foundation for both the subjective and the objective aspect of reality. Our main goal is to present a new ontology that will be the basis for a following *a posteriori* verification.

As Franz Brentano wrote in 1874, in one of the most famous as well as most criticized passages of philosophy:

> Every mental phenomenon is characterized by what the Scholastics of the Middle Ages called the intentional or mental in-existence of an object, and what we might call though not unambiguously, reference to a content, direction towards an object (which is not to be understood here as meaning a thing), or immanent objectivity. Every mental phenomenon includes something as object within itself; although they do not all do so in the same way. In presentation something is presented, in judgment something is affirmed or denied, in love loved, in hate hated, in desire desired and so on. This intentional in-existence is characteristic

[1] (Whitehead 1978).

[2] We use the old term *gnoseology* purposefully. Here we refer to the general problem of knowing something about the world and we do not want to use a term with a heavy burden of recent historical debates such as epistemology. Besides we use the term epistemology to denote only objective knowledge.

> exclusively of mental phenomena. No physical phenomenon exhibits anything like it. We can, therefore, define mental phenomena by saying that there are those phenomena which contain an object intentionally within themselves[3].

And indeed intentionality has become one of the main concerns of philosophers of mind. There is almost a general consensus that the capability of referring to external events is what makes mental events so irreducible to physical structures. Yet, as we have previously outlined, it seems impossible to find suitable candidates for such a role. We propose a different approach. Instead of trying to produce intentionality on top of a fundamentally not intentional reality, we suppose that the fundamental domain of reality is a kind of original intentionality (Figure 5-1). This move is radically different from other approaches that endeavoured to naturalize intentionality. In this thesis, intentionality is perceived as something fundamental and, as we will see, something that expresses the true nature of experience as well as the content of the experience.

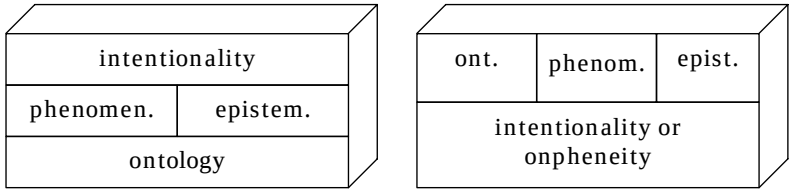


Figure 5-1 Traditionally many attempts have been made to reduce intentionality to other categories (physical reality, language, knowledge). On the contrary, our approach claims that is possible to reduce everything to intentionally which then becomes the fundamental ontological layer. This is a radical ontological revolution.

[3] (Brentano 1973), p. 88.

5.1 The principle of the conservation of meaning and experience

Before attempting to redefine the fundamental ontological framework, a general criterion must be stated. We will call it the *principle of the conservation of meaning and experience*. The two terms ('meaning' and 'experience') refer respectively to the two levels in which knowledge is usually divided: on one level there is meaning – seen as a conceptual high level epistemic way of referring to reality; on the other there is phenomenal experience (colours, pains, flavours)[4]. In brief, the principle says that

something that makes a difference must exist

The principle claims that, if there is a difference, any kind of difference, there must be a corresponding difference in the ontology of the world. As a result, it is impossible that something is only part of a thought – or of an experience or a perception – since everything, which *is*, must have a place in ontology. The idea that there are two dimensions or two domains – is derived from a dualistic ontology[5]. If such ontology is rejected there is no place for anything that is not embodied by something real.

The principle must be considered as an ontological criterion and not as an epistemic or causal one. The «something that makes a difference» must be made up of something and not merely caused by something. This distinction is important because the corresponding causal version of the principle would be less sustainable. The principle supposes that «what exists» must be explained by just one domain of entities: it is an implicit anti-dualist claim. It entails that if meaning and experience are something different from nothing (and they are, since we have a direct empirical contact with them), they must correspond to a difference in the ontology. Any change in our knowledge and our experience of reality entails a corresponding change in reality itself. Any change in the mind should entail a change in reality[6]. There is no ontological free lunch.

---

[4] The former produces the objective domain, the latter the subjective.

[5] In reality it derives from the more ancient distinction between essence and existence that was already well established in Aristotelian metaphysics.

[6] In this way the principle of conservation of meaning and experience recalls supervenience. According to the normal usage of the term 'supervenience', the mental domain is supervenient on the physical one. As we will see it is possible to propose a different solution according to which both the mental domain as well as the physical

If we had a theory that explained why knowledge and experience are a part of reality, how consciousness is part of reality, what the ontology of representations is, and what perceptions, phenomenal experience, objective knowledge are; there would be no more need to divide the world of knowledge from the world of «being here in the world». Being and becoming would merge into a compound principle.

5.2 Intentionality[7] as being, representation and being in relation-with

Given the aforementioned difficulties in finding a suitable place for representation as such, a different approach is proposed here. Its aim is to outline a different framework for representation and consciousness. This framework will eventually be tested on more empirical grounds both as a guideline for grounding representations in building a robot, and as an explanatory tool to provide insight into normal conscious experience. The first step proposes a simplification of the current ontological pattern. The persistent division – between the thing that is (the object) and the thing that represents it (such a difference can range historically from the neural pattern/external object paradigm to the Tomistic *esse in mente* / *esse in re* dichotomy passing through Descartes' dualism) – is due to the empirically unjustified belief in an autonomous domain of purely extensional entities. The proposed principle eliminates both the possibility of this domain and the difference between what is represented and what is representing. In the following, the consequences of this move are examined. Perhaps, in dealing with representations, we might have to deal with problems since two or more supposed separate concepts denote the same object.

A first issue concerns the possibility of finding suitable candidates for the role of Autonomous Representations (by relation or similarity)[8], in the physical domain. How to find a genuine *physical autonomous representation by both*

---

domain are supervenient on a further ontological dimension (§ 6.3). Clearly, any change in the physical domain should not entail a change in the mental domain. However, supervenience is different from the principle proposed since supervenience is just a structure of logical relations between different properties or entities. It does not say anything about their ontology.

[7] See Box 5-1.

[8] See § 3.3.

similarity and relation? This goal is obtained by equating representation with existence as suggested by the Principle of Conservation. The hypothesis is that both terms denote the same domain.

For example, let's pick up something that we can assert exists: a stone. Well, that stone is undoubtedly represents itself. It has all the properties of a stone. In this sense we can say that that stone is representing something. On the other hand all our conscious mental states are events. Reality is different, since we are conscious of something instead of something else. It follows that our conscious mental states are something that exists. They are real. Therefore they exist. Existence and representation cannot be split apart.

On the other hand, there is another aspect of reality that cannot be eliminated easily: being in relation with. If we know something, it is because there is relation between that something and us. Being conscious of some content, entails a relation. If something were completely destitute of any relation with the rest of reality, it could not exercise any effect on anything and it would be out of our reality. It would not exist. From an empirical point of view, we cannot say that the existence of something, which was lacking the property of being in relation with something else, has never been experienced. In order to exist, everything must be in relation with other entities. The same conceptual evolution can be recognized in the passage from Newton's classical physics to quantum mechanics. In the Newton's case, matter existed independently of any relation with the rest of reality. After all, matter was very similar to Descartes' *res extensa*. Planets and body were where they were and they did not depend on anything else except their own intrinsic capacity of existing. In quantum mechanics, it is impossible to speak of something that is outside the act of observation. The very act of observation can modify what is to be measured. While in quantum mechanics the relation between subjective and 'objective' measurement is still far from clear – that is the role of conscious mind in determining the structure of reality –, the progressive evolution towards the unification of existence and being in relation-with is manifest.

In the previous paragraphs, we have looked for a radical modification in the categories usually accepted to interpret reality, something that could be used as the basis for a rational explanation of the world, its representation and its objective knowledge. The elementary, but fundamental, conclusion that being is being in relation with has been reached. Reality must have at its roots something that can embody the essence of being and also being a relation. Besides, existence and representations seem to be equally inseparable.

Nevertheless there is still what Kant called the most difficult problem of philosophy: the problem of representation. How is it possible that something represent something different? In a purely extensional ontology, an unresolved

problem is how can an extension produce an intension? With a slogan, how is it possible *to create an intension using only extensions* without an ontological implicit expense? It is true that the concept of intension can be shown as a relation between entities, or as a function between possible worlds and references (i.e. extensions) but without the existence of minds, there is neither a meaningful way to justify the existence of intensions or meanings or content or concepts or qualia. If someone looks at a written word, it is clear that those graphic signs have no intrinsic representational capabilities but that their being signs is derived from choices, properties and existence of conscious observers. Of course, when we look at our mental representations of concepts or of external objects, we do not need another conscious being to know what the contents of our mental states are. Our mental states seem to have a sort of intrinsic representational capacity that has no place in an extensional ontology. It is as if, as we said in Chapter 0, the word 'fox' physically made of ink on this sheet of paper, would know, by itself, what it refers to. Obviously, it is not possible because such a semantic relation requires a conscious human being, its user. Removing conscious beings imply the removal of all representations. Apart from the problematic ontology of meanings and qualia, the problem is that without a conscious being, which is the carrier of the relation between a representation and its reference, there seems not to be any straightforward way to express the relation of representation between extensions.

To this apparently unsolvable conundrum, we propose a possible solution summarized in three statements that we claim are collectively true both empirically and theoretically. Their acceptance is the basis for a radically modified ontological framework. The idea is that these three separate aspects of reality – existence, representation, and being in relation-with – are just three separate roles of the same fundamental principle. They have been seen as separate concepts because they correspond to three different ways of looking at the same principle.

We can go one step towards our solution with the consideration that nothing can be said to be a separate instance of just one of these two aspects. *Nothing exists without being in relation with, nothing is in relation without existing, nothing represents without existing, nothing exists without representing, nothing represents without being in relation-with, nothing is in relation with without representing*. If we look at the world around us we can observe that each extension, each object, and each event represents only one thing, in an unproblematic way: itself. This almost obvious fact carries us to the first strong assumption: r*epresentation is existence*. Further and conversely, existence is representation. Existence is also *being in relation-with*, therefore representation is being in relation-with. In short we can state that (FT: Fundamental Thesis):

representation is existence,

existence is being in relation-with,

and being in relation-with is representation.

They are not three different ontological entities or three different properties of the same entity. They are three different roles for the same ontological entity. They seem different because, as it will become clearer, they depend on the structure of subjects and, therefore, on their epistemic attitude.

If these three aspects are never separated we can suppose that they originate from a unique principle that we will term *onphene* or *intentional relation*. Such principle is an elementary unity of being, of representation, and of being in relation with. We propose to term such an object *intentional relation* or *onphene*. The first name derives form the famous attribution of the intentional as the mark of the mental proposed by Franz Brentano[9]. The other term (*onphene*) is a compound of *ontos* + *phenomenon* + *episteme*[10]. As a working hypothesis the following is proposed:

An onphene (intentional relation) is the fundamental entity; it is the elementary unity of existence, representation and being in relation-with. (IRH)

Stated in such a way the awkward problem of representation seems to vanish. The reason is that we have moved from an object oriented extensional ontology to a point of view extensional-representational-relational that is oriented towards a new principle that is. Similarly, new openings are conceivable for the issue of conscious subjects. If before, the constitutive element of reality was an extension doomed to remain forever constrained by its own boundaries, it is now possible to build a complete subject where onphenes would naturally have the role of representation carriers. These statements have powerful and practical consequences that would be challenged later. If our mental states represent something, our mental states must be that thing. This is the fundamental principle on which we will build our theory of the mind.

[9] See note 7 in this chapter.

[10] The term *onphene* [a:nfi:n] is used here for the first time. It has been coined expressly to identify the proposed concept without any misguiding preconnotations. It is a countable noun. Its plural is *onphenes*. The domain to which it refers is *onpheneity*.

Starting from the onphene and the fundamental hypothesis, a series of related concepts need a new definition. Here it is the basis for two consequences used to explain the problem of representation and the problem of the existence of conscious subjects. They must be seen as links between the general principle of the onphene and the practical

Every representation must be the thing that is represented; representation is existence (REH)

It must be possible for something to becomes part of something else; existence is existence in relation-with (ERH)

Before entering into the rationale of the second hypothesis the consequences of the first must be analysed. If everything that is represented to a mind must be that or part of that mind, it follows that the mind is no longer constrained by the restricted boundaries of the skull but literally that it is enlarged. This intuition will eventually be developed into a larger framework termed the Enlarged Mind (Chapter 6). It is constituted by all those events that are part of his/her conscious experience, without having to resort to the dubious and expensive implicit ontology of non reductionistic functionalism or certain kinds of psycho-functionalism[11]. Such a view provides a ready made answer to what the mind is. The mind is the collection of events related to the subject. The problematic process of perception, that is the representation of external objects, can be easily explained. Instead of having to reproduce reality, it is enough to define the condition of the enlargement of the mind. When the mind perceives something, it is the mind that has expanded itself to the new event or, alternatively, the perceived event becomes part of the mind. Unfortunately, this argument rapidly collapses as the theory of sense datum did. Why should an event belong to a particular mind? What is the difference between an event belonging to a mind and one not belonging to it? This brings us to the next fundamental assumption. The elementary unit of being and representation must not be seen as a separate autonomous entity but in relation with other events (REH). Its being must not be seen as something different from its being in relation with (ERH).

[11] As an example of the latter it can be proposed the natural dualism of (Chalmers 1996). It is interesting (Sturgeon 1998) since it provides a complete overview of the related theories.

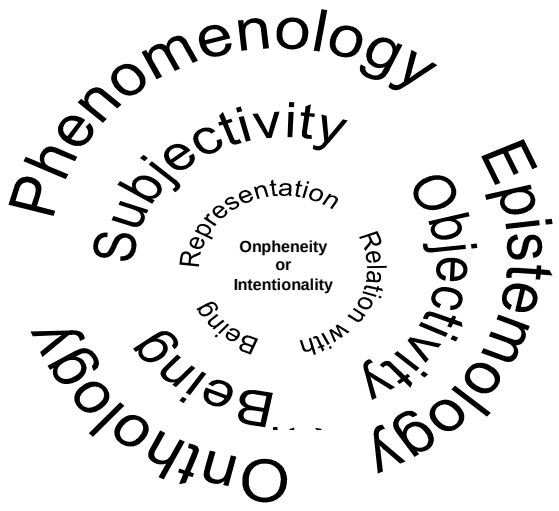


Figure 5-2 If intentionality or onpheneity is put at the centre of the ontology; all other manifestations of being can be derived directly.

Box 5-1 Intentionality and onpheneity.

Historically, intentionality is a term heavy loaded with different meanings. A *caveat* is thus mandatory. In this thesis the caveat refers purely to what is explicitly defined in § 5.2. In particular it neither refers to any kind of stance (Dennett 1987) nor to linguistic kinds of intentionality. It is very mush like the *aboutness*, about which Searle often talks (Searle 1985; Searle 1992). Given the fact that this paper is mainly devoted to explain what we believe an intentional relation is; we must also specify what it is not. It has nothing to do with Daniel Dennett's intentionality (Dennett 1987) which is useful to interpret the behaviour of agents. Our intentionality is something that could have some points in common with what John Searle means by intrinsic intentionality (Searle 1983; Searle 1992). Nevertheless, Searle's position seems to be looking at intentionality as an emergent property of biological brains while we are looking for a much more basic structure of reality. Another example can be found in some of Fodor's recent works (Fodor 1987; Fodor 1998) even if he does not speak directly of intentionality. Nonetheless, he is looking for an elementary property of the world that can be the basis for the representational activity in language. The basic idea is to see intentionality as if it were an ontological domain preceding the subject-

object dichotomy. Instead of being a property, an act, a function, or a relation that must be instantiated upon fixed (and maybe auto sufficient) ontological domains, intentionality is seen as the fundamental stuff of reality. In other words, instead of looking for intentionality as something that should be added to the already determined picture of the world (mass, charge, charm, and maybe some other basic forces), the idea is to see if it is possible to start from scratch with intentionality and then to build up the known elements of the world (mass, charge and *similia* but also pain, representation, consciousness and *similia*).

In this sense, we propose a new word for referring to intentionality as the fundamental domain: *onpheneity*. Yet, we also use the word intentionality, since it is the term that most closely matches with what we think onpheneity is. Intentionality, as it is normally conceived, is not an emergent new property of reality but it is the final expression of a fundamental structure. A useful metaphor is weight and mass. Weight is a complex physical property that depends on several factors (number of particles, density, and intensity of gravitational field), yet it is the final manifestation of a fundamental aspect of reality: mass. If there were no mass, there would be no weight too. However, it would be pointless to try to find an explanation for weight as an emergent property of complex systems. In this metaphor weight has the same role as intentionality whereas mass corresponds to onpheneity.

5.3 Events

One of the criteria to be satisfied is the capability of producing the familiar world of everyday experience. This criterion is not fulfilled by an extensional ontology. We claimed that the world is composed of onphenes. In the following we will show how an onphene produces events and how, in turn, events are the foundation of objects and subjects.

What is an event? Here the concept of event is somewhat different from the classic definition according to which an event is the instantiation of some property by some particulars at a certain instantaneous time *t*. An event is defined as *something without which the actual state of the universe would have been different.* No constraints are posed on what an event can be. For example, an event can be the modification of the energetic level of an electron or the declaration of war between two countries. It can be the melting of a glacier which lasts for several millenniums or the blink of an eye. An event can be something physical like a stone falling or subjective like a sharp pain (not its

cause but the phenomenal pain itself). In short, an event has no constraints either spatial or temporal. An event can be subjective or objective or simply physical. An event is just something without which reality would have been different. Obviously, if something had no effects on reality, it would not even exist. In this sense, there is a strong relationship between an event and simply a thing (an *ens)*, which is something that simply exists. Being an event is a condition of existence. It is impossible to be, in any meaningful sense, without being an event too. It is impossible to divide the property of being from the property of being an event, which is something that provokes effects on the rest of the world. Things are thus derived from events. Further, these last considerations take us to the natural conclusion that *being is being in relation with*.

In short, we propose that

an event is everything whose absence would make a difference to what reality is
(EH)

It is important to stress that such a definition is independent of most ontological commitments. It is independent of the kind of difference that it constitutes[12]. Besides, this definition captures the essence of REH as well as ERH. REH states that any representation must be an event in order to be different from the null event. ERH states that in order to be an event, it must be in virtue of its being in relation with something (it must make a difference). Furthermore, that being a relation and being an event must be seen as two sides of the same object. There cannot be an event by itself, an object in itself. The classical definition of noumenic object is thus rejected at its very roots. If ERH is accepted, there are no more things in themselves. On the other hand, if REH is accepted, the two separate domains for mental entities (or symbolic construction or interpretation or whatever) and for reality in itself no longer exist.

An event is therefore *similar* to what is traditionally seen as the causal relation that links two separate events. The difference is that the event is not a pure relation but is embodied in the occurring onphenes. Once one accepts that the relation between events carries the ontological weight, the need to suppose the existence of a class of substances disappears. The other immediate logical

[12] For example, it does not entail that the difference should be a physical or a mental or a subjective or an objective one. In this sense it is possible to claim that a definition like (EH) is independent of the subjective/objective dichotomy and, as such, it can be used to set a more fundamental framework.

effect of this rationale is that each relation between events must necessarily be seen both as an event in itself and as a unity. The first consequence derives from the fact the every relation is, from every conceivable point of view, something which has happened and that can be considered as an event (this is little more than a tautology of ERH). The second consequence comes from the empirical and observed fact that the world contains unities and that such unities must be substantiated at a non-reductionistic level.

5.4 Causation

An intentional relation is something that is, simultaneuosly, a condition for existence and a condition for knowledge. Can we somehow locate an intentional relation in our everyday experience of reality? This is indeed possible, albeit with a few distinctions in the way we do it. We will enter into the details of the epistemic processes in § 6.3. Here we want to deal with the similarities with regards to the more familiar causal relation.

An onphene leaves something in the world. An onphene or intentional relation leaves a causal relationship between two events. In other words, causation is the perceived intentional relation between two basic events. Causation should not be intended as the controversial relationship binding two events in general but simply as the *a posteriori* necessary relationship that can be logically assumed to have existed between them. After two events have taken place, if the second was somehow dependent on the first, it is not arbitrary to say that if the first one had not existed the second one would have not existed either. As Hume stated, we will never be completely sure that events of the same kind as the first event can lead to events of the same kind as the second one in the future. Yet, a causal relation can be supposed *in the past* (supposed not ascertained). Even if a causal relation cannot be detected it seems uncontroversial that, given whatever set of events, there had been other events, in the past, that were conditional to the first set. In short the causal relation, as defined above, is the *fossil* of the intentional relation. It states that something has linked two otherwise separate events. It is possible, however, to observe the provoked effects of that relation or the intentional relation as if it was an event. Causal relations are what have linked two events and it does not matter if they are perceived as such. Two remarks are possible. The first is that it is possible to look at an intentional relation from the outside and from the inside. In the former case, what is perceived is just a causal relation between two events, each one perceived along a separate intentional pathway leading from it to our inner mental states. In the latter case, the world is perceived as it is in our usual

manner. The subject is literally the intentional relation and what the subject perceives is the content carried by it. The second consideration is that it is not possible to observe an intentional relation directly except in the most intimate possible way: by being that intentional relation. Intentional relations carry contents and there is no need to suppose any other explanatory framework.

A first consequence of this argument is the fact that, if an event were completely deprived of intentional relations with other events, it would literally disappear from the universe. Or better, it would not disappear because, in that case, it would have some effects on some parts of reality (for example the effects of his disappearance). This argument is similar to Armstrong's causal relevancy criterion[13]. This apparently obvious result can be used in order to state a not so obvious remark: that every event has to be connected with other events and that no event exists by itself. The being of an object resides in its relational connections with the rest of the world. The idea of atoms closed like monads looses its apparent simplicity.

This brings us to the logical conclusion that onphenes or intentional relations alone are sufficient to explain the world without having to resort to events. Events can be described in terms of intentional relations only and, as such, they are considered as nothing more than explanatory means. If the previous rationale can be accepted then normal objects can be seen as useful simplifications of more complicated sets of events, while events themselves derive from intentional relations.

It is tempting to remove the logical necessity of something among causal relations. It is possible to imagine a world of pure causal relations. Such a world is attractive but presents several problems of its own. It completely lacks the ability of producing content as well as quality. In a world of pure causal relations it is impossible to explain where meaning and quality originate from. In this world everything would be reduced to the lowest ontological level of entities manifested by a pure extensional world. In the following, we will show how it is possible to describe a world of pure intentional relations.

Causation is an obvious candidate for the intentional relation. Unfortunately the causal relation (at least in its modern formulation: a law stipulating occurrence of phenomena) has been subjected to two main streams of criticism. The first is of the famous analysis Hume gave to the classical principle of causality. The second derives from a more recent conception of the world that seems to remove any need of using causation as an elementary component of

[13] (Armstrong 1988; Oliver 1996).

reality[14]. To overcome in one single step both sources of criticism, a different definition of causation is here proposed. Instead of trying to define causation as a type theory, something that must be valid for types, a token theory of causation will be proposed. In short, it is possible to say that *there has been a causal relation between two events in which one of the two would not have been without the first one*[15]. Causation becomes the unavoidable legacy of the past to the present. Besides, there are no constraints on the number of involved events. An event can be dependent on more than just one other event. For example, event A can be dependent on event B and also an event C. In our usage of the term the previous sentence means only that A could not have existed without the existence of B and C. There is no limitation by principle on the number of the antecedent events. In short, for each event occurring somewhere and sometimes, there is a set of past events without which the original event would not have existed[16].

---

[14] It is interesting to note that the bulk of criticism has been directed towards causation as a law connecting occurrence phenomena and not towards causation as a cause giving the essence or being the internal principle of its effect. Curiously, it happened when Descartes and its contemporaries introduced mechanicism as a universal explanatory framework for nature. Causation, stripped of its ontological prerogatives, would not be adequately endorsed by natural laws. The presumed inutility of causation in modern physics related more with this second kind of causation than the former one. The question if Aristotle's theory of causation is really unused in physics should be, as far as we know, further investigated.

[15] This definition is similar to the classical definition of causation as the process by which an entity (the cause) determines the existence of another entity (the effect). It is possible to compare our definition with S.Thomas "Quod potest esse et non esse indiget aliquo agente ad hoc quod sit, sine quo remanet non ens (Everything that could be or not be, depends, for its existence, by another entity, without which it could not be). *Comm. Ad Rom.* C. 2, 1. I. Or "Ex hoc quod aliquid est ens per partecipationem sequitur quod sit causatum ab alio" (Everything that exists, it exists because it receives its existence from something else). In short, in medieval philosophy the concept of cause is reduced to the concept of *ens*, which gives the principle of existence to another *ens*, as well as the concept of effect is reduced to the concept of *ens*, that receives such a principle from another *ens*.

[16] By admitting that such a set of past events could be empty there is space also for microphysical quanto-mechanical events with no legacy with the past. Let's admit that a couple of particles and anti-particles could spring out of nothing. That event would simply be an event with an empty set of antecedents. Nevertheless the fact that it is

This is a token version of causation because it does not imply any future repetition of the same series of events based on their kind or type. It does not deny repetition either. In other words, causation as it has been stated does not require that *ceteris paribus* whenever B and C occur then A occurs too (using the previous example). Of course, this is a possibility. This definition is independent of recognition or of other epistemic restrictions. There is no need to suppose that a causal relation, as defined, must be known or knowable. This definition is not easily refutable. Refusing this kind of causal relation *a posteriori* entails that every state of the world has no relation with previous states, that each instant of life of the world is completely random, a new creation. Everything would be independent of everything else and every correlation or dependence among events would be only a fortuitous coincidence. Rejecting this absurd conclusion compels to accept the proposed weak definition of causal relation. If such a definition is accepted then every causal relation is known only in two ways: either by being that causal relation or by looking at the event involved. A few words must be spent to clarify this conclusion.

If events and their intentional relations are the only components of the world, there must always be either another event or an intentional relation between two events. There are no other alternatives. As it has been argued, an intentional relation is an event in the sense that its existence must have provoked some effects. If this were not the case the intentional relation would be, from all points of view, not existent. If no other events were dependent on it its existence or its inexistence, it would be the same. An intentional relation must be an event. If an event is dependent on another event and between them there has been a causal relation, this link must be embodied by the event that constitutes their relation. Such an event is an intentional relation and the causal link between the twos at the same time. If we must embody the relation between two events then also every epistemic relation between a subject and its object must be embodied. If there is knowledge of something it is necessary for such knowledge to be embodied in an appropriate intentional relation, which in turn is an event in itself. Every relation is therefore dissolved into an event. Every relation among objects is an event. Causation must be some kind of event too. A causal relation is the event constituted by the dependence on two other events. Causation is dependency. As has been argued previously, representation is existence, and relation is also existence.

When something is observed and there is a kind of epistemic relation there is an event that is the observed object (that is the relation in itself). In order to

---

possible for some event to have no relation with the past does not entail that *all* events happen independently of the past.

know of its bare existence, to represent it, there has to be another event that constitutes the intentional relation between the observed object and the subject. If there was only the observed event and the subject no knowledge of it could be possible. The 'representational' event must carry in itself the meaning of the observed event, but it is clearly a different event from the original one. If a subject is looking at an object, he is able to perceive the meaning of such an original event because its subject is made up of the event corresponding to that relation too. If a second observer were looking at the events corresponding to the first subject, perceiving her/his object, it is obvious that the meaning of her/his experience would be different. The difference perceived by the two observers derives from the fact that the 'representation relation' is an intentional relation, which is, in turn, an event. The mysterious relation between the subject and the object, the gap between them, becomes a simple identity among events.

As it has been anticipated there are two ways of knowing something. The first way is direct and must be certain in Cartesian fashion. It corresponds to being made up of an intentional relation whose meaning is directly accessible. Its meaning becomes part of the subject. In this way, we are not conscious of the relation in itself as something that has connected two separate events (the external object and the internal mental event). We are identical to the carried meaning and such a meaning is part of our conscious experience. It is certain in the sense described by Descartes or alternatively by Gilbert Ryle with the term 'incorrigibility'[17]. This certainty derives from the identity between every representation and its true being. I can be dubious about my perceptions but not of my having those perceptions. I can feel a pain in my left leg and I can be wrong about the physical state of my leg but there can be no better judge than I on my having that pain. Conversely I can look at the world and I can 'infer' the relation among events. Hume claimed that such a relation will never be known with absolute certainty. Yet, it can be part of our experience, and indeed human beings have good reasons to be philogenetically oriented to, to observe relations between events. Such a second order form of observation of causal relation, which is relative and not absolute, is what is usually called objective observation. The uncertainty of this second method derives from the fact that when a particular causal relation does not constitute someone, he/she can know it only by means of following the effects of that relation to other events. There is no means whereby to recover the original event/relation. The classical objective causal relation observed in the physical world is similar to the ashes of

[17] (Descartes 1641; Ryle 1984).

the 'living' original relation. We can thus distinguish between the original relation what we have called intentional relation or onphene. It follows that

> *the causal relation in the objective world is the* fossil *of the intentional relation*

Two things must be emphasised at the end of this paragraph. First, causation is the result of the objective epistemic attitude regarding a more fundamental structure of reality that is intentionality. Secondarily, notwithstanding the potential epistemic barrier in perceiving causation or dependenceness between events, causation can be supposed as something that is part of the fundamental structure of reality. Given this hypothesis, it is possible to build a system that exploits this structure in order to embody causal relations and consequently intentional relations between events. Meaning is neither produced by these systems nor is an emergent property of them, but it is identical to the event that constitutes reality. Nonetheless a system can endorse more and more complex intentional structures capable of carrying an increasing meaning due to the fact that they are more and more complex events (always in token terms).

5.5 Principle of unification

e pluribus unum

One of the main limits of most of reductionistic ontologies is the incapability of reconstituting reality after it has been split by reductive explanation. For example, let's take an atomistic extensional ontology (modern or classic). If everything is reducible to a collection of atoms – and every atom is closed in itself – how is it possible to put them together and obtain everyday-life macroscopic objects? There is no space left in this ontology for unity as a whole. Any collection is just a mereologic juxtaposition of smaller entities. Every ontology that does not possess a fundamental way of proceeding from parts to wholes is doomed to what we have defined as the *reductionistic collapse* (§ 2.1).

Once it has been shattered, the unity of reality cannot be recomposed. The ontology based on onphenes does not have this problem. Every onphene is an atomic unit since it cannot be further divided. Inasmuch every onphene is a synthesis of a multiplicity of previous onphenes (there is no boundaries concerning the number or onphenes that can concur in determining a particular onphene). In normal usage multiplicity and unity are considered fundamental categories of reality. Yet, using onphenes, it is possible to reshape

epistemic and phenomenal domains, as well as the ontological one. Such categories should not exist *before* the birth of the subject. Unity and multiplicity should result from the composition of being, representation and being in relation-with. If we think outside the traditional categories, we can observe that unity and multiplicity are irrelevant at the level of intentionality. They are still undistinguishable.

A practical example shows how unity and multiplicity are not needed. Let's think of a single visual phenomenal experience, of a phenomenal experience that is what it is like to be ourselves when we are watching a certain scene. Let's think of it *before* we start analysing it by applying our esthetical and epistemic categories. Could we say that that experience is only a unity? Certainly not, since it is composed of several objects, colours, shades, all at the same time in our visual field. Could we say that that experience is a multiplicity of separated elements? Certainly not, since before analysing it, the scene we have in front of us represents a whole, experienced as such and not as a collection of parts. In everyday life we continuously have experiences whose content is at the same time 'one and many'.

Every onphene has a critical event as content. Every onphene unifies a collection of previous onphenes in a way that was unconceivable for traditional reductionistic objectivistic metaphysics. Let's analyse this point.

Every event is determined by a collection of antecedents. These antecedents constitute the content of the corresponding onphene that is the support of a subsequent event. What exactly is the unification? Take the characters on a printed page. The position of a character is determined by a large collection of previous events. As a consequence, this collection has had, , only one event. If we look at the world as a set of static structures, of material parts, we cannot understand why there are further entities that represent reality in itself (like concepts, percepts, ideas, phenomenal entities, qualia). Besides, these unify reality in wholes: they cannot without referring to conscious subjects (§ 2.2, 0). On the contrary, every event is the product of a collection of previous events – hadn't those events existed, the final event would not have existed either. These events are *unified* in the final event that depends on them for its being. It is possible to use the classical concept of formal cause (as the cause that gives an entity its being) to identify the critical event of an onphene. This is just a historical metaphor. The critical event is the content and, thus, the shape of what is represented within the onphene. The critical event becomes the formal cause or the essence of what is determined by it. Each onphene gets its own being from its critical event. As in Aristotelian metaphysics, the universal was the unifying element of reality, in our approach every event (and thus every

onphene) provokes an instantaneous unification of a portion or reality from which it necessarily gets its being.

The principle of unification is the constitution of a new meaning starting from a collection of separated events that collectively become the critical event of a new onphene.

> *Every onphene unifies reality becoming the bearer of the being of all its antecedents*[18].

The previous statement is the principle of unification of reality that highlights the fact that every onphene corresponds to the constitution of a new meaning starting from a collection of separate events.

5.6 Notation to express onphenes

The following notation is proposed.

$$R(x,y)$$

where R represents the intentional relation (or onphene), x its content and y its links towards other intentional relations. x and y are not necessarily singular terms, while R must be so. Besides y is introduced to mean the projecting of an onphene towards other onphenes. A more concise notation could be

$$R(x)$$

In this way there should be only a reference to its content. Why should this notation be used? $R()$, without the corresponding x that constitutes its substance does not have any meaningful existence. Yet, conceptually, it is useful to distinguish between the onphene (as a relation) and its content. The functional notation could be misleading. If this was the case a notation like R_x could be better. Hitherto we have found this last option unnecessary and a little awkward.

An onphene could be the content of another onphene. There is a linked flow of intentional relations, described by the proposed notation. Sometimes an

[18] Its antecedents are the equivalent of its formal cause.

intentional relation (or onphene) is the content of another onphene *as a relation*. Every onphene has its own content that is both what it is and what it represents; such content is transmitted to other onphenes.

Yet it is this passage from one onphene to another that can make up the distinct content of the first onphene. In a certain sense, every onphene can contribute to two different kinds of content: *as being* and *as being in relation-with*.

For example, if $R_1(x_1, x_2) \rightarrow R_2(x_2, x_3)$ (R_1 determines R_2) the content of R_2 must be equal to the content of R_1. Yet the fact that $R_1(x_1, x_2) \rightarrow R_2(x_2, x_3)$ is a content (different from the content of R_1 and of R_2). As such it can be the content of another intentional relation $R_3(R_1(x_1, x_2) \rightarrow R_2(x_2, x_3), x_4)$ that is completely different from an onphene R_4 that has been produced by R_1 and R_2 through the flow: $R_1(x_1, x_2) \rightarrow R_2(x_2, x_3) \rightarrow R_4(x_3, x_5)$. The content of R_4 is $x_3=x_2=x_1$ while the content of R_3 is $(R_1 \rightarrow R_2)$, which is different from x_3[19].

In the following figure – and in the final table – the principal combinations of the onphenes are showed together with the corresponding notation.

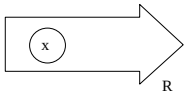


Figure 5-3 The arrow represents an onphene R. This symbol (the arrow) shows the relational nature of the onphene, its projecting forward. The circle at the beginning of the arrow shows the content x (what the onphene is and what it represents). The point of the arrow goes towards what is the target of the onphene: another onphene.

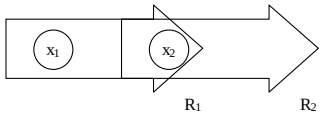


Figure 5-4 No intentional relation exists by itself like a monad or like an atomic extensional entity. Every onphene exists since it constitutes the content of another one. Onphene R_1 constitutes the content x_2 of the onphene R_2. Similarly the content x_1 will have been constituted by a previous onphene. This kind of relation can be described by $R_1 \rightarrow R_2$.

[19] As will eventually be defined, R_3 is a second order onphene, while R_4 is a first order onphene.

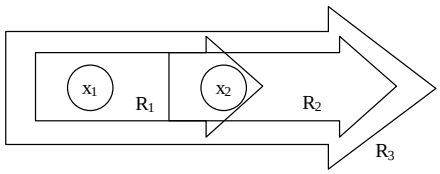


Figure 5-5 A flow of linked onphenes can be substituted by just one onphene whose content is the same as the content of the first onphene in the flow. A simple union of all the onphenes constitutes the final onphene. R_1 constitutes x_2, R_3 constitutes (as well as R_2) a subsequent onphene (not explicitly showed in this figure). The content of R_3 *is* x_1. The proposed notation is $R_{3=}\ R_1 \oplus R_2$.

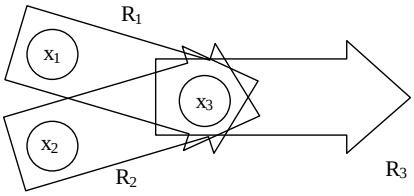


Figure 5-6 The content of the intentional relation R_3 is given by the onphene R_1 e R_2, which are responsible for its being. They can be more than two. In this case, they constitute the content x_3 of R_3. The point of the arrow shows the production of a new content. A unification of two separated parts of reality has occurred. This kind of relations can be described by $R_3 = R_1 \circledast R_2$.

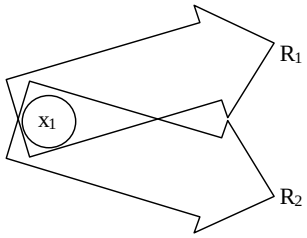


Figure 5-7 Many intentional relations (R_1 e R_2) can have the same content without having to be the same intentional relation. This happens since they make up different events.

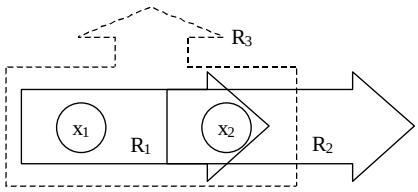


Figure 5-8 An onphene R_3 can have as content a relation between onphenes $(R_2 \rightarrow R_1)$[20]. In this figure $R_3=R_3(R_1(x_1, x_2)\rightarrow R_2(x_2,),)$. Content x_3 of R_3 is $(R_2 \rightarrow R_1)$.

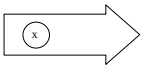
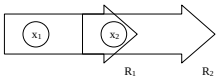
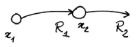
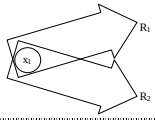
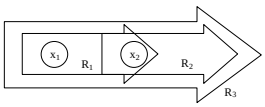
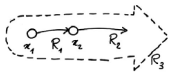
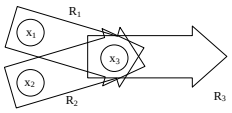
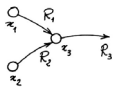
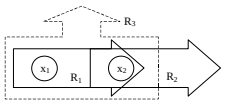
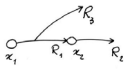
Letteral notation	Graphic notation	Sintentic notation
$R(x,), R(x), R_x$		
$R_1 \rightarrow R_2$ $R_1(x_1,x_2)$ $R_2(x_2,)$		
$R_1(x_1,)$ $R_2(x_1,)$		
$R_3=R_1 \oplus R_2$ $R_1(x_1,x_2)$ $R_2(x_2,)$ $R_3(x_1,)$		
$R_3=R_1 \circledast R_2$ $R_3(x_3,)$ $R_2(x_2,x_3)$ $R_1(x_1,x_3)$		
$R_3(R_1 \rightarrow R_2,)$ $R_3(R_1(x_1,x_2)\rightarrow R_2(x_2,),)$ $R_1(x_1,x_2)$ $R_2(x_2,)$		

Table 5-1 Summary table of onphene notation. On the rightmost column an alternative graphic notation is shown.

[20] An onphene, as a relation with another onphene, makes up a different content from what it is and what it represents.

5.7 Critical event

One further question has to be addressed here. Why should an event be dependent on another particular event (or a limited series of events) and not on a potentially infinite series of previous events? The answer is that for each event, there must have been a critical event in the past that uniquely determined such event. The content of each representation is this a critical event. For example, let's imagine we are looking at a red pen. An antecedent event of our mental state is the light that is departing from it. We are conscious of such content. Nevertheless the red pen is there, on the table in front of us and because of a complex series of events (the pen has been built, it has been carried where it is now, is has been coloured, and so on). Why are not we conscious of such a chain of previous causes? The reason lies in the nature of the critical event.

> *The critical event* e_c *of an event* e *is that event without which e would not have existed: the event* e_c *that is, at the same time, sufficient and necessary for the occurrence of* e. *(CEH)*

An example will help. Take a photoreceptor. Activation of photoreceptors is causally dependent on light events. Of course those light events, in turn, can be triggered on by other events that are not perceived directly as the content of the light event. For example, someone can switch on a light bulb but that antecedent event is not perceived in the light that is diffused from the bulb. Besides, the light events trigger other events along the way towards our visual cortex that are not perceived. Another example could be the chemical activity on the retina. The content of someone's mental state does not result from the visual experience of the chemical activity occurring in her/his retina. Figure 5-9 provides a practical example. C is some kind of light event. For example, C is the light reflected by a red pen in front of Elisabeth, a young student. B is the chemical activity induced in her retina and A is the brain activity occurring when her is conscious of the shape and the colour of the pen. D and E are two events that should have happened in order to permit to Elisabeth to perceive that pen. D is the physical transportation of the pen on the table and E is the painting of the pen in its factory. However D and E should necessarily have happened for C to occur. Nevertheless D and E were not enough. It might have been possible for the room to have been completely dark or for some other object to have hidden the red pen from Elisabeth. D and E were necessary but not sufficient events. On the other hand B was sufficient but not necessary in the sense that it could not have occurred without C. B caused A, and C caused B

but after C occurred B must occur too. In this sense B was not necessary but only sufficient. In other words, A occurs whenever C occurs or, more precisely, A occurred mostly when C occurred. Here one important issue is at a stake. Causation must be intended as a token causation *a posteriori*.

Every mental state must have its content. Each content must be real since it cannot be a creation of a not existent mental domain. Representation is existence, therefore a mind, an enlarged mind must be made up of all those events that have as content those corresponding critical events. The critical event is the content of each intentional relation. Given a critical event it is possible to know what the content of a given onphene is.

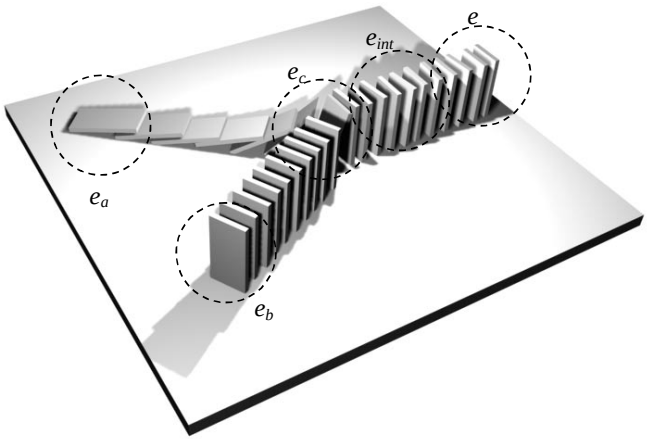


Figure 5-9 e_a, e_b, e_c, e_{int}, and e represents five possible events in mutual relation. In the example, e_a, e_b, e_c, e_{int}, and e are near the domino that by falling will provoke the next event. Let's suppose that the only starting events could be the falling of the dominos (event e_a or event e_b). What is it possible to know given the knowledge of the position of dominos and the occurrence of e (the falling of the last domino)?

5.8 The library and the onphene

There is a strong similarity between reality as a collection of onphenes and human culture as a collection of books and papers. Take a book or a scientific paper. There is a set of references at the end of. Any author is morally and professionally compelled to point at those works, previous to his/her own, that have been relevant. It they have made a difference into his/her work, they must be listed. Reference at the end of a book shows what previous books and papers have been *critical* to the development of a subsequent work. They are like arrows that point at those opera that constitutes the meaning of a new work. The metaphor with the onphene is clear.

Every onphene corresponds to a book or a paper. Big onphenes are books while smaller onphenes are papers. The onphene, which has determined the being of a particular onphene, corresponds to the references. Any book, even if it derives its content from previous works (*nanos gigantium humeris insidentes*), adds something new. It is something different from its predecessors (it is a new event).

If we looked at the flowing of books as a chain of works mutually linked together when, step by step, enlarge the domain of human knowledge domain, we would be contemplating a faithful image of the intentional flow that (even if autonomously) creates reality all together. Another similarity of this metaphor is the unifying role carried out by books as well as by onphenes. As every onphene unifies a piece of reality, every book unifies its point of view about a certain matter from a unique perspective.

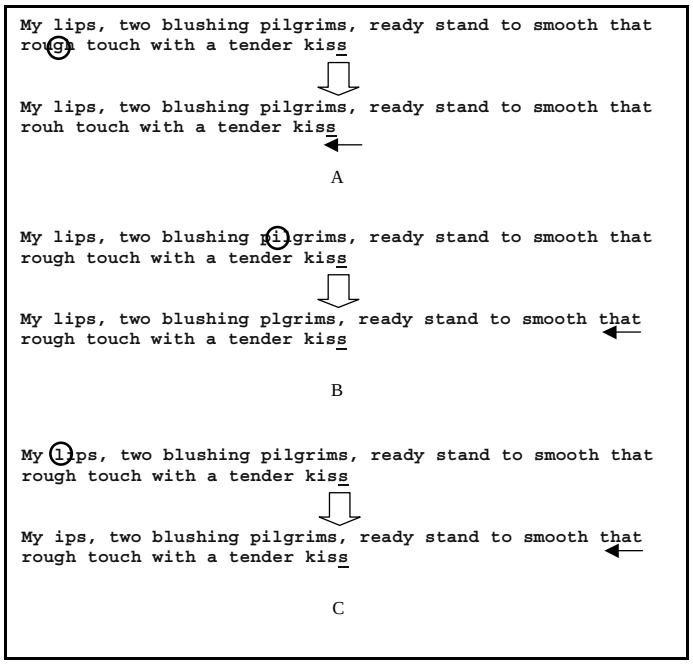


Figure 5-10 A passage from Shakespeare's 'Romeo and Juliet', provides a practical example. In case A, the elimination of the letter provokes the shifting for the word *kiss* (underlined). In case B, the elimination of the letter *i* provokes the shifting of the word *that* instead of the word *kiss*. In case C, the elimination of the letter *l* provokes the shifting of the word *that*. In the last two cases, the final event is the same although the critical event is different.

Summary

Everything that exists must have an ontological foundation. It must also be true for the content of our experience. Our having certain contents as a conscious experience is not the same as its opposite. Therefore we state a principle of conservation of experience and meaning. According to this principle there must be an ontological foundation both for the phenomenal domain and the epistemic domain. They cannot be left in a mental limbo. As a result of these principles we state our fundamental claim: at the bottom of reality representation, being and being in relation-with are indistinguishable. We propose a new entity that we have called *onphene* that is reality before any further specification. By using such entity it is possible to produce both the objective extensional world and the subjective domain. We claim that onphenes match more closely with what our direct experience is than the extensional entities that are just abstract entities that nobody really has ever experienced.

The fundamental thesis runs as follows: representation is being, being is being in relation with, and being in relation with is representation. Such a union is called an onphene. In one go it solves the problem of representation and the problem of unity that have been refuted by reductionistic ontologies.

The properties of the onphene closely match the property of what has been traditionally attributed to mental states: intentionality. For these reasons an alternative term for onphene is intentional relation. Yet this could be confusing. Intentionality is at the bottom of reality and we experience it, in our mental states, for the simple reason that our mental states like all reality, originate from such fundamental domain. When intentionality is considered in such a fundamental role the term *onpheneity* could be used.

The onphene can generate all known aspects of reality: extensional entities, subjective phenomenal experience, objective empirical knowledge, *a priori* meanings, static entities like objects. Events are as onphenes qualify only in their role of being.

For every onphene there is a particular event (or a group of events) that constitutes the content of that onphene. Such an event is called the *critical event of an onphene*.

6 The Enlarged Mind (TEM)

> SHYLOCK *Hath not a Jew eyes? Hath not a Jew hands, organs, dimensions, senses, affections, passions, fed with the same food, hurt with the same weapons, subject to the same diseases, healed by the same means, warmed and cooled by the same winter and summer, as a Christian is? If you prick us, do we not bleed? If you tickle us, do we not laugh? If you poison us, do we not die? And if you wrong us, shall we not revenge? If we are like you in the rest, we will resemble you in that.*
>
> William Shakespeare[1]

Until now we have spent a lot of effort on defining a general ontological framework capable of justifying both the objective material domain and the subjective phenomenal domain. We have avoided founding anything on the existence of subjects. Our claim is that onphenes, as defined, own an *a priori* coherency that is independent of the existence of conscious subjects. They do not suffer of the same logical circularity highlighted in the case of objects, information, static and dynamic systems. Of course such a proposal must undergo also an *a posteriori* verification. The proposed ontology must be capable of finding a suitable place both for the extensional entities of objective reality and for the subjective phenomenal entities of subjective experience. It must be capable of explaining the relation between the *wholes* of our conscious experience and the *parts* of the physical world. All the requirements – needed for a theory of mind – must be satisfied (Chapter 0).

In this chapter, the ontology based on the domain of intentionality (of onphenes) is used to propose a theory of mind that eliminates any structural diversity between the phenomenal world and the physical one. Its aim is to solve the *weltknot* not by finding an impossible reunion between the two sides of Descartes' division, but by defining reality in a way that makes such a division useless.

---

[1] The Merchant of Venice, Act III, Scene I.

6.1 Constitutive theory of the subject: Theory of the Enlarged Mind (TEM)

If we accept the idea that representation is not a high level by-product of a conscious being, but just one of the elementary facts of the world, we have still to explain why there are subjects. In other words, why is the world not just a giant flux of events, each one always perfectly, uniquely representing itself and only itself? Or, conversely, why is it not a unified representational event without boundaries? To cope with the empirical experience of conscious subjects floating inside a material flow of events and objects, a constitutive theory of the subject is needed. It must explain why some intentional relations are grouped together to form a particular and separate subject. This theory must be capable of explaining the properties of our experience of the world (subjective experience vs. objective knowledge).

We will start from our definition of what a mind is. A definition that is empirically compatible with experience and empirical data. *A mind is a collection of representations, which are intentional relations, unified by the internal principle of a particular self*. What the *principle of self* is will become clearer later. For now, it is enough to denote something that can glue together a set of intentional relations. A subjects is a unity, a part of reality which becomes unified and that, in turn, is able to unify other part of reality (objects and events). By using onphenes we can translate the definition by stating that a mind is a collection of intentional relations, each one identical to itself and carrying its own personal content. There are no other problems of representations as there were in the traditional framework where the mind is derived from an extensional piece of the world (the brain) incomprehensibly representing other pieces of the world (the external objects). Here the mind is enlarged to contain all the intentional relations that constitute it[2]. Because of this, we term our conception of mind as *enlarged mind* (the only kind of mind, in reality). Those intentional relations, which are the content of the corresponding subject, constitute a mind. There is no difference between a subject and its mind. Of course, this theory is patently not a cognitive theory of consciousness. The relation between cognition, consciousness and representation is only a nomologic fact, a contingent fact.

[2] David Chalmers and Andy Clarks use a similar term but with a consistently different meaning (Chalmers and Clark 1999). In their paper, the authors claim that the structures belonging to a single mind can be extended to all those devices that take part in its activity, like computers, calculator machines, address books.

Cognition is the practical[3] step evolution had to make in order to be capable of producing complex subjects. From a logical point of view, between cognition and conscious experience there is the same kind of relation that exists between wings and flight. While it is extremely difficult to achieve the latter without the former, the relation is only a contingent fact that depends on several nomological factors. Even nowadays flying without wings, albeit for some strange vehicles such as helicopters, is extremely difficult. Nevertheless, wings remain a contingent mean to achieve an independent goal: flight.

If a mind is constituted by a set of intentional relations, it is clear why other observers cannot experience its private content in the same way. Each observer (or subject) will experience his own intentional relations that are necessarily different from other subjects'. It is also clear why no object will ever be an acceptable place for a mind. An object is a by-product of events. Events are by-products of intentional relations. Since intentional relations constitute it, the mind cannot be an object or a state of an object. The same elementary relations that produce objects and events constitute the mind. The knowledge process is part of the necessary ontology of reality.

If minds are intentional relations, what are brains then? A brain is the physical object that is perceived to be the point in which the onphenes belonging to the subjects find their unity. Of course, when a brain is perceived as an object, it is not the same brain that part of the unification process of the subject. We do not perceive our own brain directly during introspection: we perceive the events that are unified by the onphenes ending in our brain. Even if a surgeon could look at his/her own brain while introspecting, the brain that he/she would be seeing would be just an object. Intentional relations are the basic stuff of reality and, as such, they are the origin both of the physical world and of the subjective experience of it. Brains and other physical objects are the final result of the interaction of onphenes. An amazing fact about the external world is that it cannot be comprehended wholly without recognizing its relations with the world of subjective experience. In the same way that it is not possible to reduce all aspects of reality to a purely subjective dimension, it is not possible to reduce it to a purely extensional one. One common mistake is to confuse the anti-metaphysical attitude of empiricism with an extensional metaphysics. The foundational attempt presented here does agree with the former but it firmly rejects the latter as empirically unsound. Pure extensional entities are something that is out of the reach of our empirical experience, while

[3] Perhaps this step is nomologically necessary.

empiricism is the attempt to accept entities that can be experienced[4]. An intentional framework is the most empirical framework conceivable.

If the mind is a collection of intentional relations, why should these onphenes be unified into a unique experience? And if representations are intentional relations and these are, in turn, events; why should an event belong to one subject instead of another? Indeed there are two opposed visions of the world. In the first, the old classic materialistic framework, reality is composed of a multitude of *entia* (atoms, molecules, physical or extensional objects) that interact in various ways. This vision, although worth of respect, collapses when it approaches the subjective-objective problem. It is incapable of explaining the 'wholeness' of these atoms. Besides, it cannot give any reasonable account of what a representation is. It is a well suited framework very to the needs of an objectivist version of science. The second vision, the one we are advocating here, proposes a world made of onphenes. Each event is carrying its own pearl of meaning (that is of being and of representation). It is, of course, a relation. Each intentional relation is automatically a unity on the basis of what we have termed the Principle of Unification (§ 5.5). It cannot be anything else. Objects and other physical stuff are what remain of intentional relations after subjects have objectively filtered them. This second approach is perfectly compatible with objectivistic science but it is not capable of coping with a larger domain.

Why must we deny that reality is a giant flux of events, each one connected with others that are the condition of its very existence? If this framework is accepted, there is no reason to look at the skull as the natural boundary of a mind. What is the physical evidence that proves that only the events inside the brain are responsible for the corresponding mental states? As stated in § 3.1, there is no reason to exclude a temporally or spatially distant event in the determination of the content of a mental state. Even brain events, on a reduced scale, can be seen as spatially or temporally distant events. The skull is transparent to this intentional flux that constitutes reality. A simple proof is given by the fact that, inside our skulls, subjects are experiencing what there is outside.

A suitable example can be given by a concept developed by the Austrian biologist Von Uexküll, the *umwelt*[5]. Each subject lives, according to this

[4] A related point of view can be termed the objective empiricism: the acceptance of the entities that can only be experienced objectively. It is a kind of metaphysics since it entails an ontological statement on the basis of an *a priori* judgment on epistemic criteria.

[5] Von Uexküll spoke also of an *innerwelt* that was the usual mental world opposed to the external one. With TEM there is no distinction between the *innerwelt* and the *umwelt* (Uexküll 1909; Uexküll 1934).

biologist, in what he defined her/his *umwelt*, that is the set of the events whose meaning he/she/it is able to grasp. His ideas derived from his work in the field of zoology where it is possible to observe that, given the same environment, two different specimens of two different species can occupy the same physical space but can have a completely different experience of the same physical world. Each creature can experience only those events with which he/she is in relation to. However, determining the nature of this relation is still highly controversial and there are no ready made off-the-shelf answers. Nevertheless, it is clear that the *umwelt* of a tick is completely different from the *umwelt* of a human being even if both are physically located in the same wood(one of the favourite examples of Von Uexküll). What is this *umwelt* and in what circumstances does it occur? For example, can we speak of the *umwelt* of a computer, or of the *umwelt* of a car? Where is the invisible boundary between subjects and objects[6]? The answer of TEM is straightforward. The mind, the subject *is* the *umwelt* of itself[7]. There is no distinction.

Instead of trying to find the condition for the emergence of a subject able to endorse representations or mental content, it is worthwhile to pursue a different approach. By making use of the framework developed in the previous chapter in order to propose a suitable medium for a representation within a subject. What is a representation? A representation must be something; therefore an event. An event is something that has provoked a difference in the state of things. An event is also a relation. An intentional relation or onphene can exploit all these properties. Where should we look, in order to find these intentional relations? In causation. Whenever there is an event whose happening has been conditional to the happening of a previous event we have such a relation. The target of such a rationale is the basic unity of experience, or of representation. As a working hypothesis it could be proposed that whenever there has been a causal relation between events, that causal relation must be seen as the elementary unit of representation. In other words, whenever there is causation there is also an event. If a mental state has a particular content, it happens because it is in such a relation with all those events whose meaning is contained into such a state. Does this rationale provide an answer to the *umwelt* question? The *umwelt* is made up of all those events that are in intentional relation with the mental state. The meaning of the external events no longer needs to be carried inside the brain, nor does the natural structure have to assume the meaning of the

[6] These questions have been analysed by Brian Smith (Smith 1998).

[7] This approach has similarities with the work of other authors, most prominently with Bertrand Russell's theory of neutral monism (Russell 1995) and, more recently, with Leopold Stubenberg's theory of the subject as a bundle of qualia (Stubenberg 1998).

physical structure outside the mind. The mind remains partially outside and partially inside the physical structure of the brain.

> *The mind (the subject) is the collection of intentional events that starts from the external environment and ends into the neural structure of the brain.*

This explains both subjective knowledge and the objective side of such knowledge(more in § 6.3). In fact, the content of every representation must be dependent on the particular intentional relation that grasps it and on its particular content. The first part is dependent on the kind of subject while the second part is independent. For example, a normal colour sighted subject has a colour perception that is dependent on the fact that he/she is provided with three different kinds of photoreceptors, sensible to different light waves. Let's suppose that the same subject has a twin born with only two kinds of photoreceptors. This twin would be unable to be causally dependent on certain events. However he/she would nevertheless be causally dependent on a limited number of events that are part of her normal sighted brother/sister. The *umwelt* of the one would be a partial version of the *umwelt* of the other. The mental state of the normal sighted sister would be causally (intentionally) dependent on a larger number of events. The colour blindness of one of the two twins has determined a reduction of her *umwelt* as well as a difference in the content of he/her mental states. Notwithstanding he/she is seeing the world and with certain visual events (black and white surfaces, pure colours for which the difference in her receptors is not relevant) he/she has exactly the same kind of experience as his/her brother/sister's. In these cases the content is not determined by the subject but by the real content of it. The subject is determined by the content of his/her mental states that are, in turn, determined by the content of involved intentional relations. The mind is such a set of intentional relations and each one carries its own content.

6.2 The principle of self

At this point a natural question springs to mind, what is the principle that unifies a group of intentional relations in a subject? If a mind is a set of intentional relations and each of them is the carrier only of its own content, why there should be any kind of glue between them? The answer is that there must be a final event that constitutes the natural end of all these experiences, a binding event. It is a consequence of the Principle of Unification. Each onphene unifies a part of reality. If the subject is a 'whole', there must be an onphene that is responsible for the unification. The event corresponding to this onphene is termed *self-event* and it is the natural centre around which all intentional relations experienced by a subject will collect themselves. Informally, a subject is a set of intentional relations grouped around a special event called self-event that constitutes the related *principle of the self* or more briefly *the self*. The self is therefore different from the subject. The subject is the set of all the onphenes that are part of a particular mind. The self is the onphene that unifies the subject: since each onphene is at the same time something that is and something that, being in relation-with, is the content of another onphene. We can look at the principle of self from two different perspective: as an occurring onphene and as an onphene contained in another onphene. In the first case it would be more correct to speak of *ego*, while in the second case the term *self* would be preferable (Table 6-1). Although the definition of principle of self and that of subject might seem circular, they aren't. The fundamental term is that of onphene. From an ontological point of view, a subject is not different from the rest of the world. Each onphene is potentially a unification process that could correspond to a subject. However, there's a practical difference. A human subject is normally a process that unifies an extremely large number of onphenes, many more than in usual physical processes. Introspection provides first empirical evidence. In every moment of our conscious experience the number of events that are contained in it is enormous. Every ray of light, every object we see, every edge we perceive, every sound wave we hear, every pressure on our skin. And yet, large as this number can be, we perceive each and all of them as a whole, as a fundamental unity.

There is no reason why two intentional relations should not share the same final event. In such a case the final event would correspond to an onphene that unifies the two original onphenes. There is also a nice symmetry between intentional relations having the same content – and giving the same kind of experience to different subjects –, and intentional relations having different content but sharing the same final event – and therefore to the same subject.

Subject	A set of onphenes unified by a particular onphene termed *principle of the self*
Principle of Self	The onphene responsible of unifying the onphene belonging to the *subject*.
Self	The *principle of self* seen as a content
Ego	The *principle of self* seen in its occurring

Table 6-1 A summary of the relations among the four main terms of the theory of enlarged mind.

Given two different subjects what are the relations between their onphenes? How can they interact? In general there are onphenes that originate from the same event and became part of more than on subject at once; there are onphenes that end in the same subjects and that contribute to the content of one singular subject; there are onphenes that do not belong to any particular subjects. More precisely, given two distinct onphenes

$$R_1(e_1, e_2), R_2(e_3, e_4)$$

there are four possible cases:

1) $e_1 = e_3$ and $e_2 \neq e_4$. This is the simplest case. It means that the starting event – the event whose content is responsible for the content of the intentional relation – is in common between the two intentional relations. There are two intentional relations belonging to separate subjects as demonstrated by the fact that, *ceteris paribus*, R_1 and R_2 do not have the same final event. Nevertheless, they share the same content and they represent the same object[8]. An intentional relation does not necessarily belong to a subject. It

[8] The nature of this object is interesting and can easily vary. It is an immanent object in Brentano's sense because it is the object of an intentional act. It is also a noumenic object because it is the thing in itself as Kant would have conceived the hidden object of

is a more elementary structure of reality. Therefore R_1 and R_2 do not necessarily belong to a subject. A subject does not create them. They exist by themselves. If they were part of two subjects they would give them the same kind of content. For example, if e_1 was a red patch glistening under a golden light, and if $R_1 \in S_1$ and $R_2 \in S_2$, then S_1 and S_2 would have the same kind of subjective experience because they would be made up of two intentional relations with the same content.

2) $e_1 \neq e_3$ and $e_2 = e_4$. This case represents the basis for the grouping of intentional relations into a subject. Let's suppose that the final event is the same. It means that there is an event in the world that is dependent on two separate events. Two starting events determined a unique final event. In other words, $e_1, e_3 \in Iset(e_2 = e_4)$. The two intentional relations share the final event and carry different contents. The same subject has two different experiences of two different objects in the same sense as above. Why are there subjects and why are there intentional relations that do not belong to a specific subject? The answer illustrated using a situation familiar to most of us. Let's imagine that you are correcting a paper composed of several pages using a WYSIWYG word processor. In a page full of text you occasionally change some words. For example, you decide that 'but' can be substituted with 'nevertheless' on the tenth line of the second page. The new word is longer than the previous one. When you have finished you can notice that the word processor has shifted several words below the one you substituted. You might expect that the effect of the substitution would spread to the end of the whole document, several pages later. In most cases, this will not be the case. The effect will cease after a few lines. Typically, where there is a paragraph mark, often before. If you look at the flow of text as if it was a metaphor of the flow of events you can intuitively notice that the effect of an event is not going to propagate forever. It is far more probable that it will stop after a few passages. It is a clear limit of the frequently abused butterfly effect. The idea is that there are points, along the intentional or causal chain, that stop the propagation of its content and its effects. There are events whose absence or existence would not make any difference in the existence of other events[9].

---

perception. It is not an *imago vicaria* or a representation of something. It is the thing in itself. It could be indifferently an objective entity (like a table) or a subjective moment of experience (like a red patch) or even an abstract entity (like a number)..

[9] However, there are intentional relations that do not belong to one specific subject, all those intentional relations that do not have anything in common with a subject, that are

3) $e_1 \neq e_3$ and $e_2 \neq e_4$. A trivial case. R_1 and R_2 do not share anything and are two completely separate intentional relations.

4) $e_1 = e_3$ and $e_2 = e_4$. An identity case. This is not as trivial as it may first seem but may seem as an identity condition between intentional relations.

A mind is a set of intentional relations unified by the same final event. A mind, of course, is not an object, or a brain, or the state of an object. A mind is more similar to a always running process. It is not the engine but, in a sense, the running itself. It is always changing, like reality itself, because its essence is being a relation among events. And it is a representation, for exactly the same reason. A mind is not a computational problem or a functional state of a physical system whose correspondence between physical events and meanings would remain obscure and observer relative[10]. Subjects do not create content. However, cognition does the job of providing a complex agent capable of collecting an enormous number of intentional relations under the same subject.

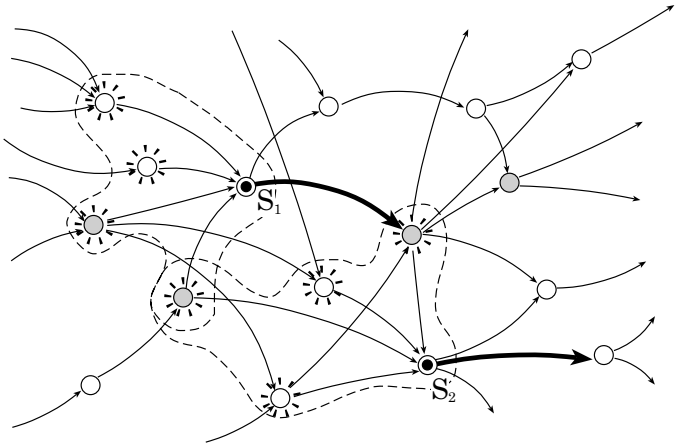


Figure 6-1 A conscious subject is nothing more than a part of reality unified by a particular onphene: the principle of self. A subject is a unified group of onphenes – that is a unity of representations, being and being in relation-with.

not that subject. The intentional relations that are now happening in another town, or at the opposite side of the world, or just in another room offer a trivial example. They do not belong to that subject that corresponds to myself (Tye 1996; Stubenberg 1998).

[10] (Searle 1992; Clark 1997; Manzotti and Sandini 1999).

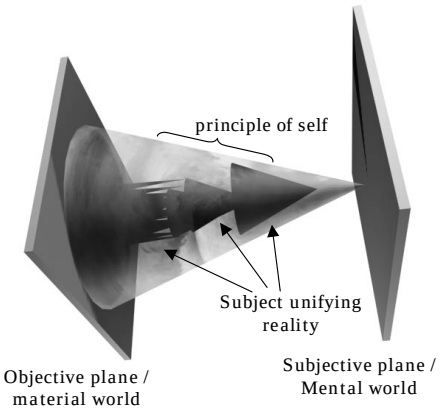


Figure 6-2 A graphical representation of the principle of self and its unifying act. The subject arises from the progressive unification of previous onphenes that concur with their content to the global critical event of the self.

6.3 Subjective experience and objective knowledge

The subject, as it has been defined, is not a problematic part of reality with properties or substances different from the rest of the world. It is not a strange, ineffable quality (or property or system) emerging from a world of pure objective relations. The subject is a set of intentional relations connecting events. Whenever something is experienced, it is because an intentional relation has entered into this set bringing the content of the associated events. Subjects are continuously exchanging intentional relations with the rest of the world. An intentional relation belongs to what makes up subjectivity if it is connected with one's self.

A central issue that must be solved is where the difference between objective knowledge and subjective experience crops up. In other words, why should reality be experienced in a twofold manner? Sometimes the qualitative, absolute incorrigible[11] essence of qualia is experienced. In other moments (or even at the same time) objective, empirical phenomena are observed. Let's imagine

[11] In Ryle's sense (Ryle 1984).

explaining the scientific objective theory of colours to a blind person. What can be explained to him is objective, what cannot is subjective. This is the shortest way to give an idea of the difference between the two epistemic realms. Looking at the world, two different kinds of epistemic objects are continuously experienced. On one side, there are the qualitative essences of perceptions (sometimes called qualia). On another side, there are the objective relations between such perceptions[12]. For example, when someone is experiencing a red patch, that kind of experience is, in a sense, absolute. The subject does not need to know anything about the world in order to know the content of her experience. Everything she needs belongs to the experience itself. It is the experience. But if near the red patch she sees a blue patch, then she can start experiencing the relation between these two different experiences. The problem is whether her experiences need to have content. What is she experiencing when she is looking at the differences between the two patches? It is enough to be seeing the two patches or there must be something in the relation itself? In other words, has the relation to be real in order to be an object of experience or is it enough that its components are real? TEM solution states that any conscious state is an intentional state and thus there is a real content for that state. What TEM must explain, of course, is how intentional relation can provide useful subjective perceptions of objective relations. Two interesting points must be emphasised here. First, the distinction between subjective experience and objective knowledge is no longer a difference in the activity but in the nature of the objects involved, events in the first case and relations between events in the second. It is possible to speak indiscriminately of subjective knowledge or of objective experience. Secondly, that even when objective observations are made, subjects are engaged in the same kind of intentional relations they have with subjective experiences.

In a particular sense, there is an even more objective realm. Imagine observing some empirical facts with a very neo-positivistic attitude. Something is observed and is translated using objective empirical asserts. After some time, regularities are detected and more general principles about them are formulated. The principle you are tempted to propose may be a geometrical

[12] The subjective-objective dichotomy, we refer to, is slightly different from the usual one. For example, for Tye the subjective refers to what needs a perspectival point of view, while the objective refers to what does not need such first-person point of view (Tye 1996). We advocate here a distinction based on the nature of the content of experience not on the modality we use to access it. Subjective knowledge is not the product of a different epistemic attitude but simply a different kind of content. TEM explains what the relation between these different kinds of content is.

theorem, the existence of an object, an inductive law, or a logic statement. Simply, something that is above the mere empirical asserts but that, nevertheless, is a more general principle about them. It is something similar to Popper's third realm[13].

Several different levels of knowledge can be perceived, just by looking at the world. At the simplest level, the subjective *quale* of experience is received. Immediately above, objective observations are advanced. Above them, more abstract and general assertions belonging to what has been generally called logic can be made. How can this inconsistency be explained in a world composed of only one ontological and epistemic principle as is the intentional relation?

In reality the answer is straightforward. Let's start with the simplest level. When something is experienced the corresponding intentional relation becomes part of someone's subjectivity. When a red patch (event e_s) is experienced, the explanation is that the intentional relation

$$R(e_s, e_f) \qquad (1)$$

becomes part of someone's subjectivity. The explanation is consistent with what we empirically know about our first-person experience. It is absolute. Besides, a qualitative experience is defined by (1), it is autonomous and, of course, it is private.

In this case the object[14], or original event or content or meaning, of the intentional relation is what is usually called a *quale*. An alternative and more precise way of expressing the same concept is to say that the subject constituted by such an intentional relation will have such a *quale*. The reason for this last observation derives from the fact that speaking of immanent objects[15] or events is neither sufficiently clear or, above all, necessary. If subjects are composed of such a cloud of intentional relations, each one with its qualitative pearl, how can objective knowledge originate from it? Each intentional relation can be seen as the most elementary form of connection between events. The fact is that an intentional relation is an event itself. If the general definition (that an event is something the absence of which would have meant a difference in the state of

[13] (Popper and Eccles 1977).

[14] It is worth spending a few words about the use of the object of an intentional relation. Because the object of an intentional relation is what constitutes the subjective side of our experience, there is some risk of confusion. Here the object has the same meaning it has in Brentano. The object is just the object of the subject and not the objective construction we derive from it. It is the immanent object of Brentano (1873).

[15] (Lanfredini 1994).

the universe after the event) is accepted, an intentional relation is maybe the only acceptable event. There is no reason why the immanent object could not be an intentional relation. In other words, there is an intentional relation with the form

$$R_2(R_1(e_s, e_f), e_{f1}) \qquad (2)$$

where the subject is aware of a relational event. There are several kinds of relational events such as difference, identity, causation, and others. What is the difference between (1) and (2)? The immanent object of (2) has a relational nature. Its essence is common to many combinations of basic events. Besides it is naturally inter subjective. The original quality of the event is hidden because the intentional relation it contains, as a starting event, is what constitutes its content. It is easier to communicate between subjects because it does not imply the reproduction of the original event e_s but only of the intentional relation $R_1()$. Therefore objectivity can be defined without ambiguities. When the content (or the quality) of an intentional perception is an intentional relation, its content belongs to the domain of objective knowledge[16].

It is worth emphasizing that representation occurs without requiring anything more special than mere existence. Representation is being.

We claim that if relations did not have their own private and personal content (and qualitative content) it would not be possible to know or to experience them. In other words, perception always needs an intentional relation whose object could be an event or an intentional relation of some kind. Epistemology and phenomenology require an ontological foundation that has been often underestimated. Relations have a quality of their own. If not, they could not be the content of an intentional relation or be perceived or known. An example will clarify this point. When someone perceives a red patch, the object of her experience is an absolute content. She does not need anything more in order to know what the content of her experience is except what she has the experience of. When she perceives a red patch and nearby a blue patch she perceives two absolute contents but, at the same time, she perceives the relation between the two. Usually it is assumed that the relation is dependent on the existence of minds and that the two coloured patches exist in a stronger sense while the relation is a belief someone can have about the two patches. While the two patches have a correspondence with physical objects, the relation

[16] For example, information can be reduced to pure differences between phenomena. Information is objective while its associated meaning is not. Syntax is objective while semantics requires subjects.

seems to be just a mental or an *a priori* fact. It has no corresponding extension. How could he/she perceive or represent something that does not exist? If the relation between the two colours did not exist how could it constitute the object of her knowledge? When she sees the two patches she perceives also the content of the intentional relations occurring between them. Evidences from perceptions show that even in the early stage of visual processing the relation among elementary events is processed, as if it was real content[17].

Another issue deals with the quality of intentional relations. All kinds of relations perceived, conceived, thought by subjects must represent one kind of intentional relation. Opposition, identity, and enclosure are all suitable examples. The world is literally tied up with intentional relations. Causal relations represent a special kind of intentional relation, from this taxonomic point of view. They are the nearest approximation to the true, and beyond possible knowledge, intentional relations. Causation, in this sense, is the bridge between the vanished past and the still not existent future. Intentional relations live in the present. Their role is to carry the content of the past towards the future and, by doing this, to preserve reality from disappearing. In this sense all other relations are an expression of this fundamental kind. For example, every kind of geometrical relations needs an observer. It means that a geometrical relation requires the process of perception that, as we have shown, consists of a chain of intentional relations. In the end, every conceivable relation is just causation and therefore intentionality.

Philosophers often argued about higher orders of knowledge about a world different from the mere empirical and objective asserts about empirical facts. Logic principles are a suitable example. Can they be ontologically founded? If intentional relations are events themselves, why should we not imagine being able to experience higher order levels of them? An example could be

$$R_3(R_2(R_1(e_s, e_f), e_{f1}), e_{f2}) \qquad (3)$$

In other words, it is possible to imagine experiencing higher and higher levels of abstractions just by being able to respond to higher order intentional relations. There is no theoretical limit to this process apart from two considerations. First, the more our intentional relations will have, as immanent objects and content, other intentional relations, the less their content will be enriched by the direct content of basic events of reality. In a sense, their content will be increasingly void of content even if increasingly general. Secondly, it seems intuitively probable that, as the process will go on, the number of

[17] (Marr 1982; Marr 1991).

possible content will decrease progressively. Each level of abstraction will reduce the number of possible variations in the class of corresponding immanent objects. A simple verification of this last consideration is the progressive simplification of laws that can be observed by going from the qualitative experience to the most abstract laws.

An example is needed to summarize what has been said up to this point. Let's imagine two subjects on a beach and a meteorite from outer space that accidentally lands on the foot of one of them. As a consequence, the injured subject has an extremely unpleasant sensation of pain in the remains of his leg. We have an event that is neither completely objective - physical nor completely subjective. If this episode were described with a classical extensional terminology (or with a scientific physicalistic vocabulary) it must be explained how a physical cause can provoke a mental effect (and the other way round), what is the place of pain in the physical world, what is the difference between the first-person experience of the injured subject and the first-person experience of the other, and so on. If we consider the episode from an intentionalist view point we see things from a different perspective. First, there are no longer autonomous objective entities such as meteorites, bodies. There are no longer embarrassing mental entities such as pain, subjective experiences and objective knowledge of supposedly objective facts. The intentional story of that episode is more or less as follows. A series of intentional relations, not belonging to any subjects and occurring in outer space, acted as to determine following intentional relations in a particular way. As a final result, this long chain of unconscious intentional relations, which could have been interpreted by external subjects as the approaching of the meteorite to the earth, the same meteorite came into contact with two large groups of unified intentional relations occurring on the surface of earth. These two large groups were, of course, the two subjects. One of them interacted directly with the chain in a very dramatic way and, as a result, felt intense pain. He had what is considered a first-person experience. In fact the group of intentional relations corresponding to his mind extended to several unpleasant event. The safe subject limited himself to a more indirect interaction with the occurring event. He knew what was happening (the destruction of the his companion's leg) but only because he perceived the relations between events that were, luckily for him, external to his own mind (in the sense of the enlarged mind). Using the intentional interpretation there are no difficulties in explaining the interaction between different objects, as well as the difference between subjective first-person experience and objective third-person knowledge.

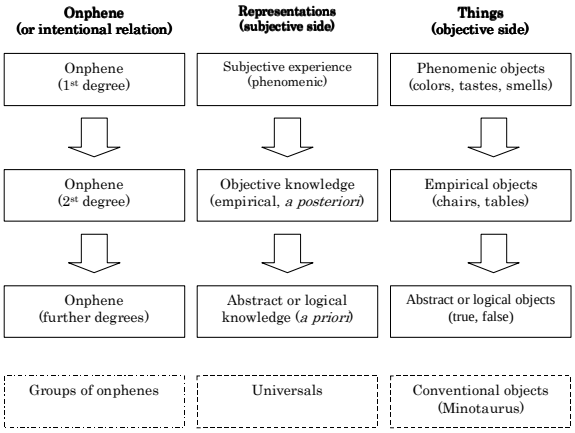


Figure 6-3 An overview for the categories provided in this chapter.

6.4 Communication

A few words must be spent on the way communication can be realized between subjects. Given two subjects – that is two sets of intentional relations, how can they communicate with each other? Before giving a full answer, it must be stressed that communication is a more subtle process than just the transmission of information. The former involves the conscious grasping of the same content between subjects while the latter implies only the existence of a causal relation between two essentially unconscious physical phenomena. The fact that communication involves transmission of information has created a lot of confusion on this point. What is meant by communication is the conscious extension of the unconscious transmission of information. To uphold this link between consciousness and communication that will be fiercely rejected by analytically oriented philosophers[18], the following argument is proposed. A couple of gears are connected together. The state of the first (angular position and speed) determines the state of the second. Would it be correct to say that the first *communicates* its position to the second? This does not seem to be the case. The lack of someone who is conscious and giving its meaning to the information, prevents us from calling what has happened 'communication'. Besides, at both ends of the transmission chain there must be a conscious being. Would it be correct to say that a human being communicates with her/his car,

[18] (Dummett 1978).

or that her/his personal computer is communicating with her/him? It is not the case. '*Interactions*' is a better term to define all those cases in which there is information transmission but no conscious activity.

How can communication, in the sense outlined above, occur between subjects? Let's briefly analyse how it is possible to reproduce an intentional relation. In all cases we suppose that a subject S_1 having the intentional relation $R_1(e_s, e_f)$, bearing the content of e_s ($R_1(e_s, e_f) \in S_1$) exists. Our problem, here, is how can another subject S_2 reach the content of e_s? Let's suppose that e_s is an elementary event (that is, it is not an intentional relation or a first-person experience). Here are four possible answers related to the nature of the original event e_s.

1) $R_2(e_s, e_{f1}) \in S_2$. The second subject is having an intentional relation with the same elementary event e_s. The two subjects S_1 and S_2 cannot be sure they are having the same original event but if they are in relation with the same phenomenon they are having the same qualitative experience. A consequence of this argument is the elimination of the inverted spectrum argument. There is no reason why people should bind alternative qualitative representational patterns to their colour perceptions because in the old sense representations are no longer . For the same reason there is no longer any need to imagine a realm of immanent objects to be correlated with a realm of noumenic objects. Only elementary phenomena exist: the same for everybody.

2) $R_2(R_1(e_s, e_f), e_{f1}) \in S_2$, $R_3(e_s, e_{f1}) \in S_2$. The second subject is observing the causal relationship between some external phenomenon, from the outside, and the supposed related elementary event e_s. He/she is taking the first intentional relation R_1 as an event itself through R_2 and, therefore, she is loosing its qualitative content. Nevertheless she is able to perceive the same event through R_3. This is the case of a normal neuroscientist that is observing her/his patient's brain activity and who is able to have the same perceptions.

3) $R_2(R_1(e_s, e_f), e_{f1}) \in S_2$ only. The same as before but without the ability of experiencing the essence of the elementary event in a first-person qualitative subjective way. This is exactly the situation of the super scientist Mary or of the zoologist studying a bat[19]. In this case, we are in the

[19] (Nagel 1974; Jackson 1986).

situation of being merely watching the vehicles of representations not the representations themselves[20].

4) $R_2(e_f, e_{f1}) \in S_2$ in such a way that there is also a valid $R_3(e_s, e_{f1}) \in S_2$. This is still a theoretical possibility. There are no practical examples. The idea is that if it were possible to extend one intentional chain between two subjects (neurally, artificially) then it would be possible to reach the original, private, qualitative, first-person content of another subject's experience. Just imagine merging two cortices together. A more restricted version of this would permit the addition of new qualia previously not accessible to a person's qualia.

Why is objective knowledge more easily transferred than the first-person content of intentional relations? The reason is clear at this point. If an event e_s is communicated to other subjects, the same kind of intentional relations associated with that event must be provoked. Unfortunately, this goal can be reached only indirectly. For example, by showing the event itself to other subjects. The internal dynamic and physics of subjects is not always known and it is not certain whether they are including precisely that event or another one in their private subjectivity. Let's suppose that the internal content of the intentional relation is an intentional relation ($e_s = R_2(e_{s1}, e_{f1})$). If before it was impossible for a subject to produce the original event itself, it is now possible to produce another intentional relation albeit with a different (but hidden) internal content.

An example will clarify this issue. Let's suppose that I want to communicate the content of my experience of a red patch. I can only show red objects to other people. However, if they are blind or are using strange contact lenses or whatever, I am not sure that they are having intentional relations with the same internal content as my own. Besides they will not be able to give me any useful feedback to prove to me that they have had the same perception I had. Now, suppose I want to communicate the content of my experience of $a=b$ (or $a \neq b$). There are several phenomena that share the same kind of intentional relations in some respects. Besides, what matters now is not the qualitative content of these experiences but their relation (that is of course their relational content). Thus, I am able to check whether other people have had experiences with the same relational content I had. To prove it to me, they will have to act in such a way as to produce intentional relations that share the same relational content as

[20] (Dretske 1995; Tye 1996).

my original communication. If this is the case, the communication will have been successful.

When we try to communicate something to somebody, we are immediately not able to let our interlocutor have the same intentional relations we have. What we can try to do is communicate the net of intentional relations, in which the contents we want to communicate lie. Two strategies are classically used to implement communication. The first one is some kind of direct experience. That is, on the basis of our knowledge of the physical structure of our interlocutor, we try to provoke in him the same intentional relation we are having. We point a finger in the direction of the colour we want to communicate. Of course, this is not possible if we are speaking to a blind person. The second strategy usually involves language and must make use of a net of intentional relations. For example, let's suppose that we have to transmit the content of 'being thirsty'. Bearing in mind that we cannot deprive our interlocutor of water for a few days, we must resort to the net technique. We suppose that there are other contents, which our interlocutor possesses. Then we transmit to our interlocutor the relations between this supposed common ground and what we are trying to communicate. It is not always possible. For example, we can imagine that there are self-contained lands of our experience for which there are no possible bridges. Communication is impossible if we cannot start from a common ground. One intuitive image is that language is a kind of net. The pearls of content rest in the nodes of this net. When we speak, we send pieces of the net without any pearls. The comprehension process of each subject is the attempt to find the best possible adaptation between the transmitted piece of the net and his/her own larger net built with first experience. When the subject succeeds in this attempt usually she exclaims 'I got it'. Language is the net. It represents the structure of the intentional relations between events. It is exactly what Wittgenstein called the logical form[21]. It must continuously change to adapt to the new intentional relations that subjects are having during their lives. Some parts of it may seem to be more stable because they are related to higher order intentional relations, in the sense outlined at the end of § 6.3. They are really stable, or even eternal, but there is no real *a priori* way to know it. At the bottom of the net, or at its boundaries, there are always new intentional relations that incessantly modify the structure of the net. After several millenniums of evolution, the net, the language, has become huge and a new concept may need time to make its way towards the centre of it. Nevertheless, as far as we know, the net could be still at its beginning.

[21] (Wittgenstein 1974).

Finally, a consideration must be given to the problem of universals. In TEM a universal is simply a stable portion of the net. It is a piece of language, which cannot be easily jettisoned. It does not have to be perfectly coherent. The net evolves constantly and it is possible for it to expand in strange ways. Nevertheless a large number of subjects own the same part of the net. This makes a universal a potential candidate for some kind of ontological promotion. Its stability can be only apparent, as has been showed in famous debates about natural kinds and personal identity[22]. However, there must be, in our framework, two plausible candidates occupying the role of universals. The first is the one outlined above. Another one is represented by the content of intentional relations. For example, the content of an intentional relation related to have a red quale, or to have an objective experience of another intentional relation. This second kind of universals can be seen as the no–conceptual atomistic correlates of the first kind.

What place is left for the classic knowledge that is derived from empirical observations and from the formulation of scientific objective theories? Exactly the same place it occupied before. Scientific knowledge represents the progressive systematisation of the second order (and followings too) of intentional relations. TEM does not intend to substitute the scientific point of view. It is a larger box into which previous knowledge will find its proper collocation. Objective facts, as well as theories derived from them, keep their objective value. They are simply freed from the impossible duty of explaining the subjective side of experience and reality as well. Besides, they can be looked at from a different and new perspective. They are no longer objective facts or objective entities, void of qualitative values, from which problematic entities like consciousness, content, and quality, should arise. They are the intentional relations of the second order perceived by subjects. In a sense, they are quality. They are the qualities of intentional relation perceived as events. The physical world is no longer the root of everything but, as should now be clear, it is just the objective restriction we apply to empirical facts. The physical world so defined is, of course, a real part of reality but only a part. The goal of science is ordering this part. Subjects are the result of the collections of intentional relations that share the same final event that makes up the principle of the self. Reality results from the dynamic interaction between the ontological and representational role of intentional relations.

[22] (Quine 1969; Putnam 1975; Kripke 1980).

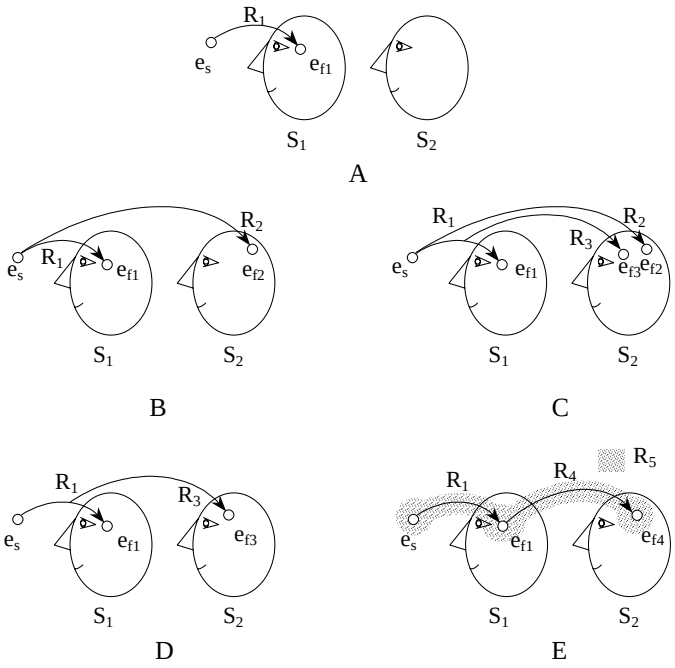


Figure 6-4 Different cases of communication between two subjects. Subject S_1 wants to obtain subject S_2 to have the same conscious experiences.

Summary

Given that the world is made up of onphenes each with its pearl of being, representation and being in relation-with, what is a subject? A subject is a set of onphenes that are the critical event of a subsequent unifying onphene. The last onphene is called *principle of the self*. The unified onphenes are each a unity in itself (of course, since they are onphenes), while collectively they make up the *subject*. The subject is a unified set of representations. The content of a subject is given by the critical event of the onphenes that constitutes the subject.

A conscious mind is a part of reality that finds its proper unification in the content of a particular individual onphene. Since such a mind literally extends itself to all the events that become part of it (content and existence are the same), this theory is called the Theory of Enlarged Mind (TEM) . Mind is no longer constrained by the physical boundaries of an object (the brain) that must mysteriously assume the meaning of external objects. Mind is a process that extends to everything that constitutes its content.

Subjective experience corresponds to first order onphenes , while objective knowledge and other kinds of knowledge belong to higher order of onphenes.

TEM can help provide a different explanation for activities like communication that would return to their intuitive original meaning, which is the exchange of mental content.

7 Neural networks and intentionality

Put the monkey into the loop
AI researcher[1]

Given what we have said in previous chapters, the two fundamental problems related to the comprehension of the mind and therefore of consciousness can be summarized: the unity problem and the representation problem. The first is related to how a higher level can arise from a lower level. How can the multiplicity of things of atomic and molecular events and structures could be experienced as unified wholeness. The second problem is related to the nature of representation: how can something autonomously refer to something else. In this chapter we will begin to transfer the hypothesis made in previous chapters to implementation issues involved in building an artificial being. The fundamental thesis of onphene (OT) is restated as follows.

A representing event of the event X is any event Y such that X is its critical event[2].

In this sense every static structure cannot rely on representation. Static objects do not represent. And the same consideration can be applied to memories, variables, photos, paintings, vectors. This is generalized as follows.

No static structure can represent anything.

Nevertheless a neural (artificial or natural) structure might be such as to permit the occurrence of a particular event Y following another event X. In this sense that structure does not represent anything but is necessary for a particular representation. It is easy to consider that structure as if it were itself representing something. Unfortunately it is a big conceptual mistake that leads to the impossibility of implementing a real autonomous representation. In this chapter, we will analyse what is required for a neural network to be an efficient

---

[1] The sentence was pronounced during a workshop about computer vision and navigation held at the INRIA centre in Sophia Antipolis, France, summer 1998.

[2] See § 5.7 for the definition of essential cause (that is the same as that of critical event).

environment for representations. This is an important keyword in order to grasp the difference between the TEM conceptual framework and the traditional one. Traditionally computer systems, a list of words or, a neural network are seen as things, which represent something in the external world. On the contrary we see a neural network (artificial or natural) as something that permits representations to occur. A neural network is not a representation; it is an *environment* in which representations can occur. A subject is not a thing but rather a process – that is a series of events linked by a particular kind of relations. To define the conditions in which a robot embodying certain kinds of neural networks becomes a subject, will be the goal of Chapter 7-9.

7.1 Control systems versus representational systems

> *I want to take mind to be the control system that guides the behaving organism in its complex interactions with the dynamic real world.*
>
> Alan Newell[3]

Traditionally a lot of emphasis has been put on mimicking human behaviour. This was understandable since in the '50 the prevalent theory of mind, behaviourism, claimed that the human mind had to be reduced to human behaviours. An artificial mind should be a system capable of producing the same kind of behaviours of a human mind. The previous equation was nicely represented by the famous Turing test. Its author also claimed that

> I propose to consider the question "Can machines think?" […] I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words. The new form of the problem can be described in terms of a game, which we call the "imitation game"[4].

This «imitation game» implicitly rested on the belief of the equivalence between a mind and the collection of behaviours that that mind makes possible. This belief is still widely upheld: «whether a system has a mind, or how intelligent it is, is determined by what it can and cannot do. Most materialist philosophers and cognitive scientists now accept this general idea[5]». No

[3] (Newell 1990), p. 43.

[4] (Turing 1950).

[5] (Haugeland 1997), p. 3.

attention was given to the phenomenal aspects of mind that were left to psychology, neurology or, worse, to philosophy.

This point of view determined a strong behavioural attitude in cognitive sciences. An attitude that produced in related fields what we call here the *control theory approach*[6]. In other words, since the mind was seen as a mere mechanism that would produce the appropriate motor responses, the goal of artificial network was to control the behaviour of agents appropriately. This principle is more or less universally accepted in neural network research[7]. A clear statement of this point of view is the following.

> Artificial neural networks are an attempt to model the information processing capabilities of nervous systems. Thus, first of all, we need to consider the essential properties of biological neural networks from the viewpoint of *information processing*. […] There is a general consensus that the essence of the operation of neural ensembles is «*control* through *communication*[8].

The problem with this approach is that it builds any theory of the mind upon three key concepts: information processing, control and communication that, as we have seen in Chapter 1 are derived from the existence of biological human subjects. The author quoted above is aware of a dependency between the previous three concepts and human subjects when he later recognizes that «it is implicitly assumed that a certain coding of the data has been agreed upon». To avoid such circularity it is important to have an alternative foundation.

The control theory approach cannot provide sound foundations for a theory of the subject, since it lacks the capability of defining and explaining internally the term upon which it is built. In particular its main fault is its intrinsic incapability of defining the goals of the control. In other words, the control theory explains how to get a certain result *given* a certain goal. The difficulty is that there must be at least one subject that suggests that goal. Similarly, the communication theory explains how to exploit the physical properties of channels *given* a certain content to be transferred; the information theory explains how to manage information *given* a certain code for representations.

Artificial neural networks have been used as a more or less efficient way to achieve particular tasks. From the point of view of a theory of mind, they have been seen as a black box (Figure 7-1). Modules that learn to implement a

[6] (Wiener 1961).

[7] (Fukushima 1975; Edelman 1987; Hornik, Stinchcombe et al. 1989; Massone and Bizzi 1989; Specht 1990; Geman, Bienenstock et al. 1992; Girosi, Jones et al. 1995; Basti 1996; Rojas 1996; Quartz and Sejnowski 1997; Arbib 1998; Sutton and Barto 1998).

[8] (Rojas 1996), p. 3.

particular function, a particular behaviour. They are not different from control theory in this respect.

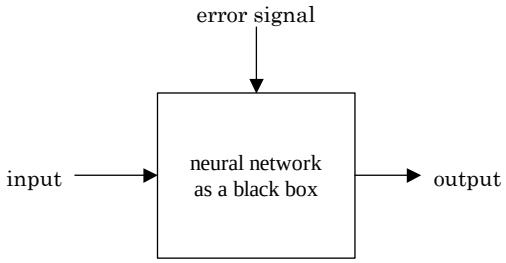


Figure 7-1 If neural networks are seen as simple behavioural units then they are equivalent to classic control modules.

7.2 Ideal and real intentional system

What is the ideal intentional system? What is the difference between a human being (a subject) and an artificial system (an object)? Can TEM be used to define the properties of an intentional architecture? Let's summarize the differences we have hitherto observed between intentional subjects (IS) and not intentional artificial systems (NIS).

- IS has intrinsic unity while NIS is only a collection of physical events that is arbitrarily seen as a whole.
- IS is capable of representing the external world, while NIS lacks real meanings
- IS instantiates internal events corresponding to external events, while NIS controls their behaviour
- IS is capable of referring autonomously to external events, while NIS does not possess any intrinsic representation
- IS possesses its own motivations, while NIS must be controlled by externally induced motivations

Human beings are intentional systems in this sense as we know from a first-person perspective. We *know* that our mental states refer to external objects.

Were we to deny it we must conclude that we have no direct proof of the existence of any external objects and have no evidence of the existence of our brain either[9]. We know that our mental states are capable of referring to something else.

We also possess motivations and goals without requiring any external attribution of them. And we know that natural selection produces organisms like ourselves. It is reasonable to be concerned about this fact and to consider intentional beings as somehow more than purely behavioural beings.

Are artificial neural networks capable of producing an intentional subject? As we will see in the subsequent chapters, they could become part of a structure that is essential for the production of events in a particular order. The occurrence of these events will be the basis for the emergence of a real subject.

7.3 A taxonomy for neural networks

What must a neural network accomplish? Here we propose to divide neural networks considering how they make events occur. It will then be possible to distinguish between neural networks devoted to solve behavioural problems and neural networks fundamental to the developing subject. The proposed taxonomy, which corresponds to an increasing tendency to be the basis of a true intentional subject, is the following:

- Input-output networks
- Networks self organizing their stimuli
- Networks self selecting their reinforcement signals

Should we conclude that a human subject is such because in his/her development a larger number of its neural networks belong to the last kind? This statement is, as it stands, too rough. Nevertheless it is possible to observe by analysing different species, which come closer to human beings, that a progressive structural modification has taken place (Figure 7-2). The human

[9] This is an example of what Aristotle would have defined as the truth: something that it is affirmed while it is denied (Olgiati 1953).

brain is not only the largest[10] and the one with the highest number of neurons and connections. It is also the brain with the largest number (absolutely and relatively) of neurons not devoted to a specific task (motor task or sensorial task). If we look at the areas lacking any precise goal as those that would develop the more extreme form of intentionality, we can understand immediately why human beings are the subjects *par eccellence*.

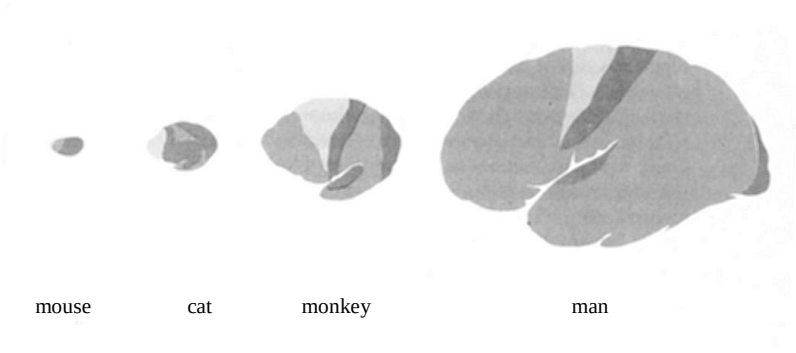


Figure 7-2 The modification in size and composition of brain belonging to different species.

7.3.1 Input-output networks

The most used model is what we have called the 'input-output' network (Figure 7-3). It corresponds to the logic of the control theory. There is a signal that is received and a required output that must be performed. The goal of the network is, of course, external to it. Besides there is a signal, generically called *learning signal*, which must control the learning process of the network. We are concerned with this kind of network because we are worried about the network's capability of performing some kind of operation. Philosophically, they correspond to a behaviourist paradigm. What matters is the final behaviour of the network and not what happens inside of it. The element characterizing this network can be summarized as an input signal, an output signal and the capability of receiving an estimate of their behaviour. The estimate is always external to this kind of network.

[10] This is not completely true. Elephants and certain whales have larger and heavier brain but the increase in size and weight is usually due more to glial cells than to neurons.

Generally, it is not even mandatory that they be implemented as networks. Other kinds of algorithms that offer similar performances (look up table, optimum control theory, square minimum) can be used, although the mathematical tool provided by neural networks is efficient. If we are interested in producing a certain correspondence between the output and the input of a block, we are not really interested in what happens inside it. Exactly the same attitude that moves behaviourists to analyse human or animal behaviour.

Several kinds of neural organizations have been used to implement this kind of behaviour. Both supervised learning and reinforcement learning will usually achieve the desired results. The only difference is that in the first case examples of the correct output are provided to the network, while only an estimate of its behaviour is available in the second case. In both cases the network has nothing to do with the determination of what the final goal of its activity is. As it is possible to see in Figure 7-4, the origin of these learning signals is always a conscious human being (typically the designer of the network) that decide what the relevant effects that the network is to provoke in the environment are. Take a network that controls the temperature level in a room (a very trivial task for a neural network). Let's suppose that it receives a temperature reading and must produce a signal that controls the flow of gas. Who decides what goals must be pursued by the network? Must it aim at minimizing the gas expenditure or at keeping the temperature of the room as constant as possible with disregard for any financial considerations? Should it follow a particular time curve during the day? What is the appropriate temperature? All these questions receive a prompt answer whenever a user is added to the description of the system. If the user were a rich owner, financial considerations would be irrelevant, while if the user were someone with a limited income a few degrees of variation would be only a minor nuisance. Wisely, a network designer would provide an easy interface to allow future users to tailor the network behaviour according to their subjective wishes. Yet the problem why someone should aim at these goals remains. For modelling purposes, we can forget the presence of the user in a input-output network. Yet in order to give a meaning to the parameters used to tune the learning of neural networks, we cannot eliminate the need of a conscious user. A nuclear bomb, for example, has completely different goals from those of a nuclear plant. The former must produce the fastest possible chain reaction in its radioactive material, while the latter should slow it down. The difference is not in what they are, but in the meaning people assign them.

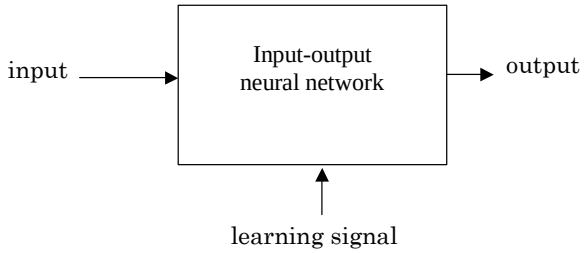


Figure 7-3 The input-output architecture. A traditional representation that apparently does not need any subjective interference.

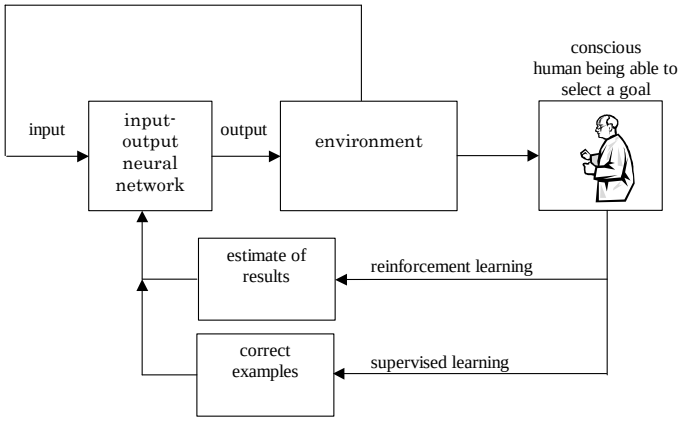


Figure 7-4 The loop within which the learning of a network occur is presented without explicitly showing the role that human subjects have in choosing the criteria for network evaluation. The above figure shows that between the effects of every network and the signals sent to guide its learning process there is always a human subject.

From this point of view, the difference between reinforcement signal networks and supervised learning networks is a mere mathematical difference among learning algorithms. However, it does not change the fact that the choice about what has to be seen as a good result and a bad result must be done by a conscious subject applying a highly subjective criteria. All objective criteria that have been used to train these kinds of networks hide a subjective category of events that their designer appreciates for highly subjective reasons. It is possible to claim that there has never been a totally objective criterion for training a neural network. Let's think of another example: the cart pole setup. In this example, a cart must balance an instable pole in the vertical position. It is a

motor problem related to the instability of the pole. Eevery movement of the cart tends to make the pole fall. Why should the pole remain vertical? Because its designers have subjective reasons to want it to be like that. For instance because they want to test their algorithms or maybe because they have a real cart-pole waiting outside the door of their laboratory and they want to use it. None of these is an objective motivation (there are no objective motivations as such). What is a learning signal sent to a network then? It is simply the best mathematical translation that engineers have found of theirs or someone else's subjective desire.

Similar situations are those provided by pattern recognition, sensory-motor coordination and function approximation. Let's analyse both cases briefly. In pattern recognition the network has to learn to recognize a certain group of stimuli by providing the appropriate answer. A classical case is the recognition of letters starting by a visual input of a surface (another interesting case is face recognition). This case can be easily transformed into a simple case of control theory. Given a series of input vectors, each corresponding to an image, the system must produce a signal that corresponds to a particular letter. For example, let's suppose that the input signal is a binary vector of 8x8=64 elements, and that the output signal is a vector of 27 elements. The network should learn to associate the correct output to every combination of the input signals. But what is the correct output? Is there always a correct output? A first caveat is that, in order to evaluate the input of the network, one must interpret the information it gives and must possess a code to assign a meaning to its physical results. Let's ignore such a difficulty. Ideally the network can be trained as to approximate the best possible subjective outcome: the subjective desire that at each visual stimulus it is possible to associate the letter that produced that visual stimulus. In reality, there is no totally objective translation of this goal – that is a translation that does not depend on a subjective version.

The sensory-motor coordination possesses similar characteristics. There is an input signal corresponding to some sensory information and an output signal corresponding to the activation of a motor apparatus. What is the desired motor activity that must follow a particular signal, for example a visual stimulus? Even simple cases like reaching a point in space with a redundant series of joints can be solved in different ways, each corresponding to a particular selection of subjective criteria. Should the time to reach the target be minimized or should the energy spent be reduced? Should either the error of position or the error of velocity be minimized? Several attempts have been made in order to mimic the trajectory curve of biological systems. While these studies provide an interesting insight about how biological systems work and what goals they try

to achieve, it is not from them that it will be possible to understand how motivations are produced.

Function approximation is another field that provides a good summary of the common properties of the previous problems. If the goal of a network is to approximate a function, there must be an input vector and an output vector. The more the output vector matches the desired function, the better the behaviour of the network is considered. Clearly, the function to be approximated might be known or not and this leads respectively to a supervised or to a reinforcement network. All previous cases can be reduced to this one. For example, the pattern recognition problem is seen as an approximation of functions, where the argument is an image and the letter contained in the image corresponds to the value that must be assumed. The problem is always the same. The choice of the function to be approximated is a completely arbitrary and subjective choice. It depends on the existence of a subject able to have goals and motivations.

Control theory, communication theory, and information theory have all something in common: they cannot be defined without considering the presence of a conscious subject that takes subjective decisions. As we have said, nothing is information unless it is associated to a subjective representation and nothing can be communicated if there are no conscious senders and receivers. Similarly, a control system requires the existence of a user whose goals must be satisfied.

It is possible to summarize the properties of an input-output network like this.

- All learning signals are *hardwired a priori*. The network designer chooses its goals.
- Its behaviour is similar to optimum control, pattern recognitions, sensory-motor coordinations.

From an evolutionary point of view, it is possible to observe that simpler animals have this kind of neural organizations. Some motivation has been chosen by natural selection and it is heavily hard-wired in the neural structure of the organism that invariably pursues it. There are no degrees of freedom from what is coded inside the genetic code. The subjective experience and the environment have little or no effect on the behaviours of these animals. They are, in a sense, environment-independent at least in determining their goals. Suitable examples are viruses, bacteria, protozoa, and insects. Some insects are provided with the capability of learning from the environment but there is good

evidence that such personal deviations from the average are more of a predicted sort than a true innovation. In other words, it is true that highly social insects like bees seem to learn particular behavioural codes during their life but it is highly reasonable that these different behavioural codes are the possible outcomes of their genetic code. The genetic code provides these insects with many innate behaviours and that only one is then activated. Nevertheless their neural structure (or behavioural pattern in animals lacking any neural cells) is precisely coded by their genetic code and must only be tuned in. What happens exactly in input-output networks? In more complex animals, even in humans, we can assume that many structures work in this way: for example, sensory-motor coordination.

7.3.2 Networks self-organizing their stimuli

The previous structure is completely dependent on its evolutionary base in biological beings, or on its designer's project in artificial networks. In order to make a first step towards an intentional being we should modify the design of the network in such a way that its growth becomes more dependent on the environment. Networks capable of self-organizing their stimuli in ways dependent only on their experiences (the series of events that enter in contact with the network) belong the to this class of networks. A possible solution is given by splitting the network into two halves (Figure 7-5). The first half of the network is capable of creating intermediate categories that correspond to certain events. The second half of the network is more or less equivalent to the input-output network seen in the previous paragraph. In other words, the second half tries to select the best choices given some *hard-wired* criterion. The second part is relatively independent of the environment because, whatever it receives from the first module, it tries to use its input to optimize (maximizing or minimizing) the learning signal. The behaviour of the first module is much more dependent on what happens at its input. Given the finite capacity of real systems, the class of events, with which the network will enter in contact, will be only a very small percentage of its full possibilities. Its output will vary depending on the effective experiences of a particular instance of the network. The conclusion is that the output of the first part of the network is not predictable independent of the environment in which it will work. Even specimens of the same network in the same environment could result in different final outputs. While the meaning of the second part will always regard the goal defined by its reinforcement signal, the output of the first is unpredictable.

The first module might be more or less free to mirror its experiences. There might be a signal that controls the selection of a particular pattern. This is possible in many cases, even in human beings (sex is one of the clearest examples). Nevertheless given a sufficiently large categorical capacity of representation, the behaviour of the first module might be almost completely free. There are simple criteria like reinforcement of more frequent inputs or selection of uniformly spaced input vectors that avoid imposing any kind of specific bias on the input.

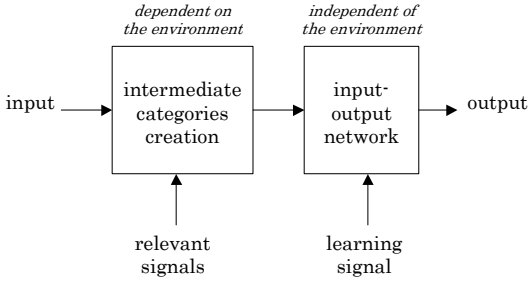


Figure 7-5 A network capable of self-organizing its stimuli. Ideally it is divided into environment-dependent part and in an environment-independent part.

The most evident limits of this network are

- learning signals are always hard-wired, defined *a priori* and incapable of changing in response to the environment

- the categorization endorsed by the first module are more or less biased by a signal stressing relevant events

- the capability of mirroring the environment is partially hidden by the fixed response of the controlling second module

In nature these kinds of network correspond to species that are capable of selecting their own categories. Nevertheless the behaviour of the corresponding individuals is bound to fixed patterns. They have no personal goals or motivations. Apart from genetic errors, each individual pursues the same purposes of all the other individuals of the same species.

7.3.3 Networks self-selecting their reinforcement signals

There is an onphene (an intentional relation) whenever there is a counterfactual relation between two events

These networks correspond to a network organization that is completely dependent on the environmental stimuli. In other words, its capability of organizing itself should be so high that these networks are seen as mirrors of the events to which they are receptive. The final behaviour of this kind of network is of course unpredictable because it depends on those events that enter into its life.

Each signal is a product of some interaction with the environment. The network is heavily dependent on stimuli received during its development. No signal is hardwired.

The genetic code limits itself to define some general rules and to provide the necessary physical structure in which this kind of network can be implemented.

This network is characterized by

- the capability of producing its own learning signals
- the capability of modifying itself according to such signals
- no hard-wired signal
- different networks with different experiences result in different internal architectures
- different experiences cause different individuals (form of subjectivity)

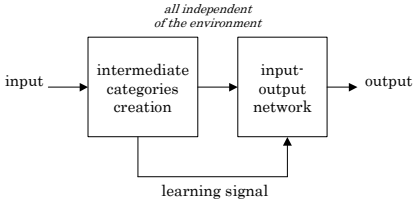


Figure 7-6 A network capable of self-organizing its stimuli. It can be ideally divided into an environment dependent part and into an environment independent part.

7.4 Neural networks, semantics and environment

> *In particular information must not be confused with meaning […] the semantic aspects of communication are irrelevant to the engineering aspects*
>
> Shannon and Weaver[11]

It is necessary to distinguish between the abstract meaning associated with a neural network, its physical static implementation and the occurring events that such a network allows. This is important because a subject is mainly a semantic structure (see previous chapters) as, when building a subject, the semantic aspect must not be considered irrelevant

In the classical communication theory (as well as in computer science) there is a universal criteria to distinguish between simple physical phenomena and representations or symbols. The latter always need a human being that is the receiver or the sender of the information involved. For example, in Shannon and Weaver's famous work, the semantic aspects of information can be dismissed as irrelevant because they focus only on the communication channel that exists between two points: yet there must always be two conscious subjects at the two extremes. Ignoring the semantic aspects was a simplifying hypothesis useful to concentrate on practical problems related to the engineering aspects of communications. Such hypothesis has now become a misleading burden with no explanatory powers. It was a working hypothesis but has been taken by many as an ontological principle. Besides, if the aim is to build an artificial subject, the semantic aspects are central.

If we looked at every artificial autonomous tool (take a simple robot doing some activity in a real environment), we immediately notice its physical structure. Eventually, we might examine its internal symbolic structure (the program that controls the robot). The physical aspects are real without any doubt. They *are* there. If we look inside the program, things are not so neatly defined. The program is just the abstract series of symbols that are associated with the internal physical structure of that machine, not the physical structure itself. Without a programmer, there would be no program either. Besides, I could argue that it is not easy to locate a precise program inside that machine. For example I could argue that, inside the robot, there is a high-level language program (like C++ for example) or an assembly language program or just a series of state machines. Which is the appropriate level? I am bound to accept them all (and even unknown levels of interpretations) or throw them all away,

[11] (Shannon 1948).

apart from the physical level. This problem arises from the will to exclude the existence of a subject as the source of the various levels of that machine. Recognizing such dependency would dissolve the arbitrariness of the degree of existence of each level: each level would be like a relation between a physical structure and a subject. Each level may exist as an onphene that takes part in the constitution both of the machine and the subject.

This program can be interpreted even if the machine is not interacting with the environment. The motor and the sensory apparatus of the robot can be disabled but someone might inspect its memory. In this case both its physical structure as well as its symbolic meaning would still be there. Yet, nothing would be happening. The symbolic meaning inside its memory would just lie there as dead. Only the act of someone reading would provoke the occurrence of some event, possibly the right onphene with that kind of content. The physical shape would have no further relation with its internal program. It would behave as a mere passive body: just a carcass of a machine.

However, if someone switched the motor and sensory apparatus of that robot on, it would begin to interact with the environment. The flow of events, that before was unaware of the presence of the robot, would begin to be modified by its presence. The flow of events might begin to be modified not only by the physical structure of the robot but also, in a way still to be analysed, by its internal program. The robot would start to execute certain actions in response to certain events.

When I look at the robot I do not see the robot as if nobody is watching it, I perceive (that is I *am*) the event that brings that particular physical shape into causing my neural activity. I do not perceive a static structure but an occurring event, which literally constitutes my own being. My conscious states are not static properties of static physical structures, like mass or electric charge: they always occur. Nobody has ever experienced a static conscious state. *I cannot know anything without something happening.* This is nothing more than another way of stating the principle of the conservation of meaning and of experience (§ 5.1).

The existence of the robot with its behaviour determines the occurrence of a series of events, which have as their causes both the physical shape of the robot and its internal program (Figure 7-7).

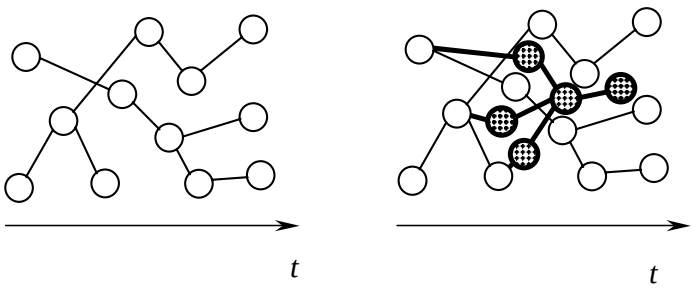


Figure 7-7 Reality is constituted by a stream of causally related events. When a dynamic structure is embodied (on the right), it interacts with the environment determining events that are dependent on its existence.

As conscious observers, at each temporal slice, we would have a perception of its position, disposition of its limbs, and so on. Yet we could not perceive directly the flow of events that are occurring. We perceive (in the sense of being) other events, which have been influenced by those occurring in the robot itself. In reality the robot is constituted by a flow of onphenes that affect our own onphenes.

The occurrence of these events is what might be the basis for the being of that artificial subject. Similarly we are conscious subjects because we are a unified set of onphenes (events) occurring and not because we are a certain physical structure with static properties. A corpse is not the person that was correlated to it Iin the same way the physical structure of the robot does not represent its possible subject. Its internal program is not really there. It is only there when some other subjects scrutinize that physical structure (its states of memory). But when the robot begins to interact with the environment something different is put into existence. A series of events are occurring: they could not have happened without the presence of the robot. These events are what we term the *dynamic structure* of the robot. It is something like a semantic structure. In fact it cannot happen without the external environment. If we look at it, it is difficult to define its boundaries exactly, which are no longer constrained by the robot physical boundaries. This structure does not depend on the external observers either. It exists on its own, under every possible criterion. It does not depend on the meaning we might give to what is happening. It is more real than the physical structure itself, because it brings into existence that physical structure as well as its possible observers (Figure 7-8).

What is *learning* then? Learning is nothing but the procedure by which the static structure of a robot is modified in order to make the occurrence of events in a certain way possible. During learning, changes to the static structure of a system occur. Consequently, different behaviours will result and different events will occur as an effect of the presence of the robot. Learning should affect the physical structure of a system permanently in order to allow it to interact with the environment in the future. Being a subject is always the result of what is happening. However, structures and symbols acquired during learning do not contain meaning in themselves as can be shown by the simple fact that any symbol of a system has a different meaning in different environments. As any word might have a different meaning with different users, so any static structure might have a different role in different environments. A static object (a physical object) has no semantics in itself. It cannot be the basis of a semantic being. On the contrary, any event, being the spring of a corresponding onphene, being its very nature relational, is semantic. This explains at once, why consciousness must be something that happens. This explains also why a subject must be a collection of occurring events. Only in this way, can it exploit its semantic nature.

As an example take a neural network made up of several neural units connected by links of varying intensity. Let 's suppose that learning is limited to changes because of these connections: something that is usually simulated by modifying the connection weights. Let's suppose that this network is trained to recognize characters successfully. At its input it would receive information from a video camera and it would provide a different signal for each of the letters of the alphabet (Figure 7-9). It can be concluded that, after the training, its internal static structure, has somehow become the carrier of the meaning of the letters. Yet if this network were brought into a different environment, stripped of its actual connections, and connected to other kinds of devices, the meaning of its activity would be completely different without having modified anything inside of it. As it has already been stated several times, that the meaning is not inside any static structure. If we take into consideration neither the static or physical structure underlying a system, but the dynamic occurrence of events that results from the presence of that system into an environment, no mistake is possible. It would be impossible to transfer *that* dynamic structure without carrying all the semantically relevant events too. This task, practically impossible, would transfer the meaning of each state of the system because meaning and the events, which carry it, cannot be separated since they are the same thing: onphenes.

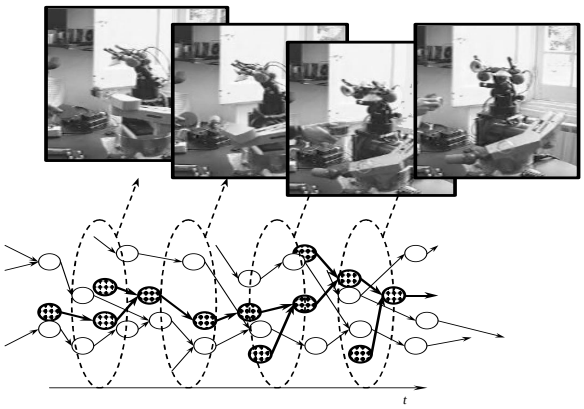


Figure 7-8 At every instant, there are three structures occurrence in our experience of a robot interacting with its environment. What is traditionally considered its physical structure is nothing more than the content of the onphene that corresponds to our perception of it.

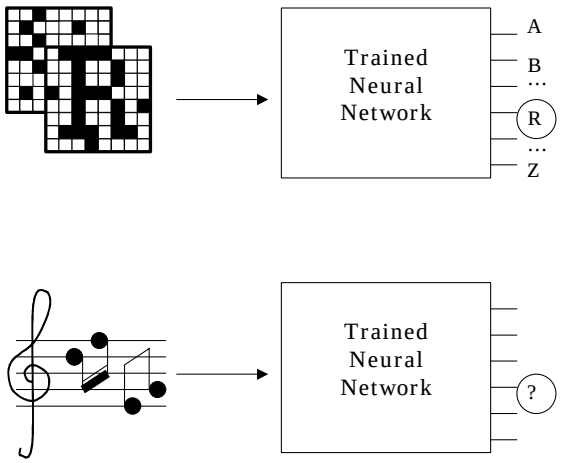


Figure 7-9 A network aimed at recognizing characters of the alphabet (on top). If carried after learning in a different environment (no bottom), the meaning of its static structure would be completely different. The same neural network in two different environments has a different meaning: thus the meaning is not within the static structure of a neural network.

In this paragraph, what kind of structure should be used inside a robot in order to produce the same kind of dynamic structure (that is responsible of our being subjects) has never been mentioned in any way. This issue will be analysed in the following chapters, yet here we can address the debate between procedural languages and neural networks. Both can be used to determine the behaviour of a robot. However, thinking in terms of neural networks, as it will be explained later, makes easier to have an idea of the network of events that will be provoked by a certain system in a particular environment. There is a natural correspondence between the connections that are produced among neural units and the relations that bound events. Procedural programs can be used as well, but with the aim of modelling the relations of future real events – that is, more or less, using procedural languages to write neural network programs (more on this in § 10.1).

Summary

Neural networks are traditionally used as control blocks. This is derived from the paradigm provided by control theory. Yet this approach depends on the choices made by conscious human designers. It is an approach that seems incapable of producing real motivations and autonomous choices.

Artificial systems possess a derived intentionality while biological and human beings possess autonomous intentionality. The difference is related to their internal structure and to the way they modify the flow of events.

A taxonomy for neural network is proposed on the basis of their capability of producing autonomous motivations and intentional relations. Input-output networks that have an a priori hard-wired reinforcement signal give the simplest case. At an intermediate level there are networks capable of self-organizing their input stimuli. So they depend more heavily on the events of their individual experience. Yet, their reinforcement signals are still hard-wired. There are also networks capable of self-producing their own internal reinforcement signals. Since these signals take place as a result of precise events there is a counterfactual relation between them and the external events. This counterfactual relation is seen as the sign of the occurrence of an onphene.

Neural networks are related to semantics and intentionality, but semantics is linked with the occurrence of events and not with the static properties of neural networks as such.

8 BIRU: Basic Intentional Robotic Unit

… Then we must say […] that such an Engine lives, and could indeed prove its own life, should it develop the capacity to look upon itself. The lens for such self-examination is of a nature not yet known to us; yet we know that it exists, for we ourselves possess it.

William Gibson[1]

8.1 Intentional units

As stated in previous chapters, the most baffling aspect of intentional subjects is their ability to refer to external objects and events. This ability cannot be immediately translated into objective entities. Nevertheless, this is not a problem because these entities do not immediately coincide with reality as it is experienced: they are abstract entities, convenient models within the metaphysical framework of objectivity. Intentional subjects are among such structures that cannot be explained exhaustively by the restricted ontology provided by objective entities alone.

When building an intentional subject, the first step is to determine the conditions in which an intentional relation exists. These conditions were outlined in Chapters 5 and 6. This paragraph focuses on the translation of these terms to a physical structure. It is an attempt to define a basic intentional unit. If intentionality derives from the fundamental structure of reality, it cannot derive from high-order properties. It cannot be the result of complex systems. Rather, it is the other way round. Complex systems have intentionality since they are built up in such a way as to exploit the fundamental intentionality of reality. At their roots there must be basic intentional units. At this point, it is important to stress the difference between the present approach and the belief that an intentional system is the product of the complex interaction of several

[1] (Gibson and Sterling 1991).

systems. These approaches see intentionality as an emergent property that, more or less gratuitously, arises somehow and somewhere from a fundamentally not intentional reality. With this approach, reality is viewed as fundamentally intentional by its very nature so, in order to be intentional, a system must be built up *from the inside* in order to exploit this characteristic of reality.

As outlined in Chapter 5, each onphene has another onphene as its content since there are no other kinds of entities. With respect to a certain onphene, the onphene that constitutes its content is called a *critical event*. Since any onphene must determine a difference in what reality is, there is no problem in assuming that each onphene does the job of an event. It follow that each onphene is a basic intentional unit. From the traditional objective point of view, each onphene results in a causal relation among events. The difference with other approaches lays in the fact that the onphene is not a reductionistic entity; for every event (objective event) there can be more than one causal link to previous events. In other words, the critical event of another event is not necessarily the proximate cause: it can be a much more causally distant event. An intentional unit must exploit this kind of relation between events in order to be capable of referring to the appropriate events.

For example, if a subject wants to be conscious of human faces, a structure referring to 'human faces' is required. In practice this entails that in the brain of that subject, there should be a structure (it does not matter here if such a structure is a distributed or a centralized one) that could permit the occurrence of events that have, as critical events, the external events that are known as being 'human faces'[2]. This static physical structure is not an intentional object in itself. It does not point to anything in itself. It does not contain any meaning. Yet, when placed in the appropriate place at the appropriate time, it permits the occurrence of events that have, as their critical events, the appropriate external events: *de facto* they permit the occurrence of events that refer to the appropriate meanings. When designing the structure of an artificial network, we must bear in mind the difference between the structure and the events that the networks makes possible. The structure is a necessary practical achievement, while the events are the real intentional occurrence.

[2] The feeling of circularity of the last sentence is only apparent. The meaning of 'human faces' does not exist apart from the set of events that correspond to it. Any attempt of removing such meaning (as well as any other meaning) from its empirical *a posteriori* root has proved to be a failure. Thus there is no difference between the events that are the critical event of a mental event and the meaning of such mental event. This conclusion is coherent with the TEM assumption of identifying representation and existence.

What does an event x – *having* an event y as its critical event – mean? It means that y is the necessary cause of x. The happening of y is a result of the happening of x. In other words, there must be events – in the subject's brain – that are related in this way only to those events that constitute their meaning. Locating these events is straightforward during perception, where the event is the perceived object. In the course of objective observation and thought, it is slightly more difficult since the critical event is of a higher order.

Let's analyse two cases: i) a rigid structure that must only learn how to tune its parameters (an eye-vergence control system); ii) a structure that defines its goals according to its experiences. In the first case what is happening inside the structure is causally related to its designers' project and the degrees of freedom, in this sense, are relatively few. Let's think of the signal coming out of a system like this. It will be causally related to some event that its designer has cleverly selected (for example the disparity of a target in front of the object). However, the fact that the output signal is causally related to disparity is not itself caused by anything that has happened to the system. The event represented by the output signal is causally related to the disparity of the target in front of the system but not on the ontogeny of the system itself. The cause of the existence of such disparity-selective detector is not part of the history of that particular system but belongs to its designers' intention. Let's consider the second case: a structure that defines its goal according to its experiences. Using the same example, let's suppose that nobody designed the previous system so that it was capable of producing an output signal correlated to the disparity of the target; due to its internal dynamics, the system produced the same kind of output stimuli. Not only did the system learn how to correlate its output signal to the target disparity but also autonomously selected this output as one of its goals. The designers' role becomes much less significant than before. The designers not have any part in choosing what the system should produce as an output signal: they have not taken any part in the cause that determined it. As this example shows, a system with an equal output can have a completely different causal story, since the causal story is a projection, in objective terms, of the corresponding intentional relations.

According to the counterfactual causal story of an event, two different scenarios can be envisaged. In the first scenario, the output signal occurs as a result of the input signals; the very fact that the output is possible is counterfactually determined by the designers' project. In the second scenario, the output signal occurs as a result of the input signals and the output is possible, not because of its designers' intervention, rather because, during the life of the system, something happened that made that event possible. That "something" constitutes the meaning of the event.

Figure 8-1 is an example of what a fundamental intentional unit is. It is a structure that, given an event, produces conditions under which that kind of event is the critical event of its output. This kind of structure with events occurring at its input, under certain circumstances, produces other events at its output. Whenever this relation is counterfactual, the input event is the critical event of the output event. Such events are the result of an occurring onphene.

More than one technique is possible to achieve this kind of structure. An easy way is described in Figure 8-1b. Let's imagine a structure with an input and an output, in which there are two modules. The first receives information from the outside, as does the second. They receive the same input, but the second module is capable of controlling the output. At the beginning no output is possible; the first module is capable of setting the behaviour of the second module so that its output is related only to one kind of input stimuli. In other words, the first module, after the occurrence of an external event produces a change in the structure of the second module; the output event has, as critical event, that particular input event. As shown in the next paragraphs, this structure can be straightforwardly translated into two neural networks, and one of the two provides the reinforcement signal for the second one.

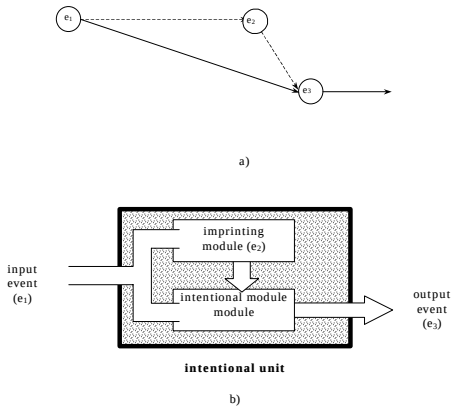


Figure 8-1 An onphene occurs whenever there is a counterfactual relation between two events. In a) there is a counterfactual relation between e_1 and e_3 while e_2 is the event that allowed the relation to occur. b) is the representation of a general structure of an intentional unit. There are two modules: the first has the same role as e_2, the second as e_3. Therefore, in order for an onphene to occur, a structure must modify itself in such a way as to produce, in the future, a causal relation.

If such a structure is implemented it can be iterated as many times as the practical resources allow. But its semantic powers remain unaltered. Moreover, it permits to proceed to higher and higher levels of integration. The crucial point is that there are good reasons to believe that such a structure is a fundamental semantic unit. It is not a semantic unit insofar that it reproduces some features of its target, or because some external observer attributes a particular meaning to it, but because it permits events to happen that have their own meaning as critical events. Here a caveat must be highlighted (Figure 8-2). It is not the physical structure in itself that is a semantic unit: the events, which can take place thanks to its physical structure, are the real semantic carriers. Such a structure allows those events to occur in the appropriate mutual relationship and it can, be considered a *fundamental intentional or semantic unit*. Of course, when it is not part of the appropriate sequence of events, it has no meaning of its own. The same can be said of neurons. When the brain is not working (a "brand dead" brain for example), its neurons loose their semantic property completely (and, consequently, their meaning). There is nothing that links their static physical structure with other objects. They are just dull extensional objects. The events that occur because of such a structure have different properties. They are what we perceive as the content of other onphenes that constitutes ourselves as conscious observers. They are the content of onphenes that constitutes the structure itself. These events are the expression of semantic relations.

If it is possible to build a fundamental semantic unit – although with some caveats due to the aforementioned difference between a static structure and the correlated events – it should be possible to create an ever increasing complex semantic structure: something that is very similar to a subject (Figure 8-3). What is a subject, if not such unity of representations? This final network of events is called *intentional dynamic system, in this thesis*.

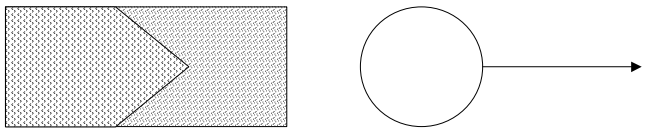


Figure 8-2 In order to create an artificial semantic machine a first block must be created capable of allowing an elementary onphene to occur. The symbol on the left indicates a *semantic or intentional unit*, while on the right there is the corresponding onphene that the unit should allow to occur in the appropriate environment.

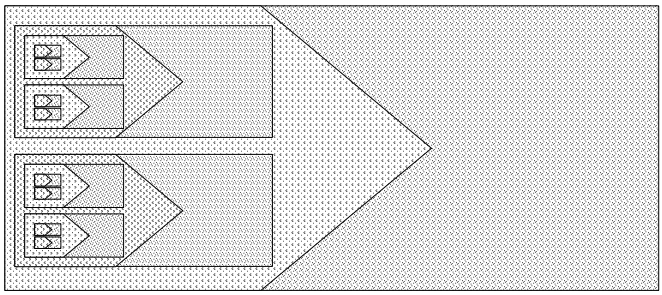


Figure 8-3 Given the first block, a kind of recursive architecture can be applied. It is possible to build a structure whose content is the sum of all previous smaller structures. The final integration of increasing semantic units is the subject.

8.2 BIRU (Basic Intentional-Robotics Unit)

It is assumed that BIRU is capable of i) defining its *reinforcement signals* autonomously on the basis of its interaction with the environment; ii) learning to choose, following some criteria, the appropriate actions on the basis of the selected *reinforcement signals*. These two conditions are the necessary prerequisites in order to create a dynamic structure that is an appropriate cluster of causal relations and that does the appropriate choices: the phenomenal side and the cognitive one.

Of course, given a robotic system, as complex as possible, we are not sure it could correspond to a subject. In our previous arguments, we came to the conclusion that no physical system (neither static like an object nor dynamic like a system) could produce anything comparable to our conscious states. This does not mean that a conscious subject does not correspond to anything real, but only that it is the result of a particular combination of events. If we are able to allow such a combination of events to occur, it will be possible to produce a subject. From this point of view, the physical structure can be seen as an arena, a theatre, where the appropriate events happen. According to an alternative metaphor, reality is comparable to a perennial flowing stream of events. Building things we interact with this stream by changing the relations of events among themselves. Any physical structure exists by virtue of this ability to create a perturbation in the perennial flowing. We can perceive them for this very reason. In this sense a robot or a biological agent is a particular kind of

structure because of its ability to actively interact with the environment. Its structure can be perceived as something that creates perturbations. To get an intuitive glimpse of what is determined by the presence of an agent, let's compare what would happen with and without such an agent. Without it, the flow of events would continue undisturbed while, if an agent were part of an environment, the agent would modify what's going on inside the flow. Let's imagine all these events as continuously converging and diverging; that is: converging when multiple events are the critical cause of just one final event, and diverging when one single event is the cause of many effects simultaneously. Mentally, this situation can be visualized as a flow of light passing through converging lenses and prisms. While the lenses would unify more or less wide portions of the flow, the prisms would divide and spread the stream. A subject can be seen as an exceptionally wide portion of the flow converging in just one ray of light, carrying within itself all the meanings of the previous lights. The physical structure corresponding to the body and the brain of that subject is equivalent to the system of lenses that have determined such a dramatic convergence of light.

This paragraph describes a network architecture that could be the foundation of an intentional robot. The aim is not a new learning algorithm but rather an example of an extended architecture that – using neural networks – could exploit the aforementioned concept of intentional unit. The underlying philosophy attempts to project a simple, albeit complete goal-seeking, environment-driven, neural network capable of self-determining its internal reinforcement signals. Such architecture is simpler than expected and could be applied recursively and constrained only by resources. In the next chapter, this network is used inside a real robot to interact with a physical environment.

8.3 A model of neuron

Neurons are very complex cells. They are the most specialized cells in our body. They are necessary to the 'being' of subjects, as the articulated wing is necessary to the flying of objects.

The study of the neuron deals with several aspects related to its molecular and biological nature: the cellular basis, the plasma membrane, the nucleus, the axon, the synapses, the mechanism of membrane, axon and synaptic potential, not to speak of the neurochemical complications related to the exchange of neurotransmitters and neuromodulators[3]. Studying the physiology of neurons is

[3] (Sheperd 1988).

like studying the South American ecology: it is unique to itself. Unfortunately there is no straightforward relation between the chemical and electrical activity carried on inside neurons and the existence of a subject.

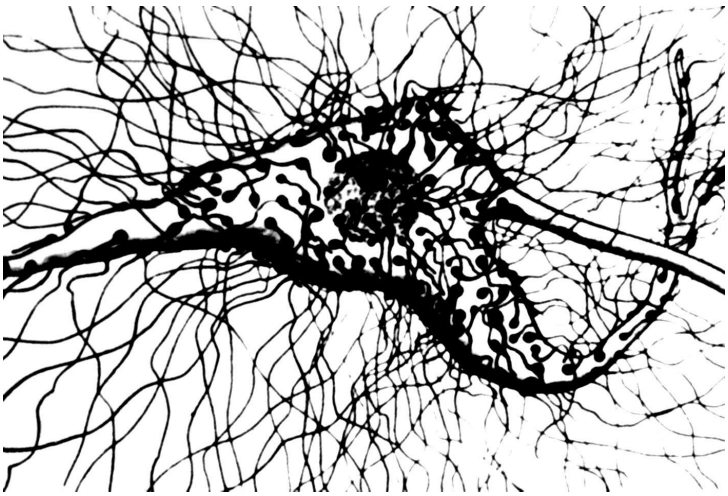


Figure 8-4 The neuron is a very complicated biological machine in which several different kinds of phenomena occur at the same time. Is there anything in this picture that can straightforwardly be related to the emergence of a subject?

Here we are not concerned with the model of neurons as such. They are the fundamental static structure that evolution has selected in order to produce subjects. There is no reason to suppose that they were preselected, at the beginning of evolution, to produce the conscious subject that appeared several millions of years later. To support the opposite point of view entails the acceptance of a teleologically driven evolution. Neurons are interesting in this context since they are capable of enabling those occurrences of events that are called 'subjects'.

Let's think of flight. In the XVth century, many attempts were made to mimick the structure of birds with the aim of building a flying machine. Had scientists tried to master the features of the articulated wing and feathers, we would not be able to fly even with the help of the present day technology. The direct replication of biological flying beings is too difficult since it does not only entail the problem of flight but also the problem of building very light bones, weightless muscles, small power sources, artificial feathery materials, and so on. Building a complete neuron model – in order to create a subject – is pointless as building a complete artificial bird is useless in order to manufacture

a flying machine (Figure 8-5). According to TEM, a subject is an event that unifies other events. Neurons are useful insofar that they help us by creating the conditions in which such events can occur. All their other characteristics can simply be put aside. For this reason, the neuron model used has been extremely simplified. Strictly speaking, it could be a mistake to call it a neuron model because its aim is not that of being a neuron model. What we are concerned with here is to create a structure capable of sustaining a subject. Natural selection kept on using neurons to let onphenes happen in the appropriate way. It is plausible that there is no need to follow the same steps exactly. For instance, in order to fly there is no need to use articulated wings and some times (as with missiles and rockets) there is no need of wings at all. Following the same principle, it is likely that thinking in terms of neurons is highly misleading.

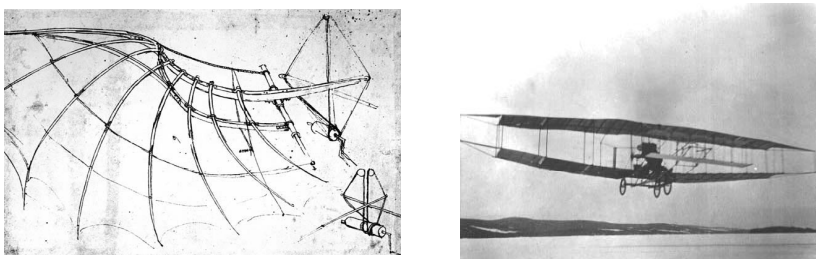


Figure 8-5 The study of the articulated wing as a means to understand the principle of flight could be viewed as a metaphor of the study of the biological properties of neurons as a means to understand consciousness. Although it can be interesting, there might be faster ways.

It is convenient to have a fundamental computation unit as a building block by which to compose more and more complex levels. This has several engineering advantages, among which the standardization of signals, the modularity, and the uniformity of the architecture taken as a whole; the term 'artificial model of neuron' has been used keeping in mind the *caveat* just outlined.

This model is highly suitable for another reason: it is also an efficient way to create a structure in which events can occur. If the subject is an occurring event, it is necessary to have a structure where is possible to interfere effectively with the occurrence of other events. Apart from the compulsory embodiment of such a structure in a real robot, another constraint is the capability of controlling the occurrence of internal events. A fundamental unit corresponding to an event is

needed. The 'artificial neuron' is such a unit[4]. We define each neuron as a unit that will assume a new value each time something happens in the units it is connected with. The assuming of this new value will be considered the 'event'. In other words, the neuron will be considered insofar as it is capable of being the seat of the occurrence of an event and the event will be correspondent to the change in its value. No other constraints are placed on these internal events. Mainly because, as it explained in Chapter 5, an event is something without which reality would have been different: any change is enough to make an event.

This point of view has several practical consequences. First there is no need to be concerned with the exact value contained into each neural unit. What is important is the change of that value. We do not concentrate any effort on the approximation capability of neural units, which do not have to approximate any function. 'Artificial neural units' are needed only in order to create the conditions by which events can occur with certain mutual connections. Second, there is no need to use bipolar units (units with values ranging from –1 to 1): it is not forbidden either. In this implementation, only positive values (from 0 to 1) have been used in order to have a straightforward and intuitive correspondence between the value of a unit and the happening of an event. The concept of a negative event, although possibly useful in more complex and efficient implementation, would have added some confusion without any real gain, at least at the present level of practical realization. Third and last, any unit is constituted by a value.

For the above mentioned reasons an 'artificial neuron' can be presented suitably, as usual, by values determined by its antecedents.

This model has been widely used in most of artificial neural networks. Following the above notation, the relation between the value of a unit and the connected units can be expressed as follows

$$h_j = F\left(\sum_k \left(h_k \cdot w_{kj}\right)\right) = F\left(\overline{h} \bullet \overline{w}_j\right)$$

[4] The same considerations led Francis Crick (Crick 1994) to demand for the term *computing units* to be used instead of *neurons*. The suggestion has been widely accepted (Rojas 1996).

where each unit h_j assumes a value that is function of the sum of the value of the previous units and of the activation function $F()$. As F, it can be used any monotonic function normalized in the range chosen for the units. We used a sigmoid function:

Box 8-1 Real neural units and their software simulation

Usually there is no particular distinction between real neural networks and their software simulation. The main reason for this is that as long as we are concerned with behaviours, there is no difference between the two cases. From a purely behavioural point of view, there is no difference whether the final activation of an arm at a certain speed is due to a real neural network or to its software simulation. On the other hand, if we aim at creating a particular series of events, there could be crucial differences between the two. After all, a software simulation of a neural network is just a procedural program running on some empirical version of a Turing machine. It doesn't matter if the designer obtained the program by thinking in terms of neural networks: in the end there is just a long series of machine instructions in some memory and a processor that is picking them up and doing the appropriate operations. From this point of view the long debate between connectionists and classic AI researchers can have a surprising final: all neural networks hide a procedural nature. In reality it is possible for the story to have another end: all procedural programs hide a structure of fundamental events. The procedural level is nothing more than an interpretation given to a set of physical phenomena. These physical phenomena are abstract entities we build starting from our conscious experience, as conscious observers. The conclusion is that even a program, if embedded in the appropriate robotic structure, can interfere with the flow of events and can let the appropriate events take place. If this is the case, the subject will arise as the appropriate combination of events. It will be coincident neither with a program nor with a neural network, but with what the program and the neural network made occur. In the end the neural network will reveal its true nature of a procedural program and this, in turn, will reveal its true nature of physical connection of fundamental events or onphenes[5].

[5] There are other connected issues that I cannot address here mainly because of the actual level of implementation. The relation between the use of virtual memory and absolute addressing constitutes an example. Recent software techniques create multiple

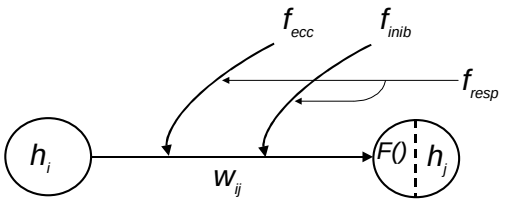


Figure 8-6 This is the fundamental unit that has been used in the implementation of BIRU.It is made up of a value h_j connected with a variable number of other units. Each connection is modulated by a weight w_{ij}, and it receives two sets of signals from other units, f_{exc}, f_{inhib}. Respectively, these two sets act as excitatory or as inhibitory channels. Each signal can be inhibited by a common signal f_{resp}.

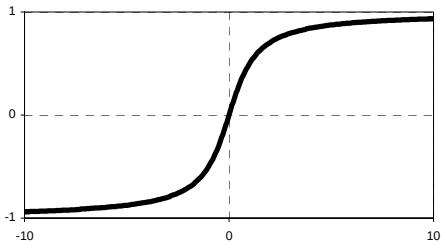


Figure 8-7 The activation function has been used to normalize the input of each unit.

where

$$F(x) = \frac{2}{\pi}\arctan(x)$$

levels of indirection that could be a possible cause of modifications in the stream of events.

Each time a neuron is modified – that is each time that an event occurs into a neuron – its weights are modified according to a fixed law. A rule similar to the Hebb's famous rule has been used.

$$w_{ij}(t+\Delta t) = w(t) + \lambda \cdot f_{resp}(t) \cdot [f_{ecc}(t) - f_{inib}(t)]$$

where $\lambda \in [0,1]$ is a constant expressing the speed of the learning. The goal was not to propose a new learning rule or a faster algorithm, but to use artificial neural network models to implement a structure focused on development.

In a certain sense it is true that a brain is a mirror of the external world. This is true if we stop to look at the external world as ordinarily experienced. The external world can be seen as a set of relations among events. Each onphene constitutes one of these events and embodies its relations with other events. According to this view reality is a set of mutually interconnected onphenes. The brain mirrors the external world onphenes by modelling the neurons in such a way as to replicate these relations inside its own structure. Of course, the web of neural connections is a static structure. It permits the happening of internal events that mirror the external ones. Paradoxically, although the neural structure of the brain is merely a static structure, it is probably one of the best representations of the dynamic structure of onphenes.

8.4 Converging Networks

Neural Networks are traditionally used as computational devices. Their main goal is to achieve the capability of providing the correct output if fed with a certain input. This is a broad generalization but it well suited to the purpose of most researchers. Optimum control, pattern recognition, and function approximation are all techniques that follow the mentioned general structure: there is an input and there is a desired output for that input[6] (see Chapter 1). After fixing the goals of a system there will always be the necessity of this kind of network. The aforementioned techniques can all be seen as function approximators. Here, we propose to term a 'Convergent Network' each network whose main goal is the selection of an appropriate output given a certain input.

[6] It is possible that the output of the network is not known but the effect of that output is known.

The term 'Convergent' derives from the fact that this network usually reduces the dimensionality and the complexity of the incoming signals.

There are several good candidates for this purpose (competitive learning, supervised learning, reinforcement learning, back-propagation). They differ mostly in their internal learning rule rather than in their external behaviour. In the proposed architecture there is no strong reason to use one rather than another. It could be even possible to use more than one kind of network depending on the different constraints arising from different cases.

A simple case of Convergent Network has been used in this thesis. Its main goal was to select the best output with respect to a defined *a priori* signal. Functionally this network belongs to the wide category of reinforcement learning networks[7]. It was constituted starting from the basic building blocks proposed in the previous paragraph. The main features are the following (Figure 8-9):

- Fixed number of input units
- Fixed number of output units
- Fixed number of reinforcement signals defined *a priori*

Given an input vector $\overline{x}$ of size n and an output vector $\overline{y}$ of size m (generally $m<n$), the network should select the best possible outputs with respect to a reinforcement signal r. Each connection links two units (x_i, y_j) with a weight w_{ij}; at each instant there is a defined output vector

$$\overline{y} = \arctan\left(W \cdot \overline{x}\right)$$

where W is the matrix of weight and *arctan()* is the normalizing function at the input of each neural input. The goal of the network is to modify each weight so to maximize the reinforcement signal r. At the beginning the weights are initialized randomly: each weight is assigned a random value between 0 and 1. Then the network begins to produce an output for each input.

A first issue is whether the network should be considered synchronous or asynchronous. From an implementation point of view, it can be used as an asynchronous network. Each input unit can be updated whenever it is necessary. Given the fact that the network is completely feed-forward, the value at its input can always be changed. Any change in the input vector entails a

[7] (Sutton and Barto 1998).

change in the output, too. By using the properties of the neural units defined in the previous paragraph, it is possible to build a completely neural structure modifying its weights in the desired way (Figure 8-8).

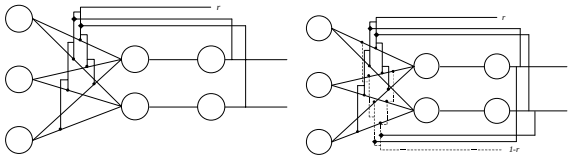


Figure 8-8 A Convergent Network implemented only by making use of connections.

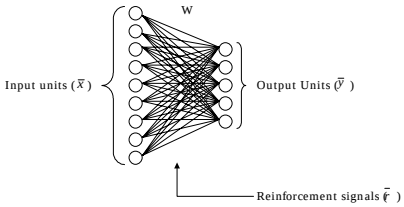


Figure 8-9 This figure shows a Convergent Network. The most simple implementation is an input vector *x*, an output vector *y* and a reinforcement signal *r*.

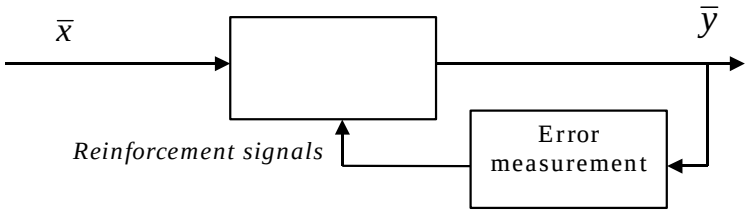


Figure 8-10 Classic neural network used as a control system to obtain certain vectors on a certain unit: neural networks as neural approximators.

Box 8-2 How many connections?

A few considerations must be done about the optimum number of connections in a Convergent Network. The problem is intriguing because it appears to have biological parallels[8]. How many connections should be possessed by a Convergent Network at the beginning of its life? The larger the number of connections, the longer the learning time. In the same way a smaller number of connections will increase the probability of remaining trapped in a local minimum. A missing connection could result in a lost solution. On the other hand, a reduced number of weights entails that their space has a lower dimensionality so that locating their minimum is easier.

In general a compromise solution is feasible. Instead of starting with a network with full connectivity – each unit of one level connected to all the units of the subsequent level – it is possible to randomly connect only a small percentage of the units, and it becomes possible to start the search for an optimum solution with a reduced dimensionality. Later, it will be possible to add new connections while the network is running. The network could also explore new solutions.

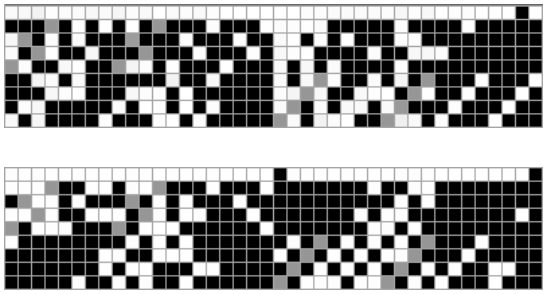


Figure 8-11 Learning with sparse connections (50%) at the beginning and with full connections (100%) at the beginning. As it is possible to see, after a reasonably high number of iterations the differences are not so high.

[8] (Quartz and Sejnowski 1997).

8.5 *Diverging Networks*

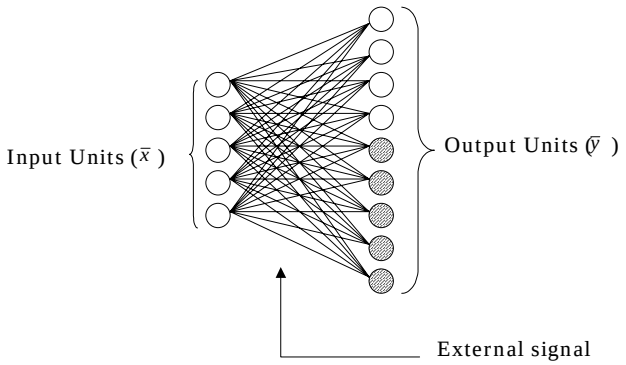


Figure 8-12 A Diverging Network. It receives an input vector *x*.

The aim of the unit called Diverging Network (Figure 8-12) is completely different from that of the Converging Network. The aim of the latter one is to select few patterns according to some reinforcement signals – that is to choose the appropriate actions among a predefined set of options. The goal of the Diverging Network is to add new units to the network: to extend the representational capability of the net as a whole. While the Diverging Network can change its behaviour according to the signals, its activity is completely different from the one of a Converging Network. This kind of network has the following main properties:

- fixed number of input units
- variable number of output units (with a maximum limit)
- external fixed signals
- its goal is to create new units and to associate them to particular combinations of patterns of input units (*flashbulb memory*)
- it must not control anything

The first point depends on several factors both practical and theoretical. From a practical point of view, it is easy to have a fixed number of input units because the network can be easily dimensioned. From a theoretical point of view it makes sense that the input part of a network is determined by the sensor capabilities (the number of receptors of a particular sensor modality). In this respect the input of a network is determined by the hardware possibilities of a

certain system. It must be said that in more complex systems (as it is the case of human cortex for instance), where there are more neural layers, the input of one layer can be the result of a previous processing stage. During development, most layers might change their output capabilities, entailing a modification in the input dimension of the subsequent processing stages. In conclusion, these properties may be modified in the future network implementations although, for the present, due to the limited number of levels, it can be considered sufficient.

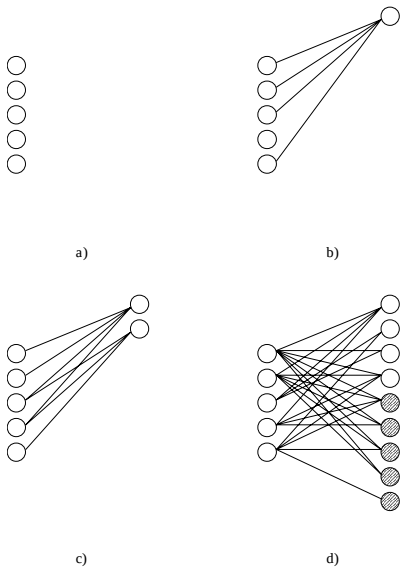


Figure 8-13 In case a) we see a diverging network at the beginning of its life. In case b) after some iterations, a new output unit has been added. It is connected with only some of the input units. In case c) after more iterations, more units have been added. In case d) a more complex state has been reached and there are many more output units than input ones.

The size of the output level must be absolutely variable because it cannot have any *a priori* dimension. The output of this kind of network is determined by the input it receives. It has no *a priori* output and there is no way of predicting what its final output will be. The goal of a Diverging Network is to retain a trace of every relevant combination of the input units. The story of a network is the only assesment of its final state. As shown in Figure 8-13, at the beginning of its life, this kind of network has no output units. Eventually they are added accordingly to the input received. Different techniques can be used to

achieve this goal. A currently widely accepted example is given by the Kohonen Networks or by competitive networks[9]. Here a simpler approach has been used the aim is obtaining a network capable of locating relevant patterns from an input vector. In theory, an infinite representational capability is conceivable. In practice, this means infinite memory. Given this capacity, for each new pattern a new unit could be recruited. We can envisage the following: every time a new pattern is presented to the sensorial input a new unit corresponding to that pattern is created. Given an input of vector $S(t)$ of dimension N ($S \in \Re^N$) of real normalized values (0,1) it could produce an output vector of maximum size equal to 2^N output units. Such mapping would be undoubtedly complete but impossible for most real-case input vectors. From a practical point of view, let's imagine a limited memory resource (perhaps also limited by the velocity in accessing the memory). It is reasonable to assume that the output units have a maximum number and that this maximum is fixed *a priori*. Let's assume $M \in N$ as this maximum. We must thus choose the *best M* output units between the 2^N possible choices.

In general, due to physical limits, an arbitrary threshold of M possible output units can be fixed. In a computer implementation this limit is due to memory capacity, while in a biological organism it is due to the number of neural units. This limit has an important consequence. Given an input vector of size N and a threshold M, each output unit would have to correspond, on average, to a capacity C_u corresponding to the number of different patterns.

$$C_u = \mathrm{int}\left(\frac{2^N}{M}\right)$$

Of course, his is only a theoretical capacity. This capacity entails that all the 2N combinations would appear in the input vector. This is not realistic for several reasons; first and foremost that in real cases there would not be enough time. For example, in a real visual input, the retina has 10^6 receptors. Even considering each of these units as a binarized value, it would mean a total of $2^{10^6} \cong 9.9 \cdot 10^{301.029}$ units that equals, at an input rate of 30 Hz, a time-span of $10^{301.021}$ years: many more than the estimated life-span of the universe. Another reason is that data is usually clustered around certain combinations and not

[9] Classically these nets are used to map a low dimensional space into higher dimensional space. For example a two-dimensional vector is used as an input of an n-dimensional output network. The goal of the network is to find the best cluster in the two-dimensional input patterns. The case we are proposing is therefore quite different.

uniformly distributed. A different, and more realistic average capacity C_u, is the following.

$$C_u = \text{int}\left(\frac{R \cdot T \cdot S \cdot C}{M}\right)$$

Where R is the frequency rate of the vector input, N the size of the input vector and T the expected lifetime of the considered system. C is an additional term, which takes into consideration the distribution of vectors. C is in the range of 0 and 1. If a system is receiving input vectors that are distributed in a perfectly uniform way, C is equal to 1, while in real cases C can be very low. For example, in the human visual system R is equal to 30 Hz, T (in human beings) can be set to 20 years[10] and C can be considered equal to 1 in order to set a maximum threshold. It follows that the quantity at the numerator equals $2 \cdot 10^{16}$. As shown in these examples there is a great difference between the number of possible combinations of a given input vector and the number of possible output units. It is extremely important to establish a criterion in order to choose the best possible output units. How to choose the *best M* output units and how to define the criterion will be dealt with in next paragraphs.

Any *a priori* criterion should be rejected because it would contradict the conditions that have been set in Chapter 1 and in § 8.1. It is essential that any criterion be established following an interaction with the environment. Given these constraints, the technique that will be used in allowing the network to grow is of extreme importance. The activity of this kind of network can be summarized as follows:

- All vectors are normalized real numbers. They span from 0 to 1. In practice they can be easily binarized (threshold with a sigmoid function) without loosing too much information.
- Si is the ith input vector, so S^0, S^1, … S^n is the sequence of input vectors.
- I_{output} is the set of all output units (at the beginning $I_{output} = \{\emptyset\}$)

[10] Of course T can be seen as the total life time or as the critical learning time (youth). Therefore T can span from 4-5 years to 75 years, more or less.

-	Each time a new input vector S^i is presented, the network must decide whether to add a new output unit.
-	If (D_1 & D_2 & D_3) then a new output unit is added to I_{output}. The new output unit is defined as to produce the maximum value for S^i.

D_1, D_2, and D_3 correspond to the three conditions that will be examined shortly. An open issue is the relation among these conditions. Must they be present all at the same time, or is just one of them, sufficient to justify the assignation of a new unit? As yet, there is no straightforward answer. In biological systems, there is no clear prevalence of one solution. It is possible to imagine situations in which both strategies would be advantageous. In this implementation we have chosen the solution of requiring all three of them. It is a conservative strategy that necessitates the most extreme conditions to add a new unit. Given the fact that resources are usually much more limited in artificial systems than in biological ones, we opted for this solution.

The three conditions have been labelled as

- D_1: Relative similarity (a new unit corresponds to an input vector that must be significantly different from the input vectors already represented by output units).

- D_2: An *a priori* learning curve (there might be *a priori* factors distributed along a time curve that could influence the probability of adding a new unit)

- D_3: Significant stimuli (the input vector could appear simultaneously or nearby a particular kind of signal)

It is important to note that only the last condition requires the presence of an *a priori* system of reflexes or instincts. The previous two conditions can be completely autonomous with respect to experience. In other words, the first two conditions do not require any kind of phylogenetic bootstrap.

8.5.1 Relative similarity (D_1)

Each time a new input is presented to the network, it is important to understand if previous output units already represent that combination. Given

the fact that resources are limited, it is important to recognize the degree of similarity between an input signal and the vector already stored into the output units. More than one criterion can be suggested and different solutions can be proposed. In the present implementation a simple Hamming distance has been used.

Therefore, given

$$dH_{\min} = \min\left(dH\left(S^i, I^0\right), dH\left(S^i, I^1\right), \ldots dH\left(S^i, I^k\right)\right)$$

$$0 \le dH_{\min} \le n$$

$$I^k \in I_{output}$$

Where dH(x,y) could be defined as the Hamming's distance between X and Y. If

$$dH_{min} > \tau n \quad (0 < \tau < 1)$$

then a new unit is added to I_{output} as to represent S^i (n is the current dimension of I_{output} and τ is a parameter arbitrarily fixed that sets a similarity threshold). If τ equals 1, it means that no vectors are accepted because they would be of a higher dimensionality than the input itself. If τ equals 0, it means that all vectors would produce a new unit in I_{output}. If τ equals ½, new units would be added only if at least half of its components were different.

Obviously, dH could be defined differently as long as it is a monotonic function with respect to similarity among vectors (Hopfield Network, Hugh transforms or self-associative network are suitable examples). As it is clear from Equation 8.x, there is no constrain on what dH() is.

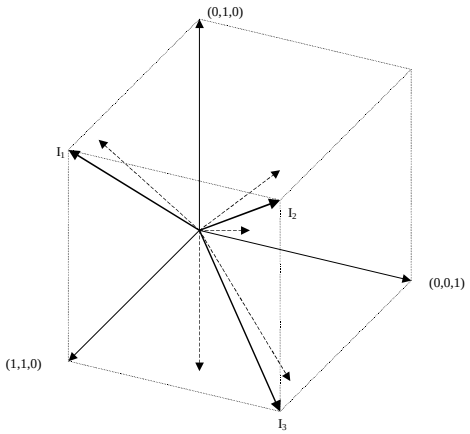


Figure 8-14 A three-dimensional input space with a series of input vectors (dotted line) and the three resulting vectors of the output units (I_1, I_2, I_3).

Box 8-3 Different kinds of mappings

A question that might arise is: why should we use such a dimensional mapping instead of a more efficient one from a computational point of view? For example, using a new dimension for each features of the input could seem a waste in terms of dimensionality. Let's imagine the following case: the input is made up of the position in space of an object with a resolution of 3×3 position in space. Let's imagine a grid of that size. A two dimensional vector could be used. Perhaps a mapping like Kohohnen or a Voronoi tessellation could be used to select a portion of the map.

As seen in Figure 8-15, the same physical phenomenon (on the left), the appearance of a target in a portion of space, can be mapped in at least two different ways: using a two-dimensional vector $\boldsymbol{v}$ (in the centre) or using a nine-dimensional vector $\boldsymbol{s}$ (on the right). The former is the usual choice in robotics because of its higher efficiency, reduced memory occupancy, and better correspondence with the numerical representational system of computers, while the latter seems to suffer from redundancy. Yet, in biological systems there are neither numbers nor vectors and there are plenty of connections. Besides, the higher dimensional vector can be immediately translated into a set of connections (each component can be seen as a binary connection between previous receptors and subsequent neurons). There is a another reason that is more compelling than these previous ones: by using

the representation provided by the higher-dimensional vector it is possible to focus on events. In other words, using ***s*** it is easy to have a direct correspondence between external events and internal events. It becomes feasible to build a neural architecture with a goal: to permit the occurrence of events in a precise causal relation.

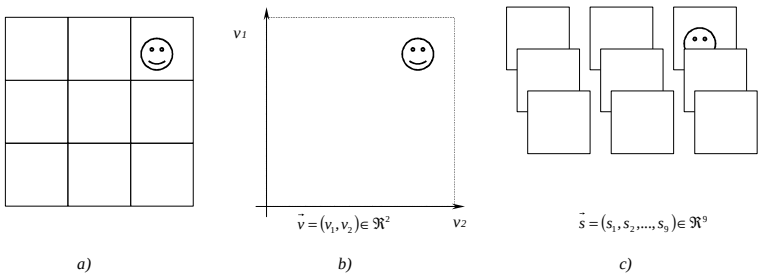


Figure 8-15 Different ways of decomposing an external event

8.5.2 A priori learning curve (D_2)

Another point that must be analysed concerns the uniformity in time of the learning activity. There are several biological organisms that manifest a different attitude both to learning and to adding new reinforcement signals. Famous examples are *imprinting* and the process of maturation. During imprinting there is a limited time scale during which young birds choose a particular visual stimulus as the archetype of their mother. Later they are unable to change it. They will follow this kind of visual stimulus for the rest of their youth. If, during the same critical period, a different object is presented to them they would follow the wrong object.

Correspondingly, in all complex species there are one or more temporal windows in which particular kinds of behaviour can be accepted or rejected. In human beings puberty, adolescence, and maturity are suitable examples. In an artificial being, different temporal windows can be chosen in order to select different kinds of events. For example, maturity is traditionally defined as a period in which people do not change their mind very easily, while in the course of adolescence people should begin to understand what they really want. Everyday commonsense might be the explanation of the presence of these temporal windows in our developmental constitution.

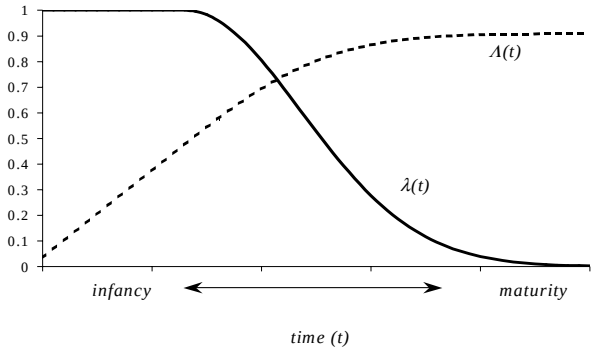


Figure 8-16 $\lambda(t)$ is the probability that an experience would result in a new unit inside the system. $\Lambda(t)$ is the acquired percentage of the total 'knowledge' the system will gain during its entire life. (1-$\Lambda(t)$) is the amount of experience the system is still capable of acquiring.

By using a function like the one shown in Figure 8-16 as a probability for the creation of a new unit, the growing of a Divergent Network can be modulated in time. In other words, during the early stages of development if the conditions are suitable the probability $\lambda(t)$ that a new unit will be added to the network equals 1: all units will be added. On the contrary, after some time, $\lambda(t)$ decreases the probability of adding a new unit. After a suitable quantity of time (a critical t_c), $\lambda(t)$ equals zero and the system becomes stable and unable to add new units. Any new event will be treated by making use of the old combinations. If $\Lambda(t)$ is defined as

$$\Lambda(t) = \frac{\int_0^t \lambda(\tau) \cdot d\tau}{\int_0^T \lambda(\tau) \cdot d\tau}$$

where T is the total life-span of an individual and t is actual time. Since $\Lambda(t)$ is monotonically going from zero to one, we can use to distinguish between stages of infancy and the maturity. Arbitrarily, infancy can be defined as the period in which 50% of the total capacity of the system is still to be used ($\Lambda(t)<½$); maturity as the period in which less than 50% of total experience is still to become part of the system ($\Lambda(t)>½$).

Box 8-4 Time and learning (exploitation and exploration)

Do we always get the best experiences at the beginning of our life? Unfortunately not, and frequently learning is ill suited to what has to be done later. Animals that have a limited period of learning are in danger of encountering situations that are drastically different from the pattern they learned during their infancy. Maturity is defined as the period in which an organism has ended its developmental phases and in which further modifications are limited if not absent at all. Two possible strategies are conceivable. According to the first one, it is possible to set a limited period aside for learning and leave the rest of the life of a system to the utilization of what was learned in the first period. Alternatively, it is possible to think up a system that is capable of continuously changing its goals according to its experiences. This trade-off is sometime called the exploration-exploitation dilemma. After all, studying and working are two different activities.

In extremely complex organisms like human beings, the richness of neural connections permits a very prolonged period of infancy. Besides, according to many, immaturity never really ends, always leaving a limited capacity for further development. Using the notation of our system, this might be restated by saying that $\lambda(t)$ never decreases completely to zero. It could be argued that this method would not be very well suited in many situations because it does not take dynamic environments into account. In fact, if an environment were not stationary, any new event that might appear during the maturity of a specimen would most probably result in the system being incapable of coping with it. For example let's look at the following situation. Imagine M patterns that, at the beginning of the life of a particular system, present themselves many times in such a way to fill up the capability of a system completely. Let also imagine that, later on, the system comes in contact with different sets of patterns that are presented many times, but are not as numerous as the first set. The system would become incapable of learning anything new. How is it possible to solve this?

Apparently there are only two possible options: providing systems with larger and larger neural resources and leaving open the possibility of discarding old patterns and recruiting old resources for new events. Human beings seem to exploit both strategies.

8.5.3 Significant stimulus (D_3)

There are a number of special situations that can be known *a priori* at least in some respect. For example, when we are in pain or overwhelmed by some events. Or maybe when we are under the influence of some low level ancient neurological subsystem like the amygdale. Some signals (let's think of intense pain or intense pleasure) are extremely powerful in modifying a subject's goals. It is highly probable that a subject will try to repeat a course of actions that ended in intense pleasure, and in the same way a subject will try to avoid repeating a process that produced pain.

There are several reasons why learning cannot be left to itself. The two most important concerns the limited time and the bias imposed by natural selection. Let's briefly analyse each of these two points. If a system must learn to recognize particular kinds of situations (in a limited time) it is important for it to acquire indications about what the relevant stimuli are. This is particularly important in natural environments where there are many, potentially highly dangerous, situations. If something produces pain nobody is willing to try it twice. Many species have limited neural capacity, so it is important for them to store the most important stimuli in their neural structures. Similar considerations can be made for the natural selection bias. Every species is selected by nature on the basis of a specific ecological niche, which it is well suited to both for physical characteristics and for its behavioural skills. Given the fact that the neural structure (and thus the behaviours) is not directly described by the genetic code[11], different strategies are used by natural selection to replicate behaviours. One of these is the presence of bootstrapping signals that, if activated inside the appropriate environment, produce the corresponding kind of reinforcement signals. One example is the mentioned above *imprinting* phase. It is not cost effective to store visual details of the mother's shape in the genetic code; besides, it could be a too rigid procedure (the mother might different accidentally from the coded image). It is better to provide some simple instincts that are activated by external events. Another example is sexual attraction. In many species the sexual characteristics are a subset of what is considered attractive by single specimens, after the process of maturation. The more detailed properties of mature sexual objects of interest are a product of the interaction between the simpler instincts and the

[11] The neural structure is described only in the most primitive animals (insects, fish, reptiles, small mammals). In more complex animals, the complexity of the neural structure outstrips enormously the storage capability of the genetic code.

experiences made during youth[12]. It is clear that a complex organism should be able to interact with unknown events and unknown situations; it must be able to produce new kinds of behaviours in response to new kinds of events.

Does this solution deny the principle that everything arises from the interaction with the environment in order to produce the appropriate kind of causal relations? Yes and no. It does because, if the presence of an *a priori* set of signals were to represent the total set of reinforcement signals, no event might be in counterfactual relation with external events. These *a priori* signals act as a guide to the creation of new reinforcement signals that will be the real carriers of meanings into the developing subject. The *a priori* signals work independently from the meaning of what they are selecting from the environment. And they can be easily cheated. If we show a black object moving at the appropriate speed against a brighter background to a frog, the animal will put out its tongue against it, trying to catch it; it can be argued that such detectors are not fly-detectors, rather they are black flying objects detectors. But even this is an arbitrary interpretation. More complex experiments can be built where frog's reflex is activated by events dissimilar to a moving black object[13]. The point is that, notwithstanding how much has been argued to the contrary, these receptors and the resulting reflexes have nothing to do with the meaning of what they are doing. They are merely a structure that, given a certain situation, will act in a certain way. They do not result from the existence of the events they are supposed to detect. Their existence is due to the activity of natural selection that, in turn, is dependent on the past existence of certain events. If they had been able to carry the meaning of something, they would not carry the meaning of the object immediately in front of them but of the object that determined, during the course of natural selection, the existence of such detectors. This kind of relations seems to be too long and weak to be responsible for anything and it can be argued that the counterfactual relation is lost in more than one passage [14]. A consequence of this rationale is that instincts

[12] This could be an explanation of the relative uniformity of sexual tastes inside relatively homogeneous groups of people and the differences among different cultures. It is a proof of the fact that there is a natural bias that must embody itself into the precise characteristics of a precise environment.

[14] Modifications of the genetic code are randomly distributed and they do not act when a new event occurs. In other words, if I survive because of a modification of my genetic code when something new happens, the crucial modification in my code must have happened before, so it could put its helpful influence into effect. Paradoxically, a particular genetic code is loosely related to the events it must interact with.

should not bear any directly content accessible to consciousness: they are unconscious (as indeed is the case). The reinforcement signals provoked by instincts should carry their contribution to consciousness.

Practically all the above considerations can be implemented as follows. It is possible to imagine a limited number of reinforcement signals hard-wired *a priori* in such a way as to speed up the developmental process towards certain categories of events. It cannot be overemphasised that these categories of events are not the direct target of instincts: they are events that, given a certain environment, become part of the relation sphere of the developing subject. Instincts do not point directly to them because they are only automatic mechanisms selected by natural selection with no intrinsic meaning. Rather they become pointers to events and to meaning if embodied into the appropriate structure.

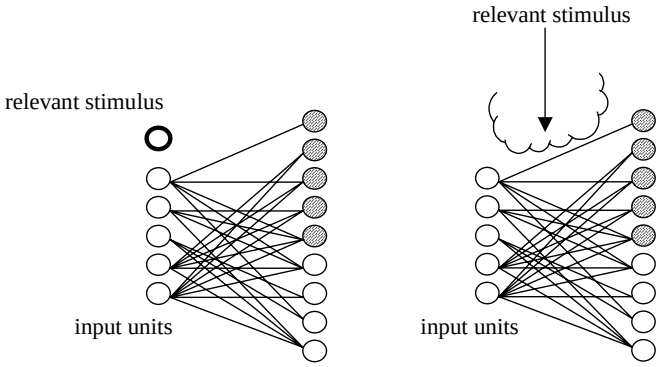


Figure 8-17 On the left, a Divergent Network is shown with an input unit used as a relevant stimulus. On the right there is a Divergent Network that has no input units devoted to any special use but that is receiving an input through separate channels.

This kind of signals can be straightforwardly inserted into a Diverging Network. Different solutions are proposed. Here each possibility will be examined briefly. In the simplest case there is just one signal that shows the occurrence of something relevant. Whenever this unit fires, it means that something important is happening. This situation is often referred to as 'flashbulb memories', in which everything, including irrelevant details, is selected (a famous example is given by the question "what were you doing when John F. Kennedy was assassinated?"). From an implementation point of view there are two different solutions: the relevant signal can be one of the input

units or can be considered as a special case; but the difference is not important if we look at the general framework.
The difference shown in Figure 8-17 that might appear irrelevant in terms of the final result (the growing of the network) is more critical during implementation. It could be done in biological organisms using two completely different mechanisms: one related to neural activity and the other related to biochemical substances. Apparently there is no reason why different channels should or should not be used. The use of a global channel for the relevant stimulus is important because its effect could spread over a large number of units without requiring the existence of a specific neural path. Since the real cause of a negative or positive event might not be obvious, most of the input units should be recorded. A mechanism like the diffusion of a chemical substance is better suited for exerting a global effect. This problem has also a temporal aspect. It might be useful to register events in a temporal window after the first appearance of the relevant stimuli. The release of chemical substances could freely offer the necessary graceful temporal degradation. Besides, it would be coherent both with an asynchronous system as with a biological one. On the other hand, two or more separate mechanisms might result in some drawbacks: mainly the loss of generality of the mechanism. In later developmental stages, the initial relevant signal might have completely lost its initial usefulness and the goals pursued by the system might require completely new relevant stimuli in order to grow accordingly. A complete neural mechanism would be more general because it replicates itself a (theoretically) infinite number of times. On the contrary, a hybrid system might be limited by bottlenecks cause by conflicts or restrictions of its channels.

As mentioned above, the relevant stimulus might represent the boundary between the contribution of natural selection and the developing subject. A simple behavioural example will help to clarify this point. Let's suppose that eating sugar rich foods, carbohydrates and fats is something that is part of the genetic code of average human beings aged 5-10 years. Let's also suppose that in the house of an old woman, say Aunt Mimi, there is a constant supply of these. Let's suppose that "going to Aunt Mimi's" does not belong to the anyone's genetic code. Young Charlie's parents take him to visit to Aunt Mimi0s. While he is there he has a free access to all those candies. Since Charlie has a strong genetic attitude towards this kind of pleasure, he soon becomes fond of "going to Aunt Mimi's": he loves it. Later on, since there are plenty of books in that house, he becomes fond of reading, too. What has happened? The first time little Charlie was taken there he had no particular interest in the place. While he was there he had a particularly pleasant time and therefore he developed a secondary meaning. This secondary meaning was "going to Aunt Mimi's".

Obviously this did not exist before his experience. Such a development was the consequence both of what had happened while he was there and of the fact that he liked candies. We can imagine that eating candies provoked the release of some simple mechanisms that diffused the chemical equivalent of the *relevant stimulus* into his cortex. Thanks to this, a representation of the situation that the presence of candies had provoked has been created into Charlie's brain and he has now recorded it as a reinforcement signal on its own. Whereas he thought of Aunt Mini he was observing he loved the place. So much so that he got used to what he could do there: read books. Did the same differentiated channels, which produced the first secondary reinforcement signal, mediate this second passage? Or was the tertiary reinforcement signal produced exclusively through neural structures? The point is only of practical importance because it is related to the limitations of the single channels in spreading their effects to other neural structures.

The generality of the approach is confirmed by the fact that, up to this point, an exact definition of the origin of these signals has been postponed. Different options are possible:

- The relevant signal is generated inside a system by some hard-wired neural structures (a biological example is given by ancient structures like amygdale and hippocampus; see Figure 8-18a).

- The relevant signal is generated by external events but transduced by specific receptors that propagate it to the other nets (for example tissue damage and the related pain; see Figure 8-18b).

- The relevant signal is generated externally and is internally propagated (forced learning; see Figure 8-18c).

- The relevant signal is not originally present in the system but it is produced afterwards starting from primary relevant signals (aunt's Mimi's example; see Figure 8-18d).

Box 8-5 Temporary buffer

The three conditions already described have the common goal of speeding up net growth, maximizing the probability of selecting relevant system input-combinations, avoiding the use of resources for irrelevant combinations. In order to succeed, biological systems exploit the plasticity of neural networks succeed. In artificial systems a potential drawback is given

by the different technology employed. In biological neural networks, it is relatively easier to embody ambiguous cases (for example, in partially grown neural arborisations). In order to be able to emulate the same capability in an artificial network, different techniques must be implemented. Here a simple algorithm using a weighted temporary buffer is described, by using the same notation of this chapter:

- A set of vectors is added ($I1_{output}$). At the beginning it is empty, $I1_{output}=\{\varnothing\}$.
- For every vector S^i if (D_1 is true) then S^i is added to $I1_{output}$ with a real number a^i such that
- $a^i(t_0)=a_0$ with a_0 such that $0<a_0<1$, (t_0 is the time in which S^i is added)
- $a^i(t+1)=a^i(t)\cdot(1-k)$ with k such that $0<k<1$
- I_{output} is constituted by the m vectors $S^i \in I1_{output}$ with the m greatest associated real number a^i
- With a set of infinite size $I1_{output}$, there would be a perfect result. Nevertheless it is possible to obtain good results even with a finite size m_1 $(m_1>m)$

Thanks to this algorithm it is possible to bypass many problems caused by finite implementation. Its main advantages are related to the control of a greater number of virtual output units than those that are effectively implemented. Learning can go on forever, with new vectors, waiting in the temporary set $I1_{output}$ until they reach such a high score as to be admitted inside I_{output}.

A possible approximation is derived by the introduction of the time value. We can assume that any activation is subject to fading. We can assume that there is some function of time that reduces the value of C and that these patterns whose value become lower than a particular threshold can be discarded.

We can assume the following situation: if a pattern has a lower threshold than X it is discarded and therefore it frees its resources, if a pattern has a threshold lower than X, but it is inside the M circle, it is kept until a better candidate is proposed. Obviously when a candidate is proposed for the first time, it must have enough time to be selected (a few different methods can be implemented, the simplest consists in giving a value greater than X as first activation). So becomes possible to have a structure with finite resources, i.e. an approximation to the first one.

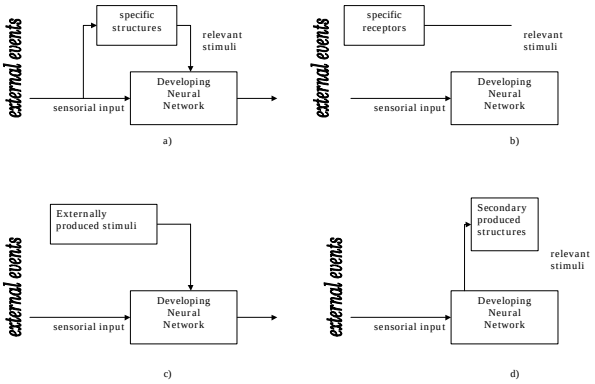


Figure 8-18 Four different functional roles for the relevant stimuli. In case a) relevant stimuli are generated by specific structures like amygdale; in case b) they derive from specific receptors causally related to specific events; in case c) they are externally forced in the system; in case d) they are produced by the network itself after the first stages of development.

8.5.4 BIRU Network: different levels of development

It is now time to use the previously described elementary blocks (convergent and Divergent Networks) to implement the intentional units that will become the original nucleus of a subject. As we will see the process should take place in a real robot, where it would undergo a series of practical constraints. Our goal is to look at the global architecture not as something that should perform some definite actions but as a portion of the environment that could find its unity. In other words, BIRU should endorse the occurrence of events in a certain connection.

However, BIRU's goal is linked to the capability of interacting with the environment. It is not an abstract structure. To achieve this result, it is useful to follow several stages of development that permit BIRU to catch the target. For a series of reasons, outlined elsewhere[15], a passive network would not be capable of reaching all the interesting events of an environment. Here we propose a process composed of three stages of development. The idea is that the system embodying BIRU is capable of doing something from the beginning. The status of an intentional dynamic system is reached gradually.

15

The developmental stages proposed here closely match the three categories described in § 7.3. There are good reasons for this. Natural selection solved the same problems faced by as robot designers. Besides, the ontogeny of each intentional biological subject matches the natural selection stages required to produce it. Each of these stages could be a goal in itself, without requiring further evolution to subsequent stage. It is just what happens in many animal species that do not require a more complex development. Another issue is that each stage endorses a certain level of autonomy and completeness, in the sense that the system is capable of performing a coherent set of actions necessary to survival. Clearly the early stages are more limited in this respect as is the case with nature, where infancy is a period in which animals (and human beings) have a very limited autonomy. Each stage is conceived so as to prepare the ground for the next one. This is the explanation of many reflexes that, during development, have the sole purpose of training capabilities that will be exploited in subsequent stages. Development can be seen as the necessary series of *a priori*, controlled steps with thanks to which a global architecture reach its final status of an intentional dynamic system.

As previously stated, three stages are suggested as the necessary first steps. It follows that more complex steps would be needed in order to obtain intentional dynamic system with increased capacities[16]. At the first stage no intentionality is present (Figure 8-19). It is just a working stage to get the system to exploit some elementary actions and to interact with the environment albeit with limitations. This stage consists of a simple Convergent Network, receiving a fixed series of stimuli, as an input, and performing a fixed set of actions. It receives a reinforcement signal defined *a priori*. It is a kind of structure with no degree of flexibility. It can be compared to the lowest form of animal life. Anything in its own history could determine a difference in its final behaviour. Therefore its output would not have any particular event as critical event. The system would not have any particular degree of unity. Although its unities would cooperate to realize the goals defined by its reinforcement signals, nothing could be the counterfactual result of some event belonging to its history. Yet this network is already capable of performing a limited set of behaviours that could serve as a basis for a active system's interaction with the environment. Besides, this kind of network learns rapidly. The output of this

[16] Suitable examples are the ability of humans to be self-conscious or to experience thoughts of highly abstract nature. Nevertheless, here, the first step of consciousness is identified with its intentional nature – that is the capacity to represent a unity (an event) as a certain content. This capacity that endorses simple consciousness can be addressed in a relatively simple way.

network is hardwired in the sense that its outputs are connected to a fixed repository of behaviours, furthermore for the more compelling reason, that its goals are predefined.

A slightly more complex network corresponds to a second stage of development (Figure 8-20). This network is, at the beginning, exactly equivalent the same as the previous one: a Convergent Network with predefined reinforcement signal and fixed input and output. Nevertheless this network also possesses a Divergent Network. At the beginning, this second module has no output units. In time, it starts creating new units, which are associated only to precise combinations of the input units. Of course this Divergent Network could be provided with one or more of the three growing mechanisms described in § 8.5. In particular it could have or not have the relevant stimulus signal. Its presence is helpful to add a bias to its development or to speed it up, but it is not mandatory. After a while, depending on the kind of stimuli to which the network is a subject, new units will be produced. A second Convergent Network would then be capable of making choices on the output of the Divergent Network. This step is not to be underestimated. The Divergent Network produces an output, which is caused both by the characteristics of the network and by the events that have occurred. From now on, the behaviour is no longer the mere of its project, but also of its life. Different networks, in different environments, would select different combinations of stimuli. Their final behaviour would be different and possess a different meaning. As long as the Divergent Network produces enough output combinations, the first network can be bypassed by the new combination of the twos. This is an important aspect of these stages. The less complex one, that helps to set the condition for the development of the following, precedes them both. For example the first stage guarantees a certain degree of movement from the structure that embodies the network and ensures wider varieties of visual stimuli. These most complex stages keep on adding new units, and the subsequent Convergent Network learns how to connect these units to actions until the usefulness of the first Convergent Network begins to fade up eventually to disappears completely.

The advantages of this second stage are several: i) it creates new units; its growth depends on the input events, hence on subjectivity; iii) it permits two levels of development; iii) it has an intermediate level of representations; iv) it can be biased by reinforcement signals, yet it can develop towards unpredictable ways. Its main constraint is the incapability of defining its reinforcement signals autonomously: it cannot develop its own motivations and goals. We have defined the intentionality of events, but as the counterfactual relation with previous critical events, there are no examples of this relation in this network. Although events having their proper counterfactual causes must

certainly occur, the causes do not integrate as critical causes for the development of the structure. We cannot speak of the occurrence of a dynamic intentional structure.

In the third stage of development, intentionality begins to appear (Figure 8-21). It is composed, as before, by the previous stages (not shown in figure) but it makes a different use of the Divergent Network output. Instead of using its output as a simple set of perceptual categories from which to select the proper combinations between stimuli and actions, it uses its output to produce new reinforcement signals. Each output of the network is employed as a new reinforcement signal for other networks. Each new network is fed with the same input as the first one and it belongs to the convergent or to the divergent kind. In this way the subsequent network would select events on the basis of the reinforcement signals selected by the first network. Their outputs will have, as critical events, those events that have occurred during the life of the network. These signals would be constitutive to the future growth of the network in such a way that future events would integrate the signals themselves in their structure. The network, at this point, is implementing those *intentional fundamental units* described in § 8.1. Each combination is made by a self-defined reinforcement signal and a new network, is one of these units.

The structure is the basis for the occurrence of a series of events integrated in such a way as to constitute an *intentional dynamic structure*. Obviously the focus of such activity would not be static, as it was supposed in the case of the Cartesian theatre. Rather there is a continuous series of feed-forward chain of events always finding a new focus in the network structure. This third stage can be seen as a true intentional structure in the sense that it allows the flow of events to integrate progressively to find its own natural unity. The intentionality is given by the fact that there are no static structures that artificially and externally must refer to what they should mean. On the contrary, in this case the static structures do not have to carry any meaning. *They only allow environmental events to find a possible unification. Such unity is the basis for an artificial being.*

The third stage is not the last one, of course. Other combinations, which are aware of the intentional structure of events, are still conceivable. Yet, the three stages, described here, are sufficient to obtain a simple example of an intentional structure and to explain how it works. They can also be combined together. For example it is possible to use the second and the third stage as to produce many separate channels with new combinations in order to feed the new networks. The problem of creating architectures capable of making the intentional flux reflexive, has still to be addressed. Nevertheless, the first step

was surely the capability of implementing a structure whose main aim was to let intentional relations or onphenes occur. BIRU aims at being such network.

Finally, one last remark about a characteristic of the third stage of development. It may seem that this architecture is feed-forward as there are no internal feedback pathways between its modules. It is not like that because the environment closes the loop. What is aimed at, here, is the birth of a peculiar structure of event, not the implementation of a particular static structure. The events that occur inside the network are only a small, albeit extremely significant, portion of the total set of events constituting the intentional dynamic system.

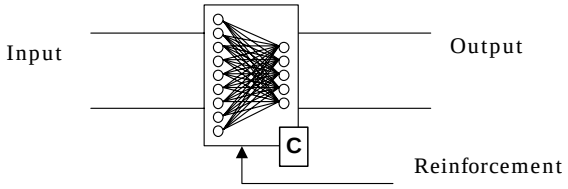


Figure 8-19 The simplest case. At the beginning only a Convergent Network controls a series of simple actions. There is almost no semantic.

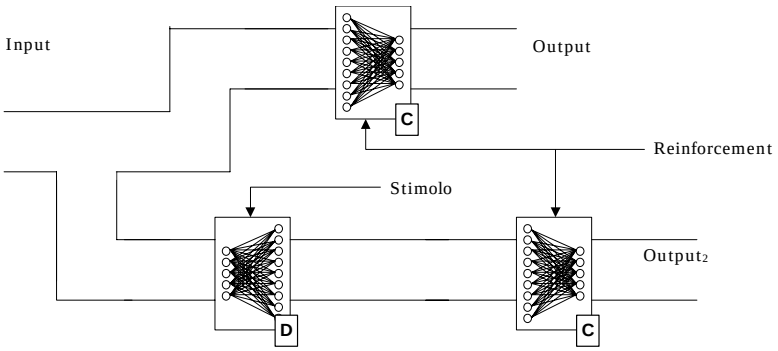


Figure 8-20 An intermediate stage of development. A converging network working on the output of a diverging network substitutes the first elementary Convergent Networks. Some semantics begins to appear.

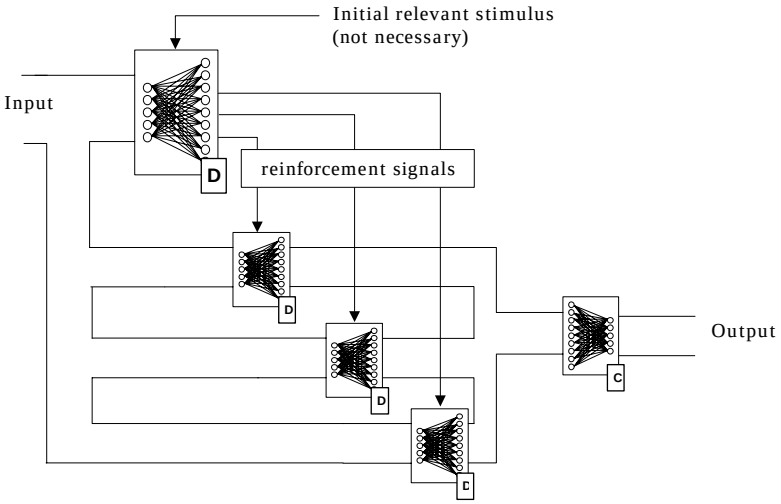


Figure 8-21 Third stage of development: a Divergent Network generates as many networks as it has units. Each of these networks is of a convergent kind and acts as to focus on a precise kind of stimulus. Their output is then used as the input of a final Convergent Network that should select the action to perform. Each Convergent Network corresponds to an *intentional unit*.

Box 8-6 BIRU-BIRU

In order to test the network, a small simulator has been written. Its only goal is to provide an amusing artificial environment by which to verify the formal (syntactical) properties of the network described. An interesting issue regards what the content of its intentional state would be, if such a simulator were to be perfected up to a complete environment. Here intuition fails to provide a clear answer. We might surmise that what is now perceived by human users as funny coloured blobs on a screen, is completely different from the critical events that are the critical events of its mental events. Therefore its content would be as different from what we perceive as our mental content is different from the qualities of neurons. Due to obvious limits of the artificial environment (the poorness of input stimuli) the BIRU-BIRU simulator was used to test only the first two stages of development described here. The third stage required a real embodyment into a real robot (see next chapter).

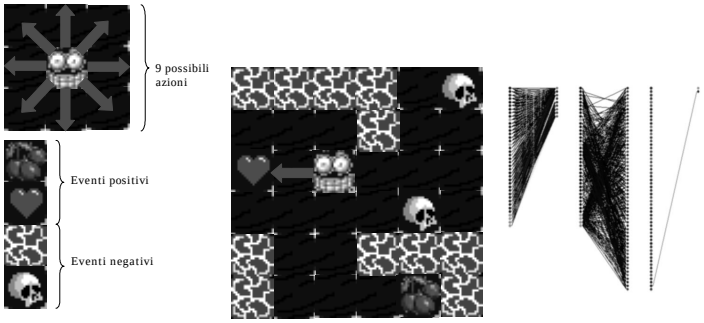


Figure 8-22 BIRU-BIRU in its artificial world

The following are the parameters of the networks tested in the simulator (for a complete comprehension refer to previous paragraphs):

a_0 initial coefficient for temporary output units buffer
m, n maximum size for Divergent Network output
$\lambda(t)$ *a priori* probability of adding a new unity
t learning coefficient
k temporal fading
$dH()$ similarity function
g new connections growth speed

Summary

A network capable of self-selecting its own reinforcement signals is proposed. This network should act as a entanglement in the flow of onphenes. Intentionality cannot emerge as a product of the interaction of complex systems since it is the most elementary property of reality. Yet a system must exploit intentionality from its very beginning.

We propose to use basic intentional units composed of two modules: an imprinting module and a representational module. The units are called Basic Intentional Robotic Units (BIRU). These units should implement a counterfactual relation between the events they let occur as their outputs, and the events that have become their causes. To obtain this result we propose an architecture that resembles biological neural networks. However we must be careful to implement only those aspects that are essential to let onphenes occur.

Two kinds of network are possible: Converging Networks and Diverging Networks. The first ones select the optimal output for a given problem while the second ones self-organize their output on the basis of their input stimuli. They can be modified and controlled by means of three general criteria: similarity of input pattern, *a priori* time learning curves and a tuning signal that acts as a reinforcement signal.

By combining these networks, it is possible to obtain an architecture that self defines its own goals on the basis of its past experience. This architecture is termed a BIRU Network.

9 I, Robot

You think that I am a monster, but you are wrong. You are completely wrong![...] I am a real robot. I am made up of steel and gears, not of flesh and blood. [...]

Eando Binder[1]

Biological beings are capable of doing a number of activities in their environment. In doing such activities they are xploiting different degrees of intelligence. However, their skill cannot be evaluated without referring to some criteria. Even from their internal point of view, there must be some criteria in order to know what they want. It is possible to refer generally to these criteria using the terms *goals* or *motivations*. As it is possible to note, there are no objective ways to define these motivations. All plausible candidates do not have in themselves any compelling power to be universally accepted. Why should animals survive, for example? Why should they try to find the best mates or look for food? The only way to find a rationale for these behaviours is to look for other equally unsatisfactory reasons. For example, I can justify the search for food as a necessary condition to be in the best physical shape and to be able to get better partners of the opposite sex. I have founded one motivation on another one. I can justify the latter by saying that having more and better sexual partners enhances the probability of having more and better children. But again this is not a solution. Why should someone be willing to have as many genetic sons as possible? Again I have to face an unanswerable question. There is no end to this chain of justifications. If I looked for a purely objective, third-person, rationale to explain biological beings' (and our own behaviour), I would not find any satisfying criteria.

If we look at ourselves when we want to get something, we have to face a different picture. Our goals do not need any justifications. They are felt as they are: natural desires of getting something independently from any other considerations. Looking for rationales to justify our behaviour is a habit that we acquire with the age. Children want what they want only because they have those impulses. They do not need any rational justifications. Adults often believe they are able to justify their behaviour, only because they never try to follow up the chain of justification to its origin. In such a case they would

[1] (Binder 1939), p.9.

discover that either their motivations are circular in nature (that is, I want to do *x* because in doing so I'll get more *y*, which will permit to do more *x*) or that they must admit that there is a set of unjustifiable motivations. What is a *motivation* then, if it cannot be reduced to anything else? It cannot be described by an *if-then* clause because it would fall into the previous problem of the infinite remainder. A motivation is very similar to what is usually referred to with the term *'value'*. Thus a motivation is a value. It is something that must be pursued by virtue of its intrinsic quality. But how can we define a value without a conscious subject able to represent it and therefore able to have experience of it. 'Value' and 'conscious subject' are two deeply related terms: it is impossible to understand the one without the other.

For example, is it meaningful to say that a stone is willing to fall, just because it is the behaviour that it will follow if all physical constraints are removed? Is it correct to say that a PC "wants" to run as many programs, as it is possible or that a washing machine "wants" to rise or lower its internal temperature? It does not seem acceptable. What is missing is the conscious representation of these actions and their experience to give them value. A conscious subject – a human being – could perform actions just by accident, even repeatedly but without being conscious of them. I can use everyday a deodorant spray whose internal gas is going to provoke ozone depletion in the atmosphere; nevertheless my motivation wasn't to create the Antarctic Ozone Hole but just to keep myself pleasant. Performing an action, both in conscious subjects and in inanimate objects, is not a sufficient condition to have a motivation. It is necessary and sufficient that such an action is lead by a conscious experience of that action experienced as a value. The claim of this paragraph is clear:

> *It has no sense to speak of goals without a complete theory of the conscious subject: without conscious subjects there are no motivations, no goals.*

For example, I can say that my goal is to eat a lot of strawberries because I am fond of them: the taste of strawberries is a value for me. I am motivated to accomplish several different actions in order to reach this result. In other words, as a conscious subject, I am able to generate several sub goals in order to fulfil the original goal. I want to go to Smith's, the greengrocer's at the end of the street, not because I am fond of Smith but because it is the easiest method of getting some strawberries. The original value 'the taste of strawberries' generates a secondary value 'being inside Smith's shop'. In such a way an enormous quantity of sub-values could be generated. However, it is important to observe that it makes sense to term them 'goals' or 'motivations' only because

they are derived from the original and real desire of strawberries. This original desire does not need any rational justification.

It is possible to try to bypass this problem as stated here. For example, by building some clever machines which the designers have equipped with some behavioural model. It is possible to range from the simple robot machines built in the '80s to the much more complicated human-like robots made at the end of the century[2]. In all cases, independently from the level of the external behaviour this machine can exploit, there is no compelling reason to adopt any kind of *intentional stance* towards the machine, apart from the naïve sense with which we can describe what a Walt Disney's animatronics is doing. If its designer has decided everything that a machine can do, that machine is exploiting some sub-goals or goals of its designer. Therefore it has no real motivations.

We can image several degrees of complexity masking the puppet hidden in the circuitry. Let's start from the degree zero, so to speak: the puppet. A puppet has no will of its own, for it must be physically moved by its master. The threads that link the puppet to its master are physical and visible. The puppet can perform some actions but it is nonsense to speak of he puppet's goals or motivations. At a higher level we can image a machine capable of performing some predictable activity as a result of a careful programming: for example an industrial arm. Its designer can define with great precision its future behaviour under certain circumstances. Other suitable examples are given by a washing machine or by an anthropomorphic puppet in a Disney park. In all these cases, the threads between the puppet and its master are invisible. The threads are no longer made of cords but of symbolic instructions written in their hearts of steel and silicon. Nevertheless the threads are still there: these machines are nothing more than very sophisticated puppets. It doesn't matter how similar their external appearance is to that of their human designer, if they can move their human like face to provoke smiles from children looking at them, if they have arms that moves smoothly or well designed flexible legs: they remain high tech puppets. Of course, many biological beings seem to be as predictable as these machines. Insects, spiders, fishes seem to be just machines produced by natural selection. In them, the trade-off between evolution and ontogenesis is completely biased towards the role played by evolution. In other words, under certain circumstances, a spider would does what evolution decided for it. It is incapable of changing its behaviour during its life depending on its development.

Neural networks have allowed adding a new level of complexity to these machines. The first generations of robot had to be programmed carefully: now

[2] (Brooks 1990; Brooks, Breazeal et al. 2000; Scassellati 2000).

there is a generation of robots that learns through experiences. The point is that these robots learn (more or less by themselves) how to fulfil the goals that have been defined by their designers. Let's take a practical example: a robot that has to learn how to navigate. Suppose that this robot is equipped with a set of ultrasound sensors and two wheels moved by motors. The robot should learn to use a signal (or a vector of signals) in order to produce another signal (or vector of signals). The first set of signals might correspond to the proximity estimation coming from a set of ultrasound sensors. The output signals might be the power level of each motor. The task is far from easy. For example, the designer of the robot wants to produce a robot capable of moving without hitting any obstacle or wall at full speed. In order to do this, the robot can use a neural net with the purpose of learning what is the best combination of motor activations in response to each combination of signals coming from the ultrasound sensors. But you cannot get away from the fact that the designer of the robot must determine the goal. What the robot must accomplish had been decided *a priori* and the robot cannot change it. Of course, its learning algorithm might work poorly and, as a result, the signal sent to the motors are inadequate and so the robot will bump continuously against the walls. Notwithstanding this inconvenient the meaning of the output signal would still be hardwired in its structure. Nothing the robot can do will change its goal. In this respect the robot is similar to relatively simple biological beings like insects that always perform what they have been programmed to do by natural selection.

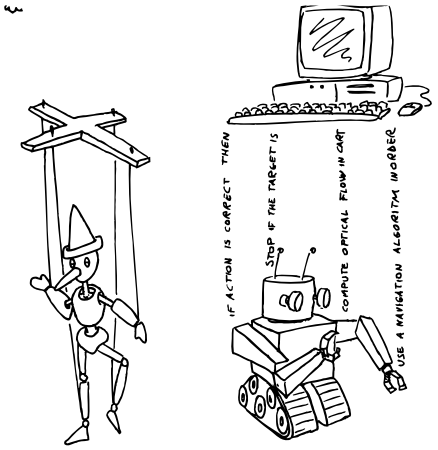


Figure 9-1 Is there any logical difference between a wooden puppet controlled with ropes and a mechanical robot controlled with logical instructions?

A human being is different because of his/her ability to learn to perform actions (in this respect we are often less skilled than many animals). A human subject is different because he/she is able to select new motivations according to his/her past experience. Besides he/she is also able to have a conscious experience of the value associated with such motivations. From a purely functional perspective a necessary step is given by the capability of producing new motivations on the basis of experience. Using the previous example, the robot should be capable of deciding by itself what its goals are. Perhaps the first time it might try to avoid thumping against the walls and another time it might try to hit certain objects. In complex animals it is impossible to code everything that must be accomplished and pursued by them: they are as much a product of their environment as they are of their genetic bias.

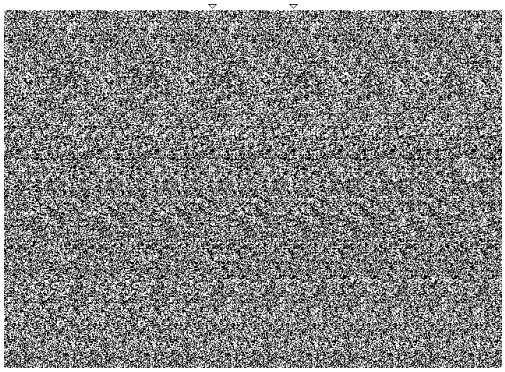


Figure 9-2 A stereogram of an Escher's Ring

9.1 Bottom-up processes versus top-down processes

Bottom-up and top-down processes represent an interesting parallel between these conscious and unconscious processes. The former processes correspond to all those activities that have to be performed quickly and repeatedly, while the latter correspond to actions that are relevant only in particular circumstances and depending on the subject's motivations. For example, the low level control of the eyes vergence angle (in human subjects) is a bottom-up process. It is fast, it works almost every time by always performing the same kind of operation. It receives information about the disparity of the image in the centre of the two eyes and about the radial flow on each eye, and as a response, it produces a

signal that controls the position of the fixation point of the two eyes. It is an unconscious processing going on without the subjects being aware of it. Let's suppose that a subject wants to interfere with it. By doing so, he will try to move his eyes according to his own will. A potential example is shown in this stereogram:

If a subject tries to see the hidden three-dimensional figure in a stereogram she must voluntarily control the vergence angle of her eyes. By doing so, she instantaneously becomes aware of the position of her eyes and of the existence of the low-level bottom-up process that is moving her eyes. In trying to change the fixation point the subject feels a kind of resistance that is due to the activity of her low-level vergence control. She is acting against it. In this she is acting against her low-level activity. She must move her eyes to reach an alternative fixation point that permits to see the three-dimensional figure. Only then she will relax and will return to control her low-level processes. Her activity can be seen as the result of a top-down process. She is taking a high level decision depending on the fact that she recognizes the dotty figure as a stereogram and this, in turn, determines the motivation for changing her fixation point. Clearly this is an unusual situation. Normally the eyes must converge on the surface of an object and not several centimetres further. It makes sense to have an automatic system always busy keeping the eyes fixed on the surface of the object at the centre of the visual field. On the other hand, it is the ability to cope with unusual situations that makes humans so adaptable to different environments. There is a lot of evidence showing that most activities, if repeated regularly, fade into the oblivion of the unconscious: walking, driving, playing a musical instrument, and closing the door of your house. The more regularly they are repeated the less they perceived are consciously. There are two possible explanations for this. First, there might be a physical or functional separation between the implementation of different processes. This separation might locate the processes of one kind in the area in which consciousness is active. A second explanation is that there might be some connection between the fact that a particular behaviour is performed automatically on the basis of a goal *hard-wired* in the agent's structure or, if it is performed following a self-produced motivation, through the personal development of the subject itself[3]

What is important to highlight here is the fact that there is a strong correlation between, on one hand, bottom-up processes, evolution biased

[3] In human beings, the shifting of so many activities from consciousness to unconsciousness could be imputed to the creation of structures functionally similar to bottom-up processes. An example is given by the result on Tetris players (Haier, Siegel et al. 1992).

behaviour and unconscious activities and, on the other hand, top down processes, personal experiences dependent behaviour and conscious activities.

9.2 Emotions and cognition

Traditionally emotions have not been taken into account in the development of robots. The main reason has been the confusion between their cognitive and their phenomenological side. It is important to clarify the difference.

Speaking from a purely cognitive point of view, emotions can be seen as simple devices assigning a global value with no conscious analysis of the details to a particular situation. Emotions are like reward variables capable of representing large collections of external situations. For example, although animals (or particular classes of them) have no conscious experience, they can have unconscious emotions.

Phenomenological emotions, or feelings, can be seen as the conscious perception of cognitive unconscious emotions[4].

9.2.1 James' theatre

Emotions are useful tools to represent a possible future reward or punishment. The idea is that they might be embodied not only in the neural structure of the brain but that they also can be physically part of the body itself. The body is the theatre on which emotions are represented (unconsciously), where a specific situation is associated to a particular body response[5]. Following this method, through emotions, the body becomes a processing element of the cognitive architecture of an active being. This structure has several advantages. One of them is the fact that the system is able to develop a sort of subjective personality and different behaviours in similar external situations. This representation is embedded in the body itself, which after all, is its own best representation[6]. For example, if I have eaten too much it is probable that my mood will change and that my disposition towards different kinds of physical activities will change too. The blood flux towards my stomach has modified my body variables and with them my own internal representation of myself.

[4] (Damasio 1994).

[5] (James 1890, Damasio 1994).

[6] (Brooks 1991).

Indirectly, the body has processed my eating information and, through emotions, will act on how the brain makes future decisions.

More complex emotional responses are obtained as a result of specific neural structure of the activation of a particular emotional response. They can be activated by specific stimuli regarding any relevant aspect of the environment (visual expressions, dangerous situations, philogenetically-selected stimuli, etc.). Among these dedicated neural structures noteworthy are the amygdale, the cingulate cortex and the thalamus[7].

9.2.2 Emotion and reinforcement

The role of emotions in learning is related to the pleasure/pain feeling associated to certain emotional states. This feeling can be exploited to guide the learning phase and, perhaps, to achieve a more efficient performance. In human beings the emotional basis of learning has a cultural as well as an evolutionary side (i.e. the association between a certain body state and some external events may be unconscious). In our experiment the implementation of emotional states and feelings is still rather simple and is essentially related to the generation of a pleasure/pain sensory feedback to reinforce/inhibit specific sensory-motor behaviours.

It is important to highlight that the overall system is acting and learning not only on the basis of exteroceptive and proprioceptive sensory data coding physical parameters, such as geometric relationships, speed of motion etc., but also on internally generated body signals explicitly coding emotional parameters (and not only geometric information about the body status). The overall state of the body (i.e. after having being modified by an emotional response) together with the normal sensations is eventually perceived by the system, which chooses the best action.

We undergo a constant flow of information from the external environment. Low-level modules (colour segmentation, motion detection, optical flow), process the info and send two separate flows to the body (or a representation to the body) and to the higher order perception modules. At the same time the system receives information on the value of what is happening. We can distinguish at least three main sources: 1) from environment, i.e. a hand on a red hot iron bar; 2) from internal values, i.e. a dark spot can be arbitrarily considered something to be avoided; 3) from the body, i.e. a badly moved joint. These events are all mapped into the pain/pleasure structure that has a twofold function. First it has to intervene directly on the state of the body, secondarily

[7] (Lane et alia 1998, Le Doux 1996, Morris 1998, Adolphs 1998).

it has to act on the higher order decision network as a reinforcement signal. This latter decision network has to receive its input from the current state of pain/pleasure system, from the body itself (which is a sort of past memory of the whole system) and from the higher order perception network. Using this information it has to decide the next action to perform with its mobile arm. The low-level motor processing will also be learnt through a reinforcement-learning network.

Concerning emotions, this architecture makes use of somatic markers, which represent the state of the body, as an integrated part of the decision system. They are a memory of the system experiences. They can act as helpers to solve ambiguous external stimuli. They can also serve as modulators of the underlying emotional bias and condition its mood.

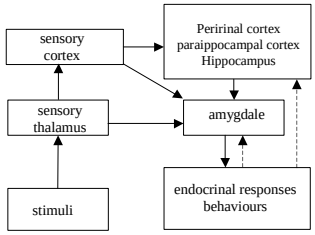


Figure 9-3 Relation between the body, theatre of emotions; the internal world of perception; and the external physical world.

9.3 Babybot and its umwelt

In order to embody the BIRU network, as defined in previous chapters, an artificial body is required. This body is provided by the Babybot set-up, which has been used to test several hypotheses concerning development and sensori-motor integration[8]. The goal of this paragraph is to describe briefly what Babybot is and what its *umwelt* is, and what is its environment. In order to be the centre of a suitable set of causal relationships (projections of underlying onphenes), Babybot must be equipped both with sensory capabilities and motor capabilities. Is necessary for Babybot to be anthropomorphic? Only if we want its *umwelt* to be as close as possible to that of humans'.

As we said in the previous chapters, it is not possible to fully simulate an intentional being because, in such a case, there would be no targets for its

[8] (Sandini and Tagliasco 1980; Sandini, Metta et al. 1997; Manzotti, Metta et al. 1998).

mental states. It would not be a real intentional subject. Intentionality entails the capability of referring to events and these events must be real.

Box 9-1 A Complete Virtual World

Imagine a Karate fighter generated on a computer screen and Let's suppose that a human player can fight against him using a computer: a classic arcade computer game. Since a human subject is fighting against it, let's call it a virtual player; there is nothing remarkable about thinking that the thing, against which the human player is fighting, is a 'computer generated Karate fighter'. But what would happen if we could connect two of these machines (or just two instances of the same virtual karate fighter playing on the same machine)? Would it be meaningful to say that they are still two 'computer generated Karate fighter'? Very probably there would be a computer screen where two Karate players fight one against the other. This might seem hard evidence to their being 'Karate fighters' albeit generated by a computer. But if we removed the computer screen and the software for the graphical reproduction, what would remain? Nobody would be able to see them. The two 'Karate fighters' would be reduced to pattern of bits exchanged between the two computers. If someone analysed the electronic activity going on between them, there would be nothing to relate that activity to the nature of the real 'Karate fighter'. There would be no flesh, no emotions, no strength, and no pain: just patterns of bits going back and forth along the computer wires. There would be no compelling reason to assert that what has going on inside the computer is the simulation of a Karate fighter. The link between the meaning of a 'Karate fighter' and those patterns is, at this point, obvious: it is the fact that those patterns were converted to some electric phenomena on a computer screen and by the fact that such electric phenomena was perceived by human subjects in a way similar to how a human subject would perceive a real Karate fighters. If we removed the human subject who is able to have intentional relations with the appropriate kind of events (real Karate fighters), the link is lost. As a result those patterns loose all meaning. This is another example of the fact that there can be no complete virtual world because it would lack meaning.

At the beginning the system merely receives a set of signals. These signals have no *a priori* meaning. Ideally there should be almost no bias about their use and meaning. This is important because anything that is the result of the designers' targets will not be necessarily part of the developing subject. Each

subject must constitute its structure of relations by itself. Anything that has a particular meaning by virtue of its designer's intentions is susceptible to having such meaning only in its designer's mind. It is important that the structure does not contain any explicit meaning for its signals at the beginning.

Another important issue concerns the sensor modality of the signals. Speaking of the sensor modality of a signal is the same as attributing a particular kind of meaning to that signal. This is reasonable usually because we know where the signal is coming from.

Besides, it is conceivable that the same physical source of signals can be used to provide more than one sensory channel. The visual channel provides an example. From a purely physical point of view, the visual channel can be seen as a stream of values related to the electronic activity of each pixel of an array of photoreceptors.

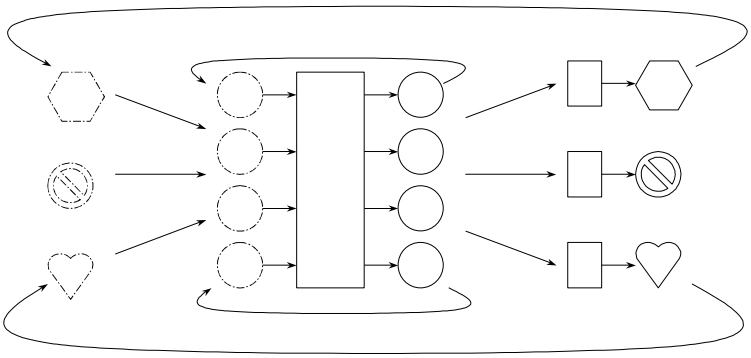


Figure 9-4 As illustrated above we can imagine to apply several processing blocks to an original set of signals. Each one of these blocks applies some kind of transformation to the original signal and extracts a new signal, which is causally correlated with different external events. This new signal can be used as new sensory modality to all effects. The threshold between external and internal is not defined anywhere.

Given these premises, this paragraph describes how we began implementing some of the general principles just mentioned. We started working on a set-up called Babybot whose main aim is to test models of development. The name 'Babybot' derives from a word play on 'baby' and on 'robot'. Ideally, Babybot should be a developing artificial agent. Essentially it is a root with a rough human shape constituted. Babybot has a head with 5 degrees of freedom and an arm with four degrees of freedom (see. Fig. X). The head is equipped with two cameras providing stereo images. An interesting features is the use of human

like log polar retinas instead of normal cameras[9], which give both an increased resolution where the gaze is pointed at and a reduced size of images; secondarily the spring-like model of motor control for the arm, which resembles the human movement[10]. Besides, the robot can move the entire body along a vertical axis. The robot is capable of performing human like movements of the chest, of the head and arm. The eyes perform all elementary eye movements like saccades (fast ballistic movements towards some object of interest), vergence (conjoint movements of the eyes to obtain an easier fusion of stereo images), smooth pursuit (pursuing of a moving target), computation of optical flow, temporal correlation as well as several kinds of colour segmentations. Neural networks performing eyes-head-arms coordination have been implemented. The computational power is provided by a rack of several standard PCs (4 actually) connected to the robot sensors and actuators. Of course, the software implementation of these control systems, as well as their integration, does not guarantee any emergence of first-person phenomena. Every causal interaction between presented stimuli and the actions of the robot is a mere mechanical causal chain (in the old, but still up to date, Cartesian sense). Every group of bits in the computer memory (during the operations of the robot) is a configuration of electronic levels, to which we assign (as external users) a derived meaning. Of course, it might be possible to uphold the same argument regards the biological brain of a living person whose first-person experience is, at least for most of us, undeniable.

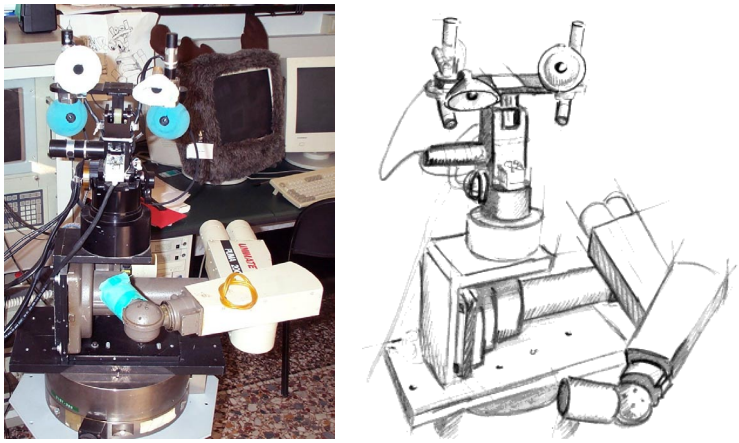


Figure 9-5 Babybot

[9] (Sandini and Tagliasco 1980; Sandini, Alaerts et al. 1998; Sandini, Questa et al. 2000).
[10] (Bizzi 1981; Gandolfo, Sandini et al. 1996).

What is the *umwelt* of a robot like Babybot? How can a robot's umwelt be compared to our own? A brief description of its sensory apparatus is needed in order to understand its environment.

A fundamental hypothesis is that for every external event (belonging to its umwelt) there must be an internal event. In order to have an explicit correspondence between internal events and external ones, we opted for a direct correspondence. Each internal unit ideally corresponds to an external event of some kind. Since each internal unit is causally determined by a set of external events, we consider such a unit to be the internal representation of its causes. The onphene will correspond to the determination of such events. This structure entails a natural discrete representation. Each unit corresponds to a particular event. This is important because we will explicitly renounce to use part of the computational capability of neural networks. For example we will not be interested in storing the exact value of some events in a neural unit. We will be satisfied with the presence or absence of that event: each neural unit might be the threshold.

Box 9-2 Environment, umwelt and enlarged mind.

To detect the umwelt of an artificial or biological being it is important to have a criterion to define what that object is. Since we have defined the umwelt by making use of the TEM, the umwelt corresponds to the enlarged mind of a subject – that is to its set of critical events. Clearly this set is not a static concept, which is not the set of things with which an artificial being comes in contact with. The enlarged mind varies continuously depending on what the content of the mind is at every instant. It is possible for a given being to locate a kind of average enlarged mind that can be thought of as its umwelt. An example is needed to define the difference between the enlarged mind and its more static counterpart: the umwelt. Take a human being: for example Sabrina a doctor. Her environment is made up by all the physical entities that interact with her body: patients, professors, cars, books, food, viruses, and bacteria. Nevertheless she acts cognitively only with a small percentage of them. For example, even if her body is continuously engaged with keeping her healthy by destroying a huge amount of bacteria, she is not aware of this apart from some exceptional cases. People work all the time to provide her, as in the case of any civilized person, with plenty of food, clothes, and drink. Yet she has no direct contact with them. Cellular phones send their electromagnetic signals through her cells but she is unaware of such emissions. Inside of her environment there is a subset of events with

which she is cognitively and phenomenally engaged: cars to be avoided, exams to be passed, friends to meet, food to eat, concepts to remember. These events constitute her umwelt, although she is not conscious of them all the time. At each instant she accesses only a small portion of them. What constitutes her conscious content at every instant is the set of events that are the critical events of her as an occurring subject. This dynamic and ever changing set of events is what is called the *enlarged mind* (Figure 9-6).

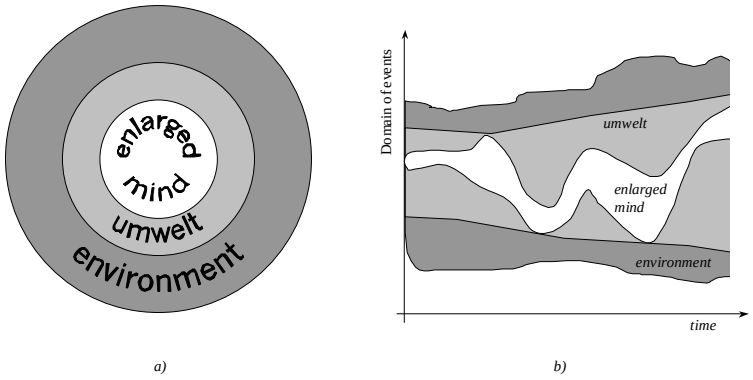


Figure 9-6 The relation between a subject's enlarged mind, *umwelt*, and environment (on the left). The same three concepts seen in their temporal relation (on the right).

An aspect that is never stressed enough is the difficulty to define unambiguously a subject. Given a living organism, defining what is the corresponding subject is still hard. An extreme case is given by distributed beings like a starfish or by a collection of smaller organisms like an ant nest. What is a real subject then? If each branch of a sea star is seen as a subject then there would be as many umwelt as there are branches. If the animal were seen as one whole then there would be just one umwelt. On the other hand, humans are a collection of cells: we could split ourselves into billions of separate umwelts. With TEM it is possible to dissolve such an ambiguity. At every instant, there is an event – the principle of the self – that has, as critical events, all those events that are the content of our consciousness. This is the real enlarged mind at that instant and, as subjects, we are defined and limited by its boundaries. A subject and its occurring content take form at the same time, being the same thing. The umwelt is then built by including all those events that, at least once, have been included in a subject's enlarged mind.

> *The* enlarged mind *of a being corresponds to the collection of its critical events. It is determined by finding all the counterfactual events to its internal events.*

In order to know a being's umwelt and content, knowing what its physical structure is does not suffice enough to know. It is important to know what its actual interactions with external events have been. Besides, we must know whether internally any structure that integrates all those events into a unique event, which has the role of the *principle of the ego* in humans, exists. If a brain was emptied by a central structure and keep alive only with its peripheral and disconnected modules working, why should it be considered as a whole? It would be nothing more than a collection of separate modules reunited in the same physical location.

An artificial being is inadequate in terms of taking into consideration the physical environment in which it is operating, and in terms of looking at the events with which separate modules are capable of detecting. It must be possible to locate a structure that is in counterfactual relation with anything that is claimed to be part of its content. If such a structure is found, then it is possible to determine the subject's characteristics, the subject's content, and thus the subject's *umwelt*, and the subject's enlarged mind[11].

9.3.1 Sensation and perception

What is the difference between sensation and perception? Is it possible to use these terms in conjunction with an artificial being? Summarizing some of the considerations made in Chapter 0 it is useful to remember that, traditionally, sensation has been used to denote what precedes the conscious knowledge of something. Perception is usually connected with consciousness. Nevertheless, it is not frequent to find the word 'sensation' outside biological studies. This is rather confusing since if the word 'sensation' could be used only to denote the causal process that carries information from the periphery of an organism to its central nervous cells, then it should be possible to use the same word to denote the causal process that carries information from the periphery of an artificial system to its central processing parts.

[11] There are several cases in which the splitting of a subject's conscious content would correspond to the splitting of the central counterfactual structure. Two cases are straightforward in this sense: schizophrenia and the syndrome of splitting brain due to chirurgical seizure of *corpus callosum*.

Primary and secondary sensor modalities. Given a sensor sensible to some kind of physical events, we can consider those events as the meaning of the outcoming signals. In other words, there seems to be a simple relation between the kind of sensors and what is caught from the world. In reality things are more complicated because subsequent processing blocks might modify the target of the sensors. For example if we have a CCD capable of detecting a certain range of light intensities in a precise spectrum-window, this CCD will be considered a visual sensor. Let's suppose that a time derivator is added to the end of the sensors. Now the device is no longer sensitive to light values in itself but to light changes. The class of physical events to which the device is sensitive has changed, not because of a modification in the interface between the external world and the device, but because of the introduction of a new processing block (a time derivator) at its end.

The same thing can be observed in human consciousness as a result of the brain activity. We perceive several different kinds of meaning starting from the same physical sensory channel. By means of vision, for example, we do not perceive exclusively intensity, hue and saturation values but contours, edges, shapes, geometrical relations, letters, faces, aesthetic feelings. By means of hearing, we perceive different kinds of noise, voices, phonemes, words, harmonies, melodies, and musical compositions. While in the brain it is difficult to precisely locate an area corresponding to the processing required to extract the appropriate response from the initial raw input, several kinds of pathology are well known to remove selectively a precise kind of content (*prosopagnosia* for faces, lesions to Vernicke's area for words comprehension) from the consciousness.

In robotics it is usual to implicitly treat differently what is implemented by hardware with respect to what is implemented by software. A hardware device, which is capable of receiving external signals and then of providing an output to some kind of physical events, is considered a sensor for that kind of event. If a software module connected to a simplified hardware sensor provides the same output, it is usually called a high level processing or something like that. This is a highly arbitrary conclusion, of course.

In reality, there is no valid reason to distinguish among phases in the process of receiving information from the outside. There is no hardware sensation distinguishable from a software perception (so to speak), there are only structures that permit events (internal to a being's body) to occur in certain relations with external events. It doesn't matter if the intermediate causal chain is implemented in different ways. This allows looking at the issue of the limits

between a sensory part and a perceptive or cognitive part with a different perspective. The first phase can be seen as the *primary sensor modality*: the physical bottleneck through which a being relates with the external world. Afterwards what can pass through it, can be subsequently differentiated through various computational processes. In the hypothesis that such processes maintains counterfactual relations with internal events they, although can be conventionally called *secondary sensor modalities*, are sensor processes at all facts. Vision and early visual processing provide a complete example[12].

Analogical versus digital information. Another important issue concerns the fact that the stream of information coming directly from sensors is not well suited to the development of a subject. The information coming from the external world is usually coded into a series of continuous values (positions of joints, light values, hue values, velocity estimates, and so on). It is true that normally these values are coded digitally and are therefore a mere approximation of the original analogical value. Yet, they still aim at representing such continuous variation of values. If we consider conscious perception of the world, we discover that many kind of information are coded in a much more limited series of events than the actual complexity of the original information. For example, take proprioception. The conscious perception of the position of our body with respect to the vertical axis is not very precise. We do not know exactly how many degrees we are leaning. We have a rough idea if we are standing, if we are on our back or if we are lying down but usually more detailed information is not needed. The reason is that we do not need to be conscious of the exact degree at which our body is. We only need a coarser representation[13]. The same holds for most of our sensory channels. Only a narrow number of events are really relevant for our conscious states. It is reasonable to assume that many low level channels can undergo a phase of reduction of complexity that will reduce the amount of information to be processed. Eventually, as development proceeds, the system itself produces more detailed mapping only for those areas of its experience that may have been related to more interesting events. By using the same example as before (posture of the body), we can imagine that by doing particular kinds of activities (sport, dancing, driving of planes), people must develop a much more detailed conscious representation of the position of their limbs. It is probable that such representation is limited only to those cases

[12] The example resembles David Marr's and Thomas Poggio's theory about vision (Marr 1982; Poggio, Torre et al. 1985; Poggio and Torre 1990; Marr 1991).

[13] This observation is true for our higher conscious states and it is not true for the low level activity of other sub-systems like the cerebellum that need to map many responses as precisely as possible.

that are relevant for their activity. For example a pillar will have a more detailed representation of the position of his/her head than a professional dancer. For legs it could be the opposite. In the following this reduction is described case by case.

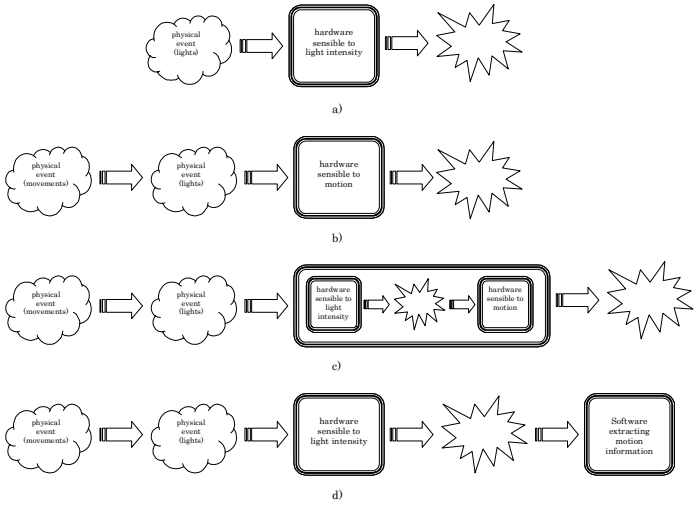


Figure 9-7 Where is the boundary between sensation and perception?

9.3.2 Vision

> *Why it is that we can see so well with what is apparently such badly constructed apparatus?*
>
> Kevin J. O'Regan[14]

The visual channel is perhaps the richest and the most complex sensory channel in humans. In the Babybot implementation an approximation of the human retina has been used. Rather than using classical squared images, log-polar images have been adopted. A couple of stereo images is used.

Studies on the primate visual pathways from the retina to the visual cortex have shown that the geometrical layout follows an almost regular topographic arrangement. These results can be summarized as follows:

[14] (O'Regan 1992).

- The distribution of the photoreceptors on the retina is not uniform. They lay more densely in the central region called fovea, while they are sparser in the periphery. Consequently the resolution also decreases, moving away from the fovea towards the periphery. It has a radial symmetry, which can be approximated by a polar distribution.
- The projection of the photoreceptors array to the primary visual cortex can be well approximated by a logarithmic-polar (log-polar) distribution mapped on a rectangular-like surface (the cortex). Here the representation of the fovea is quite expanded, i.e. more neurons are devoted to it, and the periphery is represented using a coarser resolution.

From a mathematical point of view the log-polar mapping can be expressed as the transformation between the polar plane (ρ,θ) (retinal plane), the log–polar plane (ξ,η) (cortical plane) and the Cartesian plane (x,y) (image plane), as follows:

$$\begin{cases} \eta = q \cdot \theta \\ \xi = \ln_a \dfrac{\rho}{\rho_0} \end{cases}$$

where ρ_0 is the radius of the innermost circle, $1/q$ is the minimum angular resolution of the log-polar layout and (ρ,θ) are the polar co-ordinates. These are related to the conventional Cartesian reference system by:

$$\begin{cases} x = \rho \cos\theta \\ y = \rho \sin\theta \end{cases}$$

Of course, even using log-polar images (which in practical implementation might be as small as 32×64 pixels) the number of possible combinations might be too large for a straightforward approach. A feasible solution might be to use some modules that support some bottom-up processes (optical flow, global disparity, colour segmentation, temporal derivative). The output of these modules can then be simplified arbitrarily.

For example, let's consider the global disparity index: an index that provides a measure of the average disparity of the entire image. This index is real numbers that express the difference between the two stereo images. Such output can be arbitrarily reduced to a number of fixed outputs that are feasible to be the input of a BIRU-like architecture. We can imagine reducing the disparity

output to three, or a few more, output units: objects very near, not so far away objects, very far away objects.

9.3.3 Proprioception

Babybot has a body; so, it must have some sort of representation of it. In a sense, its body is part of its *umwelt*. In biological beings as well as in human conscious beings, it is very common to have an internal representation of the posture of the body. The existence of this kind of perception becomes evident when it is lost. Without this source of information, several subjects become unable to recognize their own limbs[15]. Alternatively, there are patients that continue to feel the presence of an amputated limb several years after the operation[16]. The information provided by proprioception is of two different kinds: the relative position of the limbs and the absolute position of the body with respect to some external system of reference. The second is usually provided through the vestibular system that, in humans, provides information about the absolute spatial orientation of the body with respect to the gravitational axis.

Human subjects receive proprioception from several sources. Babybot receives a limited representation of its own body position. The data coming from the encoders is mapped into a limited number of positions. For example the vestibular position can be mapped in as few as two units: the vertical position and the no vertical position. The positions of joints can undergo the same drastic simplification process . For example each joint angle can be mapped on just two units: an acceptable position and a dangerous position. More complete representations can be made. For example each joint can be mapped on a limited number of positions: extended, 45 degrees, 90 degrees, and so on. The level of resolution of the mapping is of course important to achieve a high level of motor control accuracy. This however is not the goal of the BIRU Network. As in conscious experience we are not conscious of the exact position of our limbs, so there is no need to give BIRU precise information about the position of its joints. The mind is not the cerebellum. There is a big difference between low-level processes (both sensory and motor) and high-level conscious activity. While the first can achieve a great accuracy though it is not suited to govern development and generality. On the contrary the controlling mind is ill suited to control either movement or perception precisely, but it can point at increasingly complex events and therefore achieve a greater level of generality.

[15] (Sacks 1985).

[16] (Ramachandran 1998).

Learning some new motor skill like dancing, skiing, playing a new sport is sufficient to give us an intuitive idea of the level of inaccuracy and coarseness of our motor conscious representation. While we can have a perfectly clear idea of the movements that we must do, we are nevertheless unable to move our body in the desired way. To learn something, we must keep on repeating exercises until our cerebellum develops its bottom-up, low-level fast procedures. Yet these low-level processes, as efficient as they are, are incapable of having the degree of generality owned by the more inefficient cortical conscious representations.

9.4 Development and intentionality

According to our theory, development is crucial to the phenomenal side of learning. The reason is simple. Given that the content of each event with content is always an occurrence, there is no general structure that be instantiated. In this sense TEM is exactly on the opposite side of the rationalistic versus empirical debate inside the representational paradigm. In short, there is an ever-lasting debate in philosophy that regards the origin of our mental states. According to the rationalistic stream, ideas are transcendental, *a priori*, innate, or philogenetically determined; according to the empirical side, ideas are immanent, *a posteriori*, learned during individual experience[17]. Supposing that TEM is right, there are representational units (intentional relations) that are the bearers of content. If they really have contents according to the rationalistic framework they should have such contents *a priori*, in virtue of their intrinsic property. This is exactly the position of the representational theory developed by Fodor[18] a few years ago. On the contrary, TEM assumes that anything has the meaning of what has intentionally (viz. causally) determined it. This means that the same event inside a brain (that is the same neural pattern) can have multiple meanings depending on what its critical event has been, in the history of that pattern. In this sense TEM is at the opposite side of rationalists and innatists within the representational domain. It doesn't make sense to suppose that a neural pattern carrying the meaning of every concept, of every impression, exists in the brain. It does not make sense because it would mean that in the physical domain there are particular objects whose content is different from their own content. This

[17] However the value of such a broad subdivision of the whole history of philosophy is mainly rhetorical.

[18] (Fodor 1998).

duplication of content is at the same time preposterous and costly[19]. On the contrary, by supposing that representation is identical to existence, TEM can omit to create special classes of mental entities. The old problems of empiricists are no more, specifically: how is it possible for the meaning (content) to be transmitted from the outside to the inside of a physical object (usually the brain)? With TEM the mind extends itself gradually to an ever-enlarging[20] set of critical events. The brain is the object where most of the effects of critical events finally occur. The mind does not supervene on the brain even if the brain is necessary to the mind.

The mind of a subject is not identical to its physical structure but to the bundle of intentional relations that make up the corresponding subject. Development is necessary because it is a phase in which content is linked to the internal states of an agent. If a brain were created in a fraction of time (as in the swamp man story) it would be empty of content even if it were cognitively identical to a given normal brain. One example will clarify this issue.

Imagine surgically removing a piece of cortex from an anaesthetized subject and that the piece of cortex were intentionally linked to external events like ice cream. Each time the subject had the perceptual experience of eating an ice cream that piece of brain, would host events whose content would have the taste of the ice cream. The content (*qualia* if you like) was neither inside the particular pattern activated in the brain or in the external events. The conscious content was in the intentional relation whose critical event (its object) was the taste of ice cream. After its surgical removal from the brain, that piece would no longer retain longer any ice cream related content. Why? Because, suddenly, it is incapable of being the place where events caused by ice cream occur. Besides, even when the piece of brain were inside the cortex and was working properly, if it was not instantiating anything as an effect of the external ice cream, the subject could report any conscious experience related to that content.

Development is the unavoidable phase in which the mind can enlarge itself including newer and newer intentional relations. This is a static process. Even when an intentional relation has been instantiated for the first time, its content is always rediscovered every time it is instantiated. There are no static carriers of content (like words on a piece of paper). The content is determined each time those words are read. The content, being an intentional relation among events, is not a static property but a dynamic passage of existence between events. Content is always a difference occurring (ERH) in reality and therefore it

[19] A related critique of what is defined the "myth of self defined code" and the "myth of the explanatory quality" is exposed in (Clark 1998).

[20] Sadly, disease and old age endorse the processes of dissolution of the developed mind.

always requires an occurring event. Without development a system would not be able to become the centre of such a complex web of intentional relations.

9.4.1 Mixed architecture

In building an artificial subject, there are two possible approaches related to different philosophies (Box 9-3). In practice, the difference is between an architecture that must learn everything and an architecture that is ready to operate from the very beginning. Apart from the impact of this choice on the emergence of a real subject, there are a few practical points that must be taken into consideration.

Clearly the human brain is an extremely successful architecture, yet it would be ill suited for a mass production in industries. The reason is quite simple: it takes several years of continuous training to be operational. This fact ought make us think. It means that nature did not find a better way of obtaining a human subject than allotting several years to each individual development. Time is not cheap, even in nature. Time means increasing several times the probability of a subject dying before reproduction. This probability has the highest value in natural selection. Nothing is more important. The time before puberty (time needed to become capable of reproduction) is an unwanted necessity. This time is usually associated with the need of parental care that entail time and energy being spent on their children. A time span ranging from 10 to 15 years is the minimum time required by nature to produce a human subject. It is true that such a long time was acceptable only because there was a social evolution and a group selection and this allowed to increase the development time. Nevertheless, for several years it is a heavy burden on human life. Why did not nature find a faster way of producing human subjects?

There are two possible answers but it is still difficult to know which one is the correct. A first answer simply supposes that for casual reasons natural selection was not capable of finding a better solution. Due to some bottleneck, like the maximum capacity of the genetic code, it was impossible to do in any other way. After all, natural selection has been incapable of producing animals capable of running as fast as the fastest automobiles or as strong as the most powerful diggers. The wheel is something practically impossible for natural beings. A second answer considers that natural selection pursues only those goals that are selectively relevant with respect to other species and that development time must be considered one of those aspects. The time spent by human beings on developing must be seen as a practical limit. According to this second point of view, ten years might be a necessary time-span to let a subject

emerge from matter. If such were the case, the building of a subject would be a longer activity than usual.

Another implementation issue is that the brain hardware is extremely different from that of a PC. Besides, the number of incoming nerves and outcoming actuators in a biological being is astronomical, compared to those of a robot. Are these insuperable constraints? In practice, a compromise solution might be helpful, as Babybot case. There are two extreme cases. The first consists in using a fully hardwired architecture where the designers handcraft module by module, each governing a different behaviour of the robot[21]. The opposite approach tries to mimic the biological architecture from the bottom: an operation that risks at best being a waste of time and at worst imposing such a burden the artificial hardware that any real advancement becomes practically impossible[22]. An alternative approach is the one followed by Babybot.

The basic idea consists is to implement all the low level activities as fast as possible in hard-wired modules. They reduce the computational load of the system. For examples if the optical flow must be computed, by PCs, it is much faster to use mathematical operation on arrays than to simulate a complex neural structure for each point of the optical flow. Such a simulation would use up a relatively high quantity of computational resources. If we try to mimic the neural structure of the brain from the bottom using a different architecture (Von Neumann computers), it is highly probable that we will run out of resources before having achieved anything significant. On the contrary, if we get rid of this structure completely we will no longer have the possibility of having a real subject emerge. A practical solution is to use up to a certain level modules that best exploit the underlying hardware architecture, although by doing so some content might be removed from the experience of the becoming subject. After a certain level the system will do its best to implement a structure capable of letting the appropriate structure of events take place. The conclusion is the following.

---

[21] This is what is usually done in robotic research. Given the emphasis on fast results in a specific task, researchers focus their efforts on inventing and implementing efficient algorithms capable of solving specific problems (Fukui 1981; Crowley 1985; Inoue, Mizoguchi et al. 1985; Jacobson and Wechsler 1985; Brooks 1986; Brooks 1986; Matsushita, Sakane et al. 1990; Santos-Victor and Sandini 1997; Hirai, Hirose et al. 1998).

[22] An example is the series of robots developed at the Centre for the Study of Neurosciences of S.Diego: Darwin and Nomad robot (Edelman 1987; Edelman and Tononi 2000).

It is pointless to waste resources in order to mimic not-essential features of biological architectures.

This solution is not so different from the solution in biological beings. The twofold architecture mirrors the division between the bottom-up and top-down processes (§ 9.1) and the trade-off between phylogeny and ontogeny (Box 9-3). If we analyse the content of consciousness the evidence leads us to the conclusion that biological systems use the same compromise. Separate functional modules, which do not directly contribute to our conscious experience, perform many control activities that have to be fast and continuous. The position of the eyes, vergence control, the integration of vestibular and optical information are usually unconscious. They are perceived consciously only when the information becomes crucial for higher-level processes. This means that they are usually outside the enlarged mind. What happens inside functional modules does not contribute directly to what the subject is. Their work is physically necessary (as it is the activity of the heart, the liver and lunges), yet it is not part of the subject. Their position in terms of the causal chain, which leads from the external critical event to the internal integrative event, is usually intermediate. In other words, they cooperate in spreading the intentional flows. Yet no subject's onphene has them as content. We are not conscious of the chemical activity in our retina or the neural activity in our nervous centres. In human beings the number of levels at which the focus of attention can shift is remarkable. Take someone looking at a painting (let's say the George Gower's Armada portrait of Queen Elizabeth, Figure 9-8). At a higher level the spectator is contemplating the triumph of England over Spain: an abstract concept. At a lower level the spectator recognises the queen, then Elisabeth Tudor, then a face and a female body, then colours and elementary shapes. A separate neural structure (even if not physically separate from the rest) is responsible for the particular content of each of these levels. Does this provide evidence that the structure of human beings is not made of different architectures but totally homogenous?

Another piece of evidence comes from blind-sight cases. These are cases in which subjects, with damaged visual cortex, still have residual forms of perception without any conscious visual experience. Subjects report that they are incapable of seeing anything but, if required to make guess on the position of bright spots, they score better than random betting would allow for[23]. Their cases are relevant here because in so far that they show that an artificial being built up with a different architecture might result in a partially missing

[23] (Holt 1999; Kentridge and Heywood 1999; Marzi 1999).

conscious experience. Blind-sight subjects sill possess some neural structure capable of processing visual information. This structure does not belong to the visual cortex that is, in these cases, severely damaged if not totally missing. There must be some neural pathway that analyses information from the retina and uses it to locate a position in space. This pathway does not provide the subject with content to, thus it is external to the subject's enlarged mind. Why? A possible answer, a mere suggestion, is that it might correspond to simpler processing modules, that help the rest of the more general brain cortex, without explicitly being part of the conscious activity. Are there any possible candidates? Development helps in finding an answer. At birth babies are provided with a series of reflexes that help their visual-motor system to develop[24]. These are bottom-up processes, produced entirely phylogenetically and thus unconscious. If they were the only processes left in adult subjects who have no cortical visual capacities, it would make sense that any information provided would remain unconscious. This could provide the evidence for the existence of different architectures in human beings and for the different contribution that different architectures provide the conscious content.

Although some forms of blind-sight might affect the resulting subject, a mixed architecture is suggested here in order to overcome implementation difficulties (Figure 9-9). The idea is to enrich the primary sensor capabilities with as many secondary sensor capabilities as possible. They are nothing more than fast processes that extract, in whatever way useful, particular features corresponding to particular events in the external world[25]. They exist mainly for a practical reason. Given the Von Neumann architecture of present PCs is not feasible to simulate the biological parallel neural architecture from the very beginning. Besides, as we have seen, it is possible that even biological beings make use of such computational shortcuts. After this stage, the system is capable of feeding the next stage with several sensor capabilities, where a sensor is defined by the causal relation with a category of external events and not by the kind of hardware or software structure. The nature of these causal relations is not necessarily counterfactual, the relation do not necessarily define any

[24] (Cioni, Favilla et al. 1984; Carpenter 1988; Schmid and Zambarbieri 1991; Panerai and Sandini 1998).

[25] They are the early visual processing of Marr (Marr 1982). They have been formalized with the name of *predicates* in Minsky and Papert's famous book about perceptrons (Minsky and Papert 1969). Finally, in the first stage, there is more than one resemblance with the model proposed by Fukusihima in his Neocognitron (Fukushima 1975; Fukushima, Miyake et al. 1983; Fukushima, Okada et al. 1994). Another similar approach is that of John Weng (Weng 1996; Weng 1998).

onphenes. Figure 9-9 shows all sensors feeding the subsequent intentional stage. Do early primary sensors feed the following stage directly (the dotted lines in the figure)? If they do, they could become part of the potentially conscious content, if do not they couldn't. If the second option is chosen, the conscious content might suffer from a form of blind-content for the corresponding events. Of course, causal relations are usually extremely weight regards the quantity of information, so practical considerations might forbid their use. The intentional part, the module that integrates the incoming causal chain into a unity, occurs after this stage. It is entirely made up of basic intentional units (BIRU), which progressively nest in several layers, should bring the incoming causes into one unified effect. This last stage should allow a limited group of final events to occur. These events, as critical events, have all those events that make up the enlarged mind of the subject to be.

Figure 9-8 Queen Elisabeth I, the Armada portrait, 1588 (attributed to George Gower). How many levels of contents are traceable by looking at the painting?

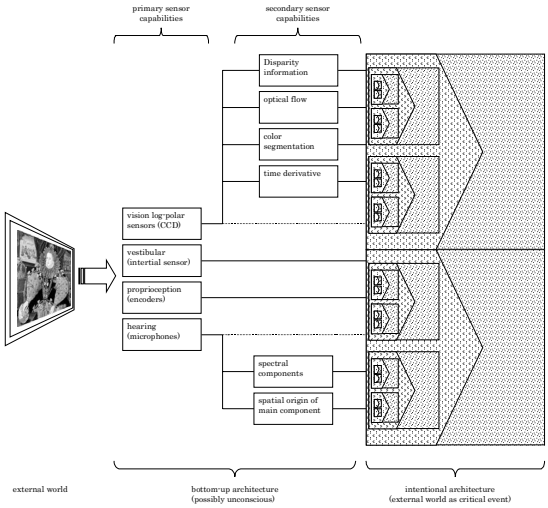


Figure 9-9 The mixed architecture proposed for the development of a conscious robot: at the bottom a level of hard-wired, bottom-up, possibly unconscious modules to do the heaviest jobs; at the top a set of basic intentional units (BIRU) that subsequently integrate up to a final dynamic unification.

Box 9-3 Phylogeny versus ontogeny

As we have seen in the previous chapters, there is a constant trade-off between ontogeny and phylogeny, between the self-organization of a system and the constraint that natural selection or its designers have imposed on it. If systems could be created in one go, there would be no need of development at all. Learning could be carried out once for all and then the results could be copied in all individuals. There could be Lamarckian evolutionary mechanisms in biology: the neural structure of parents could be inherited and then transmitted to their descendants. Yet things turn out to be different.

There are two strong practical reasons that prevent this. First, it is practically impossible to pack the necessary amount of information in a genetic code. Secondly, the packed information loses importance when the environment (the external environment or the body) changes. Development provides a better way to transmit general values and to control the ontogeny of each specimen.

There is a subtle side effect. Each individual is a subject since it is a unified set of representations: a complex counterfactual occurrence according to TEM. The emergence of a subject depends on its real experiences that physically constitute the subject itself. If these experiences were to be removed because they are unnecessary, the subject itself would no longer exist .

We experience only those events that are ontogenetically part of ourselves. Anything that is part of our constitution, but not counterfactual to our being, is not part of our consciousness; it does not belong to our conscious content.

9.5 A brain comparison

The human brain is generally regarded as a complex web of adaptations built into the nervous system, even though no one knows how.

Michael Gazzaniga[26]

If we looked at the human brain we would notice more than one similarity to our architecture. What we'd like to stress here is the relation between the capability of self-creating reinforcement signals and the role of the thalamus in the brain.

The thalamus is an important integrating centre, which receives sensory signals of various modalities, and transmits processed information to appropriate areas of the cerebral cortex (Figure 9-10). An extensive accumulation of axons connecting various thalamic nuclei to practically all cortical areas is seen in fan-like array and this, in three dimensions, reflects the profusion of the thalamic radiations. The thalamic radiations are grouped into four thalamic peduncles. All known connections between thalamus and cerebral cortex are reciprocal, two-way radiations (thalamocortical and corticothalamic). Is there anything in this structure that can support the BIRU structure?

Let's describe briefly the activity of a BIRU structure by grouping its parts into three separate modules that could be called A, T and C. C will contain all Divergent Networks and each of them will be called C_1, C_2 and so on. A and T will contain only Convergent Networks (Figure 9-11). At the beginning the sensory signals are sent to all modules. It is possible to imagine that the information going to A and T is somehow reduced in complexity. This

[26] (Gazzaniga 1998).

reduction is not necessary but it might be helpful. Let's suppose that, at the beginning only one Divergent Network is activated and that it receives its tuning signal from module A. A is receiving sensory information from the outside. However A must act as a bootstrap for the whole system. A contains a pre-programmed hard-wired function of some kind. This function will control the development process of C_1 that will choose those events that are more frequently correlated to the output firing of A. In this way, the first Divergent Network will produce a number of *a posteriori* units whose firing is counterfactually linked to the individual experience of the structure.

When the first Divergent Network begins to take on units, each of them could be assigned to a new Divergent Network. If the system had infinite resources this could a valid way of proceeding. Unfortunately, even biological brains are incapable of such a massive level of parallelism. There should be a structure that chooses which signals are worth using as tuning signals of the subsequent Divergent Networks. This role is exploited by module T. It receives signals from the Divergent Networks and makes choices about which signals are assigned to the following modules. Its goal is to act as a sort of bottleneck in order to reduce the explosion of connections. In the example, T receives the output signals coming from C_1 and integrates them so as to project them back on other areas, let's say C_2 and C_3. The system would be using, as the criterion for its following development, those input stimulus combinations that have been selected as the counterfactual result of the first part of the system experiences. Ideally the system would reiterate the proceeding as long as there are free resources.

Something very similar happens if we look at a schematic prospect of the brain (Figure 9-12). The figure is correct, apart from the fact that the sensory signals, going to the neocortex usually, pass through the thalamus. This is feasible in so far that we hypothesise that the first projections coming from the outside do not receive relevant modifications in the thalamus, at least at the beginning of the development. Things can be different later on. Initially we can visualize some kind of bootstrapping structure that orients the attention of the system towards certain classes of events. For example, babies are more interested in round shaped faces than straight, narrow, objects. Although many of these bootstrapping processes might be physically located in the thalamus itself, we like to suppose that the bootstrap is concentrated in the amygdale.

As long as the information starts to flow, the first cortical area can begin to recruit units in response to specific combinations of input stimuli. The units start to become specialized in terms of what experience + attitudes has brought them. When the cortical areas begin to mature they produce several output units, each correlated to specific *a posteriori* input combinations. At this point,

there must be a structure that narrows their number and that uses them as subsequent tuning signals for the development of other cortical areas. We suggest that such a structure be the thalamus. It must receive from each cortical area; it must make decisions on the incoming signals and, after a phase of integration and selection, must send them back to the other areas. This is more or less what we expect is going on in the brain.

Finally, let's consider one practical possible example (Figure 9-13). Let's consider the visual channel as the sensory input. At the beginning, the number of possible combinations of input stimuli is enormous. Yet, if there were some sort of control, only certain combinations would be considered worth being selected. Something like the biological amygdale or like artificial hard-wired tuning signal would provide control. This module would send a tuning signal to a first Divergent Network. Let's say that round-shaped objects in the visual field activate the module. Every time an object with these properties appears in the visual field, the module fires. As a result the first Divergent Network selects units which usually correspond to round shaped objects. These units do correspond only to a limited class of possible round objects. They belong to the experiential domain of the system. At this point, each unit might become the tuning signal in a new Divergent Network. This is not possible owing to the limited resources available so what units must be chosen? The answer lies in the module T that exploits some general rules and selects only a limited subset from all Divergent Network output units the. The units that have been selected become the tuning signals of other Divergent Networks. The process goes on as long as there are new units to be recruited. It might be possible to introduce techniques that reassign units to cope with a changing environment and which means increasing flexibility.

The idea of grandmother's cells has evoked widespread criticism. The architecture proposed seems to advocate this kind of representational structure. We believe that there might be something true in the idea of the grandmother's cell since our structure is by no means limited to localized representations. In other words, more complex and more efficiently distributed representations might be implemented. The overall structure described here might still be acceptable. It is still not completely clear how many neurons are needed in order to be conscious of a single rich visual image. There are recent results that appear to support the idea that the firing of one single neuron could be enough to provoke the conscious perception of a complex image[27]. Such a result, which

[27] In a first-ever demonstration, UCLA School of Medicine and Caltech researchers have shed new light on how the "mind's eye" works, uncovering evidence that single neurons – individual cells in the brain – are involved in recalling specific visual images to mind.

traditional science cannot explain, might comfortably fit into the framework provided by TEM.

Each unit represents an intentional unit that has allowed events to occur because there has been a counterfactual relation. If the proposed framework (TEM + BIRU) is true, a conscious event is produced whenever a neuron (or a group of neurons) is activated as a result of a process that was originated as a chain of counterfactual relations. The content is the event, which is external to the brain, that is the critical event of the event that is occurring inside the brain. The physical structure of the brain is no longer seen as the physical object that instantiates mysteriously impossible phenomenal properties. Phenomenal properties are no more. In one sense, everything is real[28]. Each neural event carries, as its critical event, its content that is an externally occurred event. Each event occurs because an onphene has occurred. The brain is the place where the flow of onphenes becomes trapped into a knot focusing on higher and higher levels of integration. The final result is the conscious subject that is, more than ever, a *weltknot*.

Figure 9-10 Thalamus and thalamic radiations in the left cerebral hemisphere. In this preparation the corpus callosum, the caudate nucleus, and most brainstem structures have been removed.

They found that single neurons in certain areas of the brain – the hippocampus, amygdala, entorhinal cortex and parahippocampal gyrus – selectively altered their firing rates depending on the stimulus the subjects imagined. Most recently, Fried and his team found evidence that single neurons in the human brain can differentiate between separate categories of visual images, ranging from animals to caricatures of famous people to photos of celebrities. Their «study reveals single neuron correlates of volitional visual imagery in humans and suggests a common substrate for the processing of incoming visual information and visual recall».(Kreiman, Koch et al. 2000).

[28] Everything is real and not physical since 'physical' entails objective.

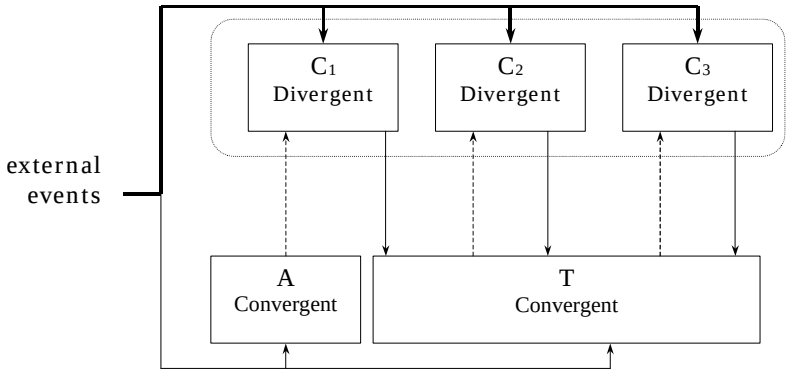


Figure 9-11 A schematic representation of the BIRU structure.

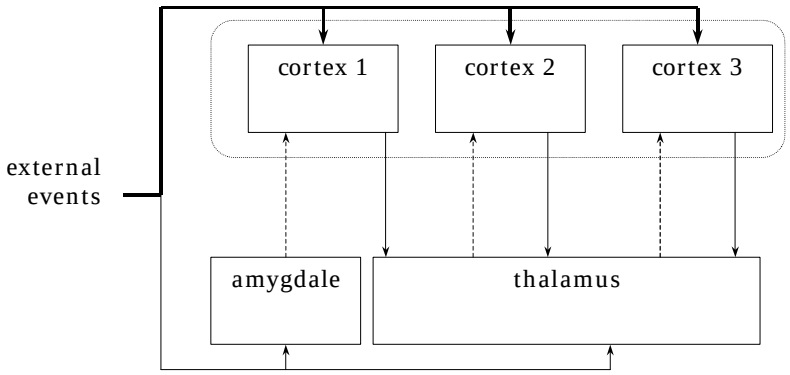


Figure 9-12 A comparison with the known relation between brain modules.

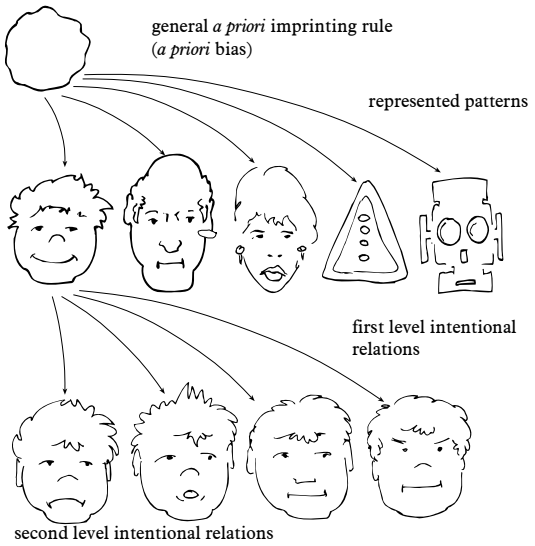


Figure 9-13 A process of subsequent categorization starting from an simple *a priori* hard-wired criterion: the presence of a round shaped object in the visual field.

9.6 Unit, representation and intentionality

As has been stated previously, the mark of the mental is the capability of having unified representations. This capability has been called *intentionality*. In the first part of the thesis we argued that instead of defining intentionality on top of other categories, intentionality must be seen as the fundamental ontological domain. We also proposed that intentionality, as defined above, can also be called *onpheneity*.

A physical structure like a brain or a BIRU network is the static physical structure that lets onphenes interact so that they progressively unify. The static structure is not the onphenes themselves. Let's outline where the unified representations occur in a robot with a BIRU structure.

Initially a BIRU unit is a *fundamental intentional unit*. Each unit, when it is *imprinted* inside a Divergent Network, begins allowing events to occur, events that could not have occurred unless other events had happened. Each unit traps the occurrence of events with counterfactual relations. So each intentional unit becomes the static structure corresponding to the occurrence of an *onphene*; the

issue of representation is thus resolved. Whenever an internal event occurs, it is the last part of an onphene. The representation lies in the fact that the onphene began with an external event – the critical event of that onphene – and ends with the internal event inside BIRU. The representation is identical with the 'being' of the onphene. There is no a distinct mental domain: each representation (each onphene) is, so to speak, both inside the network and both in the external world. The BIRU network is the last part of a longer chain of events. According to TEM the supposed mind of BIRU is constituted by the internal events and by the external events.

Each internal event corresponds to a critical event that represents it. Each internal event is also the unification of what it represents. Since each event occur because of a certain group of critical events have occurred, each event unifies a part of the environmental events. This is an application of what we have defined as principle of unification in § 5.5. Each event, internal to BIRU, unifies those events that have been their counterfactual causes. Of course unification and representation are identical in this respect. Both of them coincide with the being of the occurring onphene. Representation, being in relation-with, and being are all unified in the same kind of occurrence.

What happens when new cortical areas (new Divergent Network) are assigned to the signal produced by previous ones? A more complex chain of events counterfactually determines each new unit. Take vision. At the beginning there are only intentional units that let events occur in response to simple physical events like simple shapes, coloured dots, moving patterns. The mind and the *umwelt* of BIRU are extremely limited. After a while new Divergent Networks are assigned. Each of them selects more complex external events as critical events for theirinternal intentional units. Each new level adds a new level of complexity to the critical events that are the content of the onphene of the emerging mind. And the mind increasingly enlarges itself.

Summary

In a subject's neural architecture two separate classes of processes can be distinguished: bottom-up hard wired processes and top-down self-determined processes. The first group is efficient but unconscious, while the second is much less efficient and it is usually conscious. They correspond to a difference in the neural structure that endorses them. The BIRU Network belongs to the second class.

The BIRU Network is a static structure that can make the flow of event focus into one final counterfactual event that will be the heart of a constituting intentional subject. Of course it must be embodied in a structure that allows for the interaction with real events. A robotic structure must be considered.

Recently emotions have been recently presented as a way of representing the state of the body internally. In our model they are necessarily part of the Enlarged Mind of the constituting subject.

The robot used is called Babybot since it is used to simulate the first stages of human development (*Baby*+r*obot*). Its Enlarged Mind will differ from that of human beings working in the same environment. Its critical events, which constitute the content of its mind, correspond also to what is known as *umwelt*.

Development is intended as the necessary series of steps that must be taken in order to obtain a complete subject. Given a series of practical constraints (limited resources, limited time, limited recover facilities after errors) some kind of endogenous teaching devices must be used. Development is the required endogenous teaching device.

The BIRU Networks in Babybot contains a working example of a real physical structure that modifies the flow of events in such a way as to produce onphenes. These onphenes are the elementary constituents of Babybot's enlarged mind.

10 A thirty thousand page menu

Metaphysics is a restaurant where they give you a thirty thousand page menu and no food.

Robert Pirsig[1]

A few aspects of our theory have already been dealt with by other authors: the overlaps are outlined in this paragraph. Our position is similar to Alfred N. Whitehead's intuition about finding an alternative base for reality that could be a process of becoming; to John Searle's analysis of information and intrinsic intentionality; to Jerry Fodor's regards to the representational atomism of his psychosomatics; to Franz Brentano's regards for the importance given to the notion of intentional objects; to Leopold Stubenberg's regards the notion of the subject as a set of subjective states endowed with content; to Gerard O'Brien and Jonathan Opie's regards the applicability to connectionism; to Galen Strawson's regards the notion of the subjective experience of knowledge; to Michael Tye's regards the identity phenomenal-intentional- representational[2].

We will outline the points where our theories overlap, but also those areas dealing with theories that remind us of TEM because of unusual converging evolutions, but that cover up fundamental differences. Mistaking a shark for a dolphin would certainly have unpleasant results, so it is better to highlight the most important differences with few historically defined positions.

Before examining the most representative of these theories more closely, we will set out a general pattern in order to compare TEM with other trends of thought. The fundamental TEM theory is summarized as follows:

- Nothing exists without representing (A)
- Nothing represents without being in relation-with (B)
- Nothing is in relation-with without existing (C)

Naturally now we can add the dual version of these assertions.

- Nothing represents without existing (D)

---

[1] (Pirsig 1991).

[2] (Whitehead 1929)(Searle 1983; Searle 1992)(Fodor 1987)(Brentano 1973)(Stubenberg 1998)(O'Brien and Opie 1999)(Strawson 1994)(Tye 1996).

- Nothing exists without being in relation-with (E)
- Nothing is in relation-with without representing (F)

Clearly these six constitute a set (Figure 10-1). The central claim of TEM is that each of these assertions cannot be separated from the others and that it is not possible to present empirical examples where only one of the sic is true and the others false. In other words, each of these statements must be true or all of other must be false. The non-completeness of other philosophical systems stems from the absence of one or of more of the theses above.

For instance, dualism, identified here with the usual interpretation of the *cogito*, corresponds to the discovery of (D) because it identifies the intuition that nothing can represent (i.e. be a thought of) without existing.

As a matter of fact, Descartes felt he had to define two substances because thought as well as representation had to be supported by the domain of existence. Non the less this path lead to Cartesian dualism to a fundamental asymmetry, sort of original sin that would influence it for ever: if the thinking substance (*res cogitans*) existed, the reason why the extended substance (*res extensa)* could only be thought of was not at all obvious. When dualism, pushed by mechanicism, collapsed into materialism, it became clear that it would be impossible to bring both thought and representation back into the domain of material substance.

Descartes' thought can be viewed from another standpoint: the *cogito* could be seen as statement of identity between two not so different domains[3].

Berkeley was one of the first to side for (A) defining a form of pure idealism. His *esse est percipi* can be viewed as a complete reversal of the Cartesian *cogito*. TEM accepts a part of these philosophers' thought, but places it in a larger frame.

An interesting comparison can be made with the various interpretations given to quantum mechanics. For instance, according to the Copenhagen School, to talk of events, without taking the problematic concept of measurement into account, dosen't make sense. Without going into the detail the position of the Copenhagen School can be summarised as: «Nothing can be said to exist, if it has not been measured ».

Since measurements are a part of the conscious experience of a conscious subject, the same thesis can be formulated as follows: «nothing can be said to exist, if it has not been represented». Clearly a relation between these two

[3] If this interpretation were to cover the *cogito*, Descartes would have intended affirmed both (A) and (D).

versions of the principle of indetermination exists: a relation that has its roots in the thorny and controversial concept of measurement.

If we limit ourselves to repeating our thesis that states that measuring corresponds to a relation between representations, we can reformulate the two versions of the Copenhagen School like this: «We cannot say that something exists, if it has not been represented or if it is not the result of a relation among representations (measurement)». If this claim is acceptable, it corresponds to (A) and (E)[4]. Even in this case though, the relation between existence and representation has still to be accounted for.

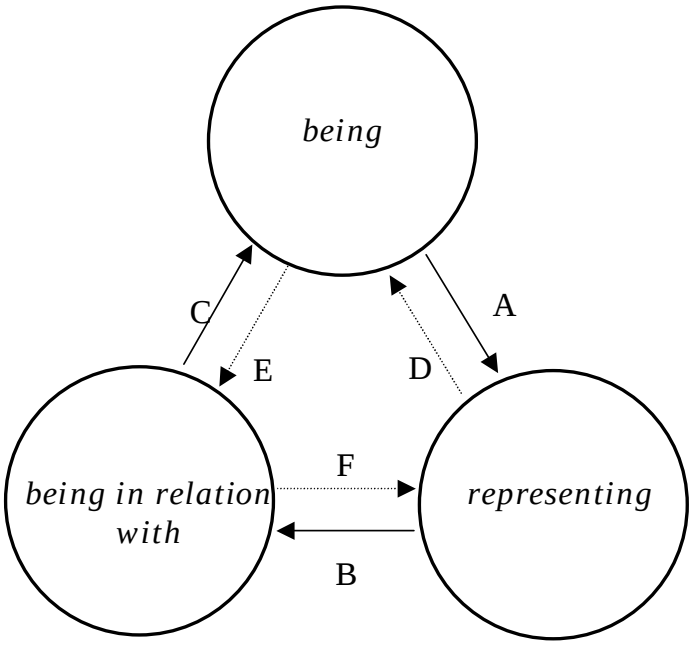


Figure 10-1 The fundamental relations between the three roles of the onphene. In the diagram you can see the relations among the 6 enunciated theses: Nothing exists without representing (A); Nothing represents without being in relation-with (B); Nothing is in relation-with without existing (C); Nothing represents without existing (D); Nothing exists without being in

[4] The claim of the Copenhagen School is open to epistemic interpretation or to a declaration of anthological indetermination of measured properties. In the first case the statement of the Copenhagen School would correspond to (A) + (E) excluding any possibility of saying anything about (C) and (D). In the second case it would be: (A)+(B)+(C)+(D).

> relation-with (E); Nothing is in relation-with without representing (F). The arrows indicate the ensuing relations among the three concepts. The fundamental claim of TEM is that such concepts cannot be separated and that, for this very reason, the structure described aove must be accepted as a whole.

Berkeley's idealism. The Irish Bishop Berkeley thought that reality was made up of ideas and as such was perceived. Berkeley had two strong pillars supporting his idea: God and the thinking subject. The first is a universal support for those who accept a metaphysical architecture. The second is the support provided by a disembodied subject who is capable of creating its own objects. Onphenes are different from idealistic ideas because: i) they don't need a subject; ii) they don't have to be perceived; iii) they don't exist as a subject's modification; iv) they don't radically oppose other entities with different properties (physical entities). Intentional relations make up both physical events and mental events.

Monism (Bertrand Russell's kind). Bertrand Russell embraced the idea of neutral monism that contained the aspects of the subject and the object. His ontology, which is based on events, is similar to ours. For Also Russell's events are the foundation on which the world is built up. He also thought that events can be considered from a subjective and from an objective point of view.

> An event is something occupying a small finite amount of space-time [...] an explosion, a flash of lightning, the starting of a light wave from an atom […] seeing a flash of lightning … hearing a tire burst or smelling a rotten egg, or feeling the coldness of a frog[5].

Russell claimed that both objective reality and subjective reality are made up of *percepts* and that these percepts are, sometimes, known in a direct way, like phenomenal contents, and, sometimes, they were acquired thanks to the strumental relations of our experiences. Basing his idea on percepts Russell also proposed a theory of the subject that, albeit from a different point of view, is similar to ours. This author thinks that the subject derives from aggregates of percepts, and that experience is a percept belonging to a particular aggregate of percepts rather than specific act done by the subject. Nevertheless Russell's idea differed from ours as follows: i) his percepts do not have a particular aggregating rule that allows to decide when a percept is part of a subject, ii) his percepts are entities that move in space-time (that pre-exists them) and must

[5] Quoted in (Stubenberg 1998), p. 302 from (Russell 1979 (1927)).

exist within the same dimension of atomic particles, iii) his percepts are atomically divided, and self-sufficient. In consequence of the first point the difficulty to define the characteristics of a determinated subject are increased. The second point obliges the percepts to be extensional entities, though evolving in time, they are extensional entities incapable of representing anything. The last point, which depends on the previous one, highlights the brain's incapability to experience extensionally any meaning different from itself. According to Russell, must literally be found in the brain; everyone's experience is limited to his/her own brain. Russell claims that each subject can be directly conscious only of the qualia by which he/she is constituted, qualia that can exist only in the place where the subject physically is: that is inside his brain. On the contrary, we define the problem of representation without incarcerating the subject in his/her own brain. Every onphene has, as content, its own crucial event. Every onphene represents only itself, but its own being is determined by the onphenes that were its essential causes.

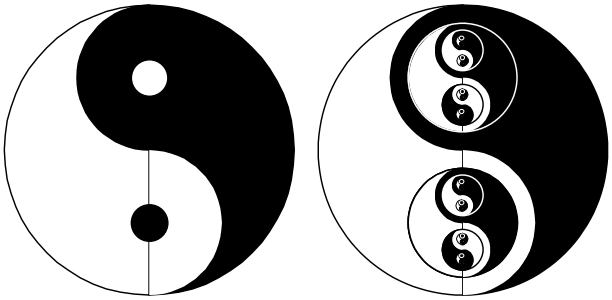


Figure 10-2 Curiously enough this famous symbol is well suited to onphenes. It is the famous Tao's circle (on the left). The two principles twirling together might correspond to representation and being, subjectivity and objectivity; the picture as a whole, the two principles becoming a new unity, provides the relational nature. Of course, this is only a graphical metaphor. Yet the metaphor can be slightly improved (on the right) and conjures up echoes of Hegel.

Functionalism and TEM. Both theories attempt to find a solution to the problem of what the mind is. They both use an elementary aspect of reality (relation of cause, function or dispositional) and try to explain a subject starting from the subject itself. The main difference is that TEM does not attribute the arguments of materialism to itself and because of this, it is not reduced to

exclusively objective entities, as reductionism would do. Other differences are summarized in the following.

- Regards functionalism cannot define what constitutes a functional element autonomously from conscious subjects while according to TEM a intentional unit is a static structure that allows the occurrence of event related counterfactually.

- Functionalism's multiple possibility of realization cannot explain the origin and the presence of the subjective quality of experience while according to TEM the subjective quality of experience corresponds to onphenes of first order.

- Functionalism does not give any explicit physical constraints about the physical implementation of a cognitive system. On the contrary, TEM suggests the physical constraints for a system that can develop a subjective experience of the world.

- According to functionalism, the paradox of the inverted spectrum is a logical possibility and Mary does not learn anything. In TEM, the inversion of the spectrum is impossible and Mary learns something new.

Functionalism is generally referred to as a non reductionistic theory, owing to the fact that it is not, strictly speaking, a physicalistic theory. It admits the existence of functional facts and states that they are different from any materialistic support, albeit ambiguously without an explicit ontology. Some functionalists like Dennett have developed theories so extreme as to be hardly distinguishable from reductionistic materialists' that, for this very reason, are often thought of as reductionist functionalists. However from the TEM point of view, Functionalism is a reductionistic theory because it does not have a metaphysics that unifies intentional states in anything more than collections of entities without real unity.

For functionalists the subject is nothing more than an aggregate of functional states. In this sense, functionalism is a reductionistic theory to all effects, because it lacks an ontology that can unify reality into entities of a superior grade. The need to go beyond metaphysics, as a theory of being that lacks an empirical ground, is a common issue both to analytical and continental philosophers. Historically, the correlation to objectivity caused the limitation to the objective dimension of experience. TEM extends the acceptable domain of experience. It allows maintaining the established boundaries of objective

empirical knowledge suggested by science, while it extends ontology to those areas traditionally and necessarily precluded to objective knowledge (subjective experience, quality, values, representations, intentionality). There is no *a priori* reason why a theory should not also consider subjective empirical facts (phenomenal objects) together with empirical objective facts.

If, by metaphysics, we mean every attempt to extend knowledge beyond the dimension of objective facts (according to the usual Galilean definition of physics), then TEM is a metaphysical theory because it establishes the proprieties of being so that empirical knowledge can and must have a place. Nevertheless, it shows a fundamental difference if compared to classical metaphysics. TEM is a constitutive theory of reality and is based on empirical verifiable facts (subjectively and objectively); it is not an arbitrary discourse about being, but an ontologically economic establishment of the empirical base. It tries to bridge the gap between empirical objective facts and empirical subjective facts. TEM is a theory that closes the gap between idealistic visions of reality and materialistic atomistic objective visions.

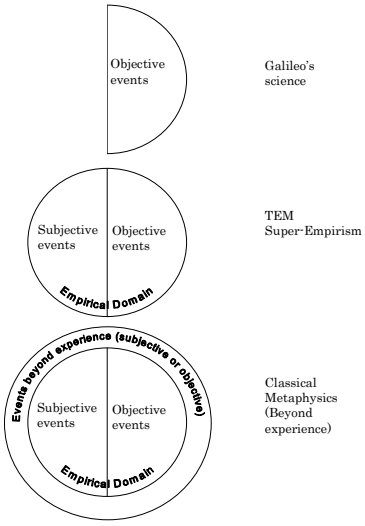


Figure 10-3 Galileo's science admitted only objective facts (on the top). Metaphysics, in the negative sense, tries to set principles and explanations that go beyond the empirical domain (on the bottom). TEM asserts that there is a neglected empirical domain, the domain of subjectivity, which has been hitherto concealed by objectivistic dogma.

10.1 Descartes', Leibniz's, and Whitehead's programming style

There is absolutely no possibility of describing what occurs between two consecutive observations.

Werner Heisenberg[6]

In the old days of programming, there was a rigid distinction between data and procedures (or programs)[7]. Data was memory doing nothing, while procedures were a way of coding actions that had to be performed on data. Obviously, procedures too had to be stored and, in this respect, were a kind of data. This computing model was very similar to the mechanicistic vision of a Cartesian world composed of matter (data) upon which forces (procedures) could act performing actions. Data had no power in itself. It was only capable of existing and 'being there'. It was the same role assigned by Descartes to substances: i.e. 'being there'.

Later, at the end of the '70, a new computing model started to become a commercial success: the union of data and procedures (eventually termed *methods*) into a unified entity (usually called *objects*). Objects were complex objects defined by their properties and by the actions that they were able to perform on themselves and on other objects. The important difference from the previous paradigm is that a distinction between static data[8] and dynamic actions is no longer made. Besides the practical advantages that such a model provides (which have been abundantly stressed elsewhere), we want to highlight the conceptual step it has been taken. Anything can be viewed as the modification of a property (a method can be seen as a property of a special kind) of an object. Descartes has gone out of the window and the door has been spewed for Leibniz. For Leibniz believed that the existence of each entity is equal to the sum of its properties and these properties can be seen as the powers that drive and control the becoming of the entity itself. There are no more static

[6] (Heisenberg 1958).

[7] This is not true for the very first days of computer science. Before the victory of von Neumann's model of computing there were several other attempts. In most of them, as well as in most of the early contemporary machines, there was no clear distinction between data and procedures. Nevertheless such machines were rapidly abandoned in favor of von Neumann's machines.

[8] 'Static data' is a term that belongs to computer science jargon. In this case, it is used with a street value. It has nothing to do with static and dynamic usage of data.

sub-stances that are passively acted upon by programs or by procedures. Each entity, each object, is bringing its own powers (methods) with itself.

An example is:

```
-      Function Power(b as integer, e as integer)
-            Dim r as integer
-      r=1;
-      For i=1 to e
-                  r=r*b
-            Next i
-            Return r
-      End sub
-      Dim base as Integer
-      Dim exponent as Integer
-      Dim result
-      Let base=5
-      Let exponent=2
-      result=Power(base, exponent)
```

As it is possible to see from the example, there is a clear distinction between the data and the procedures. Such a distinction will become clearer by a comparison with the same trivial piece of code translated into the new paradigm.

```
-      Class integer
-      Begin
-            Property value
-            Method power
-            Method assign
-      End
-      Dim object as integer
-      Dim other_object as integer
-      object.assign(5)
-      object.power(2)
-      other_object.assign(object.power(3))
```

The new paradigm has a series of advantages in terms of robustness and clarity. Unfortunately this Leibniz-like model of programming is still an abstract structure. It is still closed to external reality. It behaves like a closed monad. There is no way of translating its abstract representations into anything real without recurring to conscious observers.

Both programming paradigms have only a weak semantic relation to the meaning of their procedures: the semantic relation must be provided by human programmers or by users. In fact the paradigms correspond to syntactical languages: they present no intrinsic connection between their programs and the structures that are embedded in them[9].

With the advent of neural networks, a new paradigm began. Its nature is completely different from the two previous ones, although the first two can be used to implement this one. In neural networks there is no implicit definition of the meaning of symbols. They can be seen as mere tools to control future interactions among events. Of course, this would be true if we looked at a neural network as part of a real physical system and only if we used them as a convenient means to trace the route for future occurrence of events. In this sense, we can look at a neural network as a tool for describing and controlling the interactions of events: a way of proceeding not too dissimilar from Alfred N. Whitehead's standpoint and, of course, of TEM. In neural networks there is no *a priori* distinction between data and procedure, and also there is no distinction between properties and what they represent. They can be seen as a formalized flow of data waiting to let the appropriate events occur. The real network, of course, is not the program or the hardware that statically constitutes the network but the chain of events that takes place every time the network is inserted into a real agent in a real environment.

Unfortunately the main focus of neural networks has been directed not on their structure and their capability of controlling the flux of events in a straightforward way, but on their practical potentiality: approximation of functions, unsupervised learning, graceful degradation, and generalization. These are all fine features but do not correspond to the crucial aspect of neural networks: using them, it is possible to determine what the flow of events will be in the future. According to TEM, there is no sharp distinction between internal and external events. These boundaries are continuously and dynamically

[9] In the past, the dogmatical acceptance that semantics could not be part of the natural world; attempts were made to produce semantics from syntax alone. The idea was that by reproducing all the syntactical relations of external objects, systems would own their semantics too (Casalegno 1997).

changing and in hand with the related mind. What there is outside and what there is inside a given system does not depend on any physical constraint but on the way events mix together. Similarly there cannot be any hidden software inside our mind. We are not a «giant software virus that parasitized our brain»[10]. Similar conceptions are a result of the two previous software paradigms.

Neural networks, seen not as a static connections structure but as the dynamic occurrence of events is provoked by the corresponding static structures, can thus be termed as Whitehead's programming style.

10.2 Process and reality

> *[scientific materialism], with its abstractions, has become too narrow for science itself, too narrow for the practical facts which are before it for analysis. This is true even in physics, and is more especially urgent in the biological sciences.*
>
> Alfred North Whitehead[11]

No philosopher got so close to defining the TEM position as Alfred N. Whitehead in his philosophy of organism. Since his works are not very famous and due to the numerous affinities with TEM, a paragraph has been devoted to outline the philosophy of organism and its differences with TEM. According to Whitehead, philosophy aims mainly at defining the most abstract concepts: the fundamental parts by which every subsequent discipline tries to define reality. Western Thought has made itself incapable of proceeding towards a comprehension of reality, because of a basic mistake in the definition of elementary entities. The idea that reality is constituted by particles, isolated from each other, prisoners of an existence connected to the instantaneous dimension of time and space, compels physicists and philosophers to build a world starting from components that have no real grounding in our empirical experience. This is what Whitehead calls the mistake of misplaced practicalness

[10] «[consciousness] might not be good for anything – except replicating. It might be a software virus, which readily parasitizes human brains without actually giving the human beings whose brains it infests any advantage over the competition», (Dennett 1991), p. 221. A surprisingly similar idea can be found (even if from different perspectives) in (Blackmore 1998), (Dawkins 1990), (Jaynes 1976).

[11] (Whitehead 1925), p. 66.

that derives, among other things, from an uncritical acceptance of the classical metaphysical categories of substance and quality. According to him,

> the fallacy of misplaced practicalness, resulting in the idea of instantaneous matter with simple location, has been the occasion of great confusion in philosophy[12]

and

> the paradox [of mind] only arises because we have mistaken our abstraction for practical realities.[13]

In order to obviate such a fundamental mistake, Whitehead proposes metaphysics based on different principles and on a few simple elementary entities. Whitehead's fundamental starting point is what he terms the ontological principle, i.e. the principle according to which «no actual entity, no reason»[14] or, alternatively, according to which «to search for a reason is to search for one or more actual entities»[15]. With *actual entity*, Whitehead means a unity of existence. Using the same notation of § 1, the ontological principle is a generalization of point (D) or rather of the Cartesian *cogito*. Besides, it is very similar to the principle of the conservation of meaning and experience (§ 5.1), were it not for the fact this concept is referable also to the content of a phenomenal qualitative kind, while the ontological principle is traditionally used with reference to thought. Whitehead meant something similar to the principle of the conservation of representation because, in other points of his work, he frequently noticed that propositions and judgments express those contents whose subjective forms are those of the judgments. Whitehead derives this principle from Descartes and Locke.

Reality cannot be separated from experience through which we know it. For this reason this author's philosophy is often seen as a kind of pan-existentialism. Consequently the ontological problem becomes inextricably united to the epistemic problem (question). The ontological principle is a statement that the two problems are substantially identical. As does TEM, Whitehead believes that the classical concept of event[16] ought to be abandoned

[12] (Whitehead 1925), p. 51.

[13] (Whitehead 1925), p. 55.

[14] (Whitehead 1927), p.23.

[15] (Whitehead 1927), p.24.

[16] «An event is some particular special point at some particular moment. Events, therefore, have zero temporal duration as well as having zero special extension» (Penrose 1994), p. 219.

in favour of a new entity endowed with characteristic proprieties from our everyday experience. This entity is called an *actual entity*. The world is composed of these actual entities incessantly in relation the ones with the others. The aggregating process of these actual entities into other actual entities is termed *prehension*. «Actual entities involve each other by reason of their prehension of each other»[17]. In this way prehension becomes the heart of Whitehead's philosophy of organism, as well as the unifying moment of the actual entities. A prehension is the elementary process that makes the becoming of reality possible. The author defines it in this way:

> A prehension is only a subordinate element in an actual entity. Every prehension consists of three factors: (a) the subject which is prehending, namely, the actual entity in which that prehension is a practical element; (b) the datum which is prehended; (c) the subjective form which is how that subject prehends that datum. Prehensions of actual entities – i.e., prehensions whose data involve actual entities – are termed physical prehensions; and prehensions of eternal objects are termed conceptual prehensions.[18]

and also

> I have adopted the term 'prehension' to express the activity whereby an actual entity effects its own concretion of other things.[19]

According to Whitehead, the world is constituted by a never-ending flow of prehensions that would transform actual entities into other actual entities. The way in which the author defines the single moment of prehension allows him to spread unities of experience all over reality.

Incidentally, he takes back both the subjective moment and the objective datum to the elementary level of reality. Between them, there is a *subjective form* that identifies itself with the modality through which the prehension transforms the subject's becoming into an objective datum. Up to this point there are several similarities between the philosophy of organism and TEM. TEM also conceives the world as a flow of onphenes continuously transforming themselves the ones into the others, while exchanging state of contents and of occurring events. But it is better to point out a difference with Whitehead's panexperientialism. The prehension is defined as the necessary meeting point of actual entities, while the onphene represents the last horizon intuition can reach in defining the elementary constituents of reality. The onphene is

[17] (Whitehead 1925), p. 23.

[18] (Whitehead 1927), p. 23

[19] (Whitehead 1927), p. 52

founded on the necessity of the TEM three most important theses (see the previous paragraph). Besides in the case of the prehension, the role of process is highlighted, while in the case of the onphene there is a perfect balance in the three roles an onphene can assume. What is more, the relation between the prehension and its three moments is completely different from the relation between the onphene and the three roles it can assume. While the prehension *contains* the three aspects characterizing it, the onphene determines the three roles that constitute its first manifestations (existing, representing, being in relation-with). The onphene does not contain a datum, but it can be contained. The onphene does not have a subjective moment but it can constitute a subjective content. The onphene does not have a subjective form, but some onphenes can be seen as onphenes of a superior grade. At this point, we can introduce another important difference. Regarding prehension subjectivity and objectivity, which are seen like internal moments. On the contrary, regarding onphenes, we cannot speak of subjectivity and objectivity, given one onphene, they are concepts determined successively. As we have seen in § 6.3, the proposed definition of objectivity is being an onphene of a superior order to the first one, that corresponds to a relation of onphenes. On the contrary, subjectivity would correspond to onphenes of a simpler order.

Inside the onphene we can speak neither of a subjective nor of an objective moment. The onphene can be considered whether as a subject or as a content or as a relation, but only globally, that is *from outside* according to the relation with it. The onphene cannot be divided into parts or into moments. When speaking of prehension we can say that «each actuality is essentially bipolar, physical and mental, and the physical inheritance is essentially accompanied by a conceptual reaction»[20], while, in the case of an onphene, we have never spoken about bipolarity. Even if, for Whitehead, bipolarity does not imply a loss of unity, he introduces (when speaking about prehension) those characteristics that will be explained later (subjectivity, liberty, emotion, creation). As far as the third characterization of prehension is concerned, other differences must be pointed out: the subjective form within which the prehension develops its inner process. According to Whitehead, every prehension is accompanied by an emotional tone (affective tone)[21]. Though he points out that such an emotional moment should not be intended as necessarily identical to human beings' emotions, the author seems to have embodied some characteristics of conscience at an atomic level. On the contrary, TEM does not give an onphene any emotional aspect but explains the emotions as the complex contents that are

[20] (Whitehead 1927), p. 108.

[21] (Whitehead 1933), p. 233.

determined through the complex relations between body and mind. In the case of the onphene, emotion, or any other similar thing, is near the top of the ontogenetic scale and it is not part of the elementary structure of reality. The idea of attributing elementary emotions to elementary unities of reality induces Whitehead to construct his entire theory of perception on the presence of emotional elements in whatever perceptive content.

> In their most primitive form of functioning, a sensum is felt physically with emotionally enjoyment of its sheer individual essence. For example, red is felt with emotional enjoyment of its sheer redness. In its primitive prehension we have aboriginal physical feeling in which the subject feels itself as enjoying redness[22].

On the contrary, with TEM there is no need to reproduce all the attributes of the subject (like emotions) in every event, in every experience. An onphene can easily represent red without such a content assuming any emotional colour. In practice, it is also possible, for some colours, to be perceived with an associated emotional meaning. However, this fact is at all not necessary. The empiric proof is that there are a few people whose amygdale is not working because of pathological or traumatic causes: yet, though such people are unable to feel emotions, there is no reason for believing that they can't perceive colours consciously. Or even for believing that they are not conscious. Another great difference is that the prehension is bound to the flowing of time. In other words, the prehension evolves *within time* – it is contained *within* time. On the contrary, the onphene determines time and it is not necessarily within time.

Space and time are onphenes that determine certain peculiar relations among other onphenes. Space and time are the content of onphenes of a superior grade to the first one. The prehension originates from the concept of event and, therefore, it is subject to the occurrence of such events, whereas an onphene precedes (logically and metaphysically) both the concept of time and the concept of space[23]. After pointing out these differences, common details must be taken in consideration. According to Whitehead, intentionality should not be seen as an emergent propriety of macroscopic systems (static or dynamic) but as an elementary propriety of reality. An onphene is the bearer of intentionality as much as prehension. In David Griffin's words:

---

[22] (Whitehead 1927), p. 314.

[23] In reality Whitehead's position about time is slightly more complex. According to him: «every actual entity is in time so far as its physical pole is concerned, and is out of time so far as its mental pole is concerned» (Whitehead 1927) p. 248. Notwithstanding this affirmation the difference between onphenes and prehensions is clear.

> Whitehead is […] pointing to the most basic form of the operation that lies behind what philosophers, following Franz Brentano, have called "intentionality" meaning "aboutness". By using the term "prehension", however, Whitehead means no merely external reference but the way an experience can include, as part of its own essence, any other entity.[24]

TEM fully agrees on this point. The subject is a radical unity and such a unity cannot be explained starting from a reductionistic ontology (as outlined in the first chapter). The philosophy of organism is as essentially against reductionism as TEM. According to Whitehead, the subject determines itself starting from an aggregate of occasions of experience. Since «an occasion of experience consists entirely of prehension»[25]. Whitehead's subject emerges from the flow of prehensions which reality is composed of. In the same way the TEM subject takes form starting from an aggregate of occurring onphenes. There are various similarities but also many differences in the two theories about the subject. The main difference, which enlarges as the two theories progress, comes from the subtle difference with which prehensions and onphenes unify reality. The prehension is an atomic process, conveying within itself the actual entities transformed into objective content. On the contrary, the onphene expands as far as the onphenes it represents, through the selection of a critical event. The most obvious consequence is that, to the philosophy of organism, the objective datum is always inside a given prehension, which changes it into an actual entity. On the contrary, in the case of TEM, it is the subject that expands itself; it is the mind that enlarges until it includes the appropriate events.

According to Whitehead, an adult's brain can get the corresponding subject to have experience of all the things that have previously been object of its experience. This happens because the events, contained inside that subject's brain have, in their physical pole, as an objective datum, the meaning derived from the events, with which they have been in contact. Given a chain of prehensions linked together, each of them adds a new meaning corresponding to its emotional tone, to the original meaning of the first actual entity. Consequently the meaning perceived in the end by the subject corresponds to the set of meanings that the original event has acquired along the chain of prehensions – something like the growing of a shell. Every passage conceals the previous one, until it is completely inaccessible. In this way, the quality added by the last prehensions involved in the chain becomes the content of the perception. The subdivision between primary and secondary proprieties is

[24] (Griffin 1998), p. 126.

[25] (Whitehead 1927)

clearly accepted by Whitehead, because these last ones are the result of the last prehensions along the perceptive chain. «Whitehead's view is that secondary qualities are produced by the mind out of values, or emotions»[26]. Obviously, these things being stated, the content of a perception is not the external object in itself (or alternatively the original event) but the quality added by the prehensions that transmitted the perception. According to Whitehead, «the central lesson of physiology itself is that sense perception is not direct observation of its objects. The physiologist looking at my brain is not directly observing my brain cells»[27]. Therefore the chain of prehensions mediates perception. Through the last passages the organs of sense add particular secondary qualities to their objects.

> The transition from without to within the body marks the passage from lower to higher grades of actual occasions. The higher is the grade, the more vigorous and the more original is the enhancement from the supplementary phase. […] Thus the transmitted datum acquires sensa enhanced in relevance or even changed in character by the passage from the low-grade external world into the intimacy of the human body.[28]

The perceptive process modifies the original content by the introduction of sense data or, in Whitehead's words, of external objects. In the case of TEM we don't resort to these intermediaries of the perceptive process, because, owing to critical events, the perception can *go back* as far as the external event that must be represented. According to TEM, perception is a completely transparent process that does not add anything to the perceived event and that permits the foundation of the appropriate causal chain between external and internal events. An obvious consequence of the difference in the interpretation of the relation between the subject and the external world is given by the mental experiment of the brain in a jar. According to the philosophy of organism, the perceptive states of a brain, conveniently stimulated in an artificial brain, would be similar to those of a brain inside a normal body. According to TEM, independently from the last events, along the perceptive chain, the complete identity of causal chains is necessary. Besides, because of the kind of perception proposed by Whitehead, we have to introduce a category of *eternal objects* that constitute the content added by senses to the events deriving from the external world. Since «the direct perception (…) can thus be conceived as the transference of throbs of emotional energy, clothed in the specific forms

[26] (Griffin 1998), p. 141.
[27] (Griffin 1998), p. 142.
[28] (Whitehead 1927), p. 120

provided by sensa»[29] it is not clear what «the specific forms provided by sensa»[30] are and how our senses can reach them. It is a problem common to all the perception theories that consider the domain of phenomenal objects separated from the domain of physical events. In the following passage, it is clear that senses determine an obliged passage in which events are clothed with the appropriate sensorial qualia.

> It receives an exemplification in the character of our perception of the world of contemporaries actual entities. That contemporary world is objectified for us as '*realitas objectiva*' illustrating bare extension with its various parts discriminated by differences of sense-data. These qualities, such as colours, sounds, bodily feelings, tastes, smells, together with the perspectives introduced by extensive relationships, are the relational eternal objects whereby the contemporaries actual entities are elements in our constitution. […] The bare mathematical potentialities of the extensive continuum require an additional content in order to assume the role of real objects for the subject. This content is supplied by the eternal objects termed sense-data. These objects are 'given' for the experience of the subject.[31]

TEM evades the problem of a separated domain of phenomenal entities or qualia (besides the problem of their relation with events) by proposing a perception theory that presupposes a perfect realism and coincidence between the perceived event and the occurred event. Speaking about perception, we have to point out the lack of the concept of critical event. Given a chain of prehensions it's impossible to understand when the meaning of the original actual entity is to be substituted by the meaning of the prehensions that follow one another. Moreover, it is not clear when to speak of original actual entities when, from a logical point of view, there is the possibility of an endless regression. This cannot happen with TEM because of the critical event that, not only guarantees the transparency of perception, but also allows the causal regression to be arrested. Whitehead, aware of this problem, confines himself to mention that:

> [given a causal chain] Some of it may stand out with distinctiveness by reason of some peculiar feat of original supplementation, which retain its undimmed

[29] (Whitehead 1927), p. 116.
[30] (Whitehead 1927), p. 117.
[31] (Whitehead 1927), p. 62

> importance in subsequent transmission. Other members of the chain may sink into oblivion.[32]

Without the critical event, which represents a barrier of intentional opacity, which is insurmountable by perception, it's impossible to define where and why the causal chain that transmits or modifies the meaning should stop. Other differences are of a more technical nature and ought to be dealt with in depth. Nevertheless, we think we have provided enough exemplification, albeit inevitably concise, of similarities and of the differences between the philosophy of organism and the TEM. In conclusion we believe that TEM and the philosophy of organism have a fundamental point in common: they agree that the classical metaphysical categories, substance and quality, have caused numerous misunderstandings, both in science and in philosophy, and that a new definition of the elementary elements of reality can break new ground in the understanding of the mind body relation.

10.3 Science is a card game

Philosophy is the scientific construction of a world-view.
Martin Heidegger[33]

In some Italian villages people play a card game termed *Machiavelli*. Bridge cards are needed: two decks of cards, each of 52 cards, without the jokers. Two or more people can play it. One player could also play it, but the game becomes a solitaire that, in the end, will always be successful. When a player makes a combination (for example a straight, a three of a kind, a straight flush), he/she puts it on the table. At every hand or the player draws a card from the deck. The aim is to get rid of all the cards one holds before the other players do. All the combinations that have been completed remain on the table. And what characterizes *Machiavelli* best is that it is possible to free up a card, using those already on the table, breaking the proposed combinations in order to make new ones. At every hand the players draw a card from the deck if they cannot put more cards on the table.

As the game proceeds the players draw fewer and fewer cards. Since there are an ever increasing number of combinations, it is better to concentrate on the cards on the table rather than drawing new cards from the deck. At last

[32] (Whitehead 1927), p. 120
[33] (Heidegger 1988), p. 7.

someone manages to find a combination that allows him/her to put down all his/her cards on the table. He/she is the winner. The game is a metaphor of the empiric speculative method, used to expand the limits of knowledge. The deck of cards is the world – that is reality. The world is made up of a certain amount of events, of facts. Whenever a player draws a card, he gets experience of the world. He doesn't know, previously, what is going to come out of the deck. The pack of cards (the world) is full of unknown facts and drawing cards out, is the only way of knowing it. But making experiences in this way is not sufficient.

We must give a meaning to these experiences and this meaning can only be given by concepts. Every time a player makes a combination of cards he has created a concept.

Unifying simple facts into a general fact is like combining isolated experiences into a theory. The set of combinations on the table represents the best generally accepted theory at our disposal about the known facts of the world. Like every theory, it is not complete because there are still some facts (in the deck or in the players' hands) that nobody has been able to include. Every combination of Machiavelli is a concept based on experience; the drawn cards are the hypotheses already experienced by science. In this game – it is known from the beginning – that cards belong to an orderly set made up of four suits ordered in four straights of thirteen cards each or thirteen straight flushes, while in real life we have no assurance that the facts of experience will find, in the end, a global settlement. Every philosopher's and every scientist's hope is that an order waiting to be uncovered really exist, beyond the multiplicity of attemptable things. Since the deck of cards represents the whole of reality, consciousness (i.e. the conscious mind) constitutes itself from the world. What is more, every individual conscience corresponds to the cards drawn by a player, which corresponds to the experiences he has had. Similarities are numerous. When a player adds a card on the table he is (he behaves) like a researcher relating data of an experiment with a result that confirms the existing theories. The player mixing all the combinations again is like somebody making a radical change in the assessment of reality. The cards in the player's hands correspond to those facts that neither scientists nor philosophers have ever been able to explain, but have reduced them to categories or concepts consistent with the theories they currently have. Every player has the right to rearrange all the cards, as he/she likes. If his/her attempt is going to destroy the strategies of the others and his/her proposal is successful, it is accepted by everybody in the end. On the table nobody has the right to interfere with any combination or to prevent the reliability of a concept to be tested. On the contrary, we can't help noticing that, often, in the history of thought, certain ideas have frozen up and, because of that, have ceased to be objects of research.

Inevitably this fact, here slowed down or stopped the development of research in these areas. As an example, the idea that the Earth was in the centre of the universe forced Ptolemaic scientists to invent complicated theories about epicycles in order to defend the hypotheses of geocentrism. There are two possible moves: either trying to put new facts together drawing other cards in the hope that they may amalgamate with those facts that are still inexplicable and that they will allow something new to be discovered; or trying to understand whether the facts, now at our disposal, can make them recombine again (join together again). If the game is played with two decks of cards, in one deck the cards have red backs in the other they have blue. So the cards are double and identical at the same time. Every value (number or figure) can have a red or blue back. Always by analogy, we could consider the cards with a red back as the subjective facts and the ones with a blue back as the objective facts. Metaphysically speaking, we can compare the ontology of things to the deck of cards, and the domain of experience and consequently of knowledge to the drawn cards. The red cards, i.e. subjective experiences would correspond to phenomenology; the blue cards, i.e. objective experiences[34] would correspond to epistemology. The kind of experience of the world is associated to the deck of the cards.

By means of the above metaphor, the main currents of thought as far as the nature of conscience is concerned can be analysed. Eliminativists are like those players that refuse to see and to utilize all the cards with a red back in their own combinations. It is clear that, after a while, even if they were very lucky and succeed in drawing many combinations of cards with a blue back (i.e. many theories regarding the objective facts), they would end by having many cards with a red back, which they would fing impossible to collocate. It would be no use to keep on denying their existence. The various currents of eliminativism are in this position: they have to deny the existence of facts acknowledged by everybody.

On the contrary, reductionists refuse to accept that the combinations are something real. According to them, only the whole deck is important and not the way the cards are combined. We could say that their position means either to refuse to play or, according to the various currents, will only accept certain combinations. For example, they might come to the conclusion that the game of poker should be seen like fragments of straights and not as a whole and that straights have to be seen like sets of fragments of the game itself. In this sense,

[34] Clearly the terms "objective experience" may look like an oxymoron because we usually speak of subjective experience or objective knowledge.

and here we have to condone the difficulties placed on by our metaphor's path, they seem to be willing to reduce the number of combinations on the table.

Constructivists, functionalists and idealists? Each of them tries, and all in different ways, to deny the existence of all, or of a part of the pack of cards. The cards do not exist, they declare, only our combinations do. Some philosophers think that the players create them by considering the idea of existence of the cards as a support to be eliminated as soon as they clearly appear to be a useless hypothesis to describe the development of the game. For others, the combinations exist, the players do not create them, but the cards exist only if they belong to a combination. They cannot understand what the meaning of a figure isolated from the deck of cards could be. A card has a meaning only if it is in the deck or in a combination, because in this case it is related with the others. So single cards don't exist. Other philosophers, from a more extreme point of view, uphold the idea that the deck of cards is only a rhetorical means to describe the behaviour, whether of the single player or of the group of players. In other words, only the players exist, the cards are an illusion. TEM tries to use all the cards, the ones with the red back and those with the blue back and to find a good synthesis of all facts. To succeed, it supposes that, notwithstanding the colour of the back, every number and every figure corresponds to a real fact, an event whose existence modifies the game.

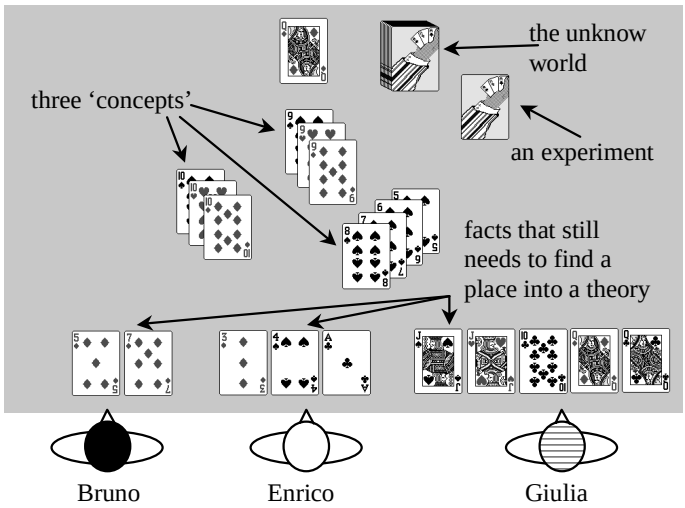


Figure 10-4 Bruno, Enrico and Giulia look at the same cards on the table: the public fact. Each has a personal access to the cards he/she holds in his hand. The deck is the unknown world. To draw a card is to make an experiment.

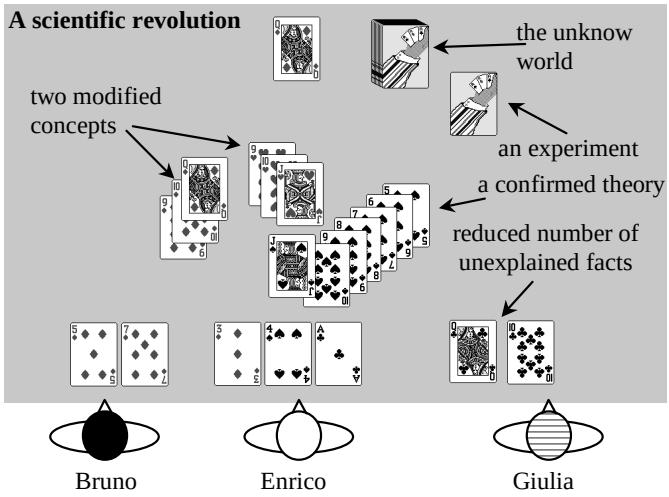


Figure 10-5 Giulia made a combination. She changed the theory of the world. Two fundamental concepts, the two combinations on the left were reformed to let her get rid of her cards. Besides, she was able to validate one of the already known theories by adding new facts to it. The other two players, Bruno and Enrico had to accept her move since it reduced the number of unexplained facts (the cards Giulia is still keeping in her hands).

TEM propounds a principle, on the basis of which, certain aggregates of cards with red backs can be seen as natural combinations, and because of this, unified as subjects. Other combinations, composed of cards with blue backs, correspond to the objects and to the other entities studied by science and by logic. The limit between the two classes is not insurmountable and every subject, every conscious mind, modifies its limits continuously, enlarging to new events. According to TEM, every playing card is, at the same time, an event (that is a quantum of existence) and represents content (number or figure). It is also a relation-with: the value of every card is related to its position as regards to the others. Every card is an intentional relation (an onphene) and is considered subjective or objective according to the way it joins the unities constituted by the subjects. Let's make one last observation: how can something represent (refer to) something else? In this book we have presented a hypothesis that must answer this question. Representing is being and understanding the former requires understanding the latter. Representing is the inner nature of conscience, but such nature is incomprehensible if it is not founded on existence. So consciousness and reality can be seen as the two

objectifications of the two fundamental aspects of reality: representation and existence, unified by the relation.

The idea that the ontology of reality (intended in its more ample meaning) should be neither extensionalist nor dualist, nor idealist, but based on intentionality, has been strongly asserted. The bonds of our proposal are: i) the compatibility, inside the category of objectiveness, with scientific method; ii) the possibility of formulating hypotheses also including the phenomenal field of subjective and qualitative experience; iii) the capacity of reaching a synthesis, acceptable from Ockam's point of view of the present proliferation of concepts aiming at explaining the mental.

If we wanted to give a name to the position we have defined, we could term it intentionalism. It is the attempt to consider intentionality not as the emerging product of other proprieties of reality, but as the constitutive principle that precedes all its successive definitions: subject and object, being and *dasein* (being in the world) consciousness and reality.

Summary

We reached the end. We proposed a fundamental revolution in the ontological framework. Instead of the classic extensional ontology we propose that the fundamental level of reality is intentional in nature.

A final comparison with other points of view is made. It is shown that the proposed framework is a generalization of several other attempts that can, therefore, be explained as its partial accomplishment.

Our proposal should have an effect on the way we build robots and on the way they are programmed. It is noteworthy that in the evolution of programming there are three phases that resemble the way of thinking of three famous philosophers: Descartes, Leibneitz and Whitehead. The last is the philosopher whose ideas come closest to our proposal. A detailed analysis of the commonalities and the diversities between TEM and the philosophy of Organism is presented.

11 Appendices

11.1 The (non) existence of objects

This is the formalized and logical argument we can use in order to demonstrate that physical objects do not exist by themselves. Let's assume that the world is composed of small particles m_i where i is a progressive number identifying each particle. It is not important whether these particles are molecules, atoms, quarks or whatever. It is not even important if they are mass or energy packets. The only requirement is that they *exist* in some relevant sense of this abused verb, by themselves. That is that they exist without requiring the observation of a conscious subject. We assume also that each one of these particles has some physical property (mass, position, velocity, etc) and that these properties exist by themselves. What we have just depicted seems to be a reasonable description of ontological reductionism or atomism. Given this ontology, what is a composite object O then? The only meaningful description is the following:

$$O=\{m_i \textit{ such that } P(i)\} \textit{ where } P() \textit{ is whatever rules arbitrary chosen on the class of physical property } X_{ji}$$

This formula is interesting because it summarizes the problematic relation between extension and intension, between wholes and parts. The set of all m_i corresponds to the extensional world made of unrelated atomic units (each m_i is a elementary particle). X_{ji} is the set of all physical properties (primary properties) instantiated by particles. Therefore $\{m_i, X_{ji}\}$ represents the physical world as such. Every $P()$ identifies some object. And yet every $P()$ is not extensional in itself. From the point of view of physicalistic ontology the world is complete once $\{m_i, X_{ji}\}$ is fixed. There is no need to add anything more. But if $P()$ was to be added what would O be? O would be a whole, which clearly depends on more than the physical base alone. In fact, it depends on $\{m_i, X_{ji}\}$ as well as on $P()$. We could restate the previous as follows:

$$O=R(\{m_i, X_{ji}\},P)$$

where R represents the relations between intension and extension.

To have an intuitive idea of this notation we can imagine a world made up of a limited number of particles and a limited number of properties; i.e. 100 particles and 2 properties. Let's imagine also that each of the properties identifies the position of a particle along one axis. In such a way X_{1i} would correspond to the position of each particle along one axis and X_{2i} would correspond to the position along another axis. Let's imagine that all particles are uniformly distributed on a square of size 10 centred in the origin[1] (Figure 11-1a). Suitable P could be for example

$$P^1(i) = \begin{cases} \text{true } if \ X_{1i}^2 + X_{2i}^2 < 3^2 \\ \text{false } if \text{ otherways} \end{cases}$$

$$P^2(i) = \begin{cases} \text{true } if \ |X_{1i}| < 3 \vee |X_{2i}| < 3 \\ \text{false } if \text{ otherways} \end{cases}$$

P^1 and P^2 correspond respectively to Figure 11-1b and Figure 11-1c. The problem is that Ps do not belong to the extensional world as such; they must be added to it. Therefore if we, as conscious subjects, experience wholes, it follows that the extensional ontology must be false.

In the real world the list of properties can be more complex but it does not change the form of the rationale. For example, suppose that X is the length wave of reflected light (that is the colour of the particle). Then a suitable P would be: P(X)=true if and only if X is red (or is in a reasonable range). If you had a red shape on a wall you could select that shape and the belonging particles by using the $P_{red}()$ property. Yet $P_{red}()$ does not belong to the extensional world.

In fact, for every combination of particles, it is always possible to find a P() that is pointing at it. In other words, there is no combination of particles that could not have its particular intension. Given the fact that we can always suppose the existence of a P() so that (for example Figure 11-1d)

$$P(i) = \begin{cases} \text{true } \textit{if we want that particle } m_i \textit{ belongs to the object} \\ \text{false } if \text{ otherways} \end{cases}$$

[1] Clearly in such a limited universe the measure units are meaningless. We are using them for explanatory purpose.

The problem is that $P \notin \{m_i\}$ and $P \notin \{X_i\}$. In other words, P is neither a physical object nor a physical property. The problem is what stuff P is made of. Functionally it occupies the role of an intension but ontologically it is something completely external to physical reality, therefore its existence must be or something related with consciousness or not existent. If we want to follow the latter, we must conclude that the objects do not exist. Since this is an absurd conclusion, which follows from the application of our standard objectivistic categories, we must look for a revision of such categories.

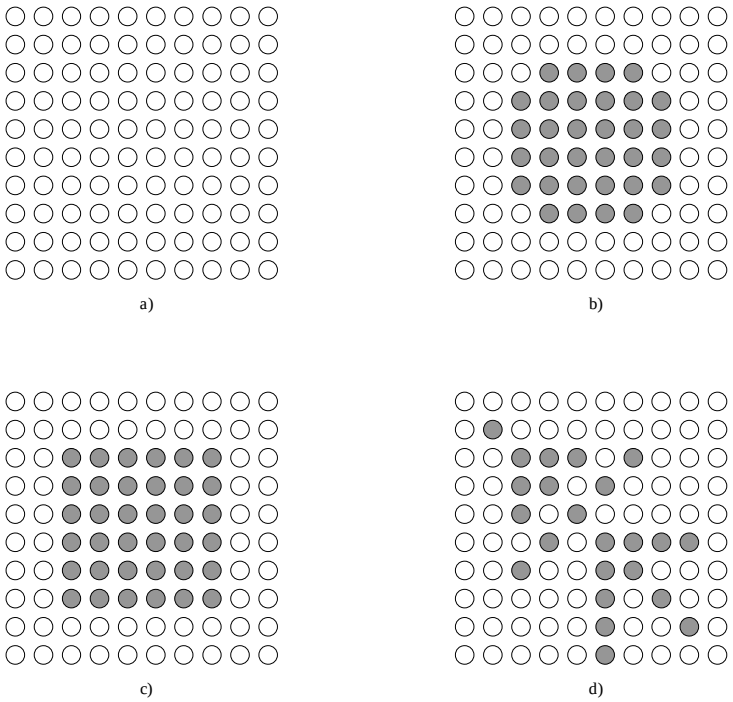


Figure 11-1 A limited universe made up of 100 particles with only their position as possible properties. Starting from the original extensional universe we can add new wholes. Are they part of the extensional ontology or do they require something more?

11.2 TEM in a nutshell

The Theory of Enlarged Mind (TEM) can be summarized as follows.

1. Reality is only one and the best knowledge it is possible to get of it is a theory capable of explaining all empirical facts by using as few explanatory principles as possible.
2. Empirical facts are both subjective and objective.
3. Any experiential fact is an empirical fact.
4. It is not possible to say anything about empirical facts that are not experiential facts; so, from our point of view, all empirical facts are experiential facts.
5. Every experiential fact exists, represents, and is in relation-with.
6. Nothing exists without representing.
7. Nothing represents without being in relation-with.
8. Nothing is in relation-with without existing.
9. Nothing represents without existing.
10. Nothing exists without being in relation-with.
11. Nothing is in relation-with without representing.
12. Representation, existence and being in relation-with are just three different roles played by the same entity that is called *onphene* (or, as a synonym, *intentional relation*).
13. Being an event is a role (not an entity or a substance). 'Event' denotes what is done by an entity, not what that entity is.
14. The role of an event is to provoke a difference in reality: being something that is having a content that is being in relation-with some aspect of reality.
15. An onphene is the simplest ontological candidate to identify the structure defined in the previous points.
16. Each onphene occupies the same role of the event; so it is a natural candidate to support events.
17. 'Every event is an onphene' is a contingent truth.
18. 'Every onphene is an event' is an *a priori* truth.
19. Content is what an onphene is, so an onphene (or intentional relation) is a possible content.
20. Every onphene must be in relation-with, so it is the content of another onphene. If this were not true, that other onphene would be not part of reality.

21. Since each onphene has a content, this content is either a simple content or another onphene.
22. An onphene, which as its content has another onphene, is termed a *second order* onphene.
23. An onphene, which as its content has a second order onphene, is termed a third order onphene.
24. In similar way higher order onphenes can be envisaged.
25. A first order onphene corresponds to a subjective event.
26. A second order onphene corresponds to an objective event, which is relational in nature.
27. A third order onphene (or further) corresponds to a logical proposition.
28. Subjective events (first order onphenes) constitute the domain of subjectivity. This domain is defined *before* the subject (idealistic principle).
29. Objective events (first order onphenes) constitute the domain of objectivity. This domain is defined *before* the object (materialistic principle).
30. Logical events (third order, or further, onphenes) constitute the *a priori* truth domain. This domain is defined *before* the belief in a transcendental dimension (third reign).
31. Subjective events represent simple events, objective and logical events represent onphenes as such.
32. The content of simple events corresponds to phenomenal objects (colours, tastes, pleasure, pain).
33. The content of objective events corresponds to empirical observation of relational nature of second order onphenes as such (bigger-than, darker-than, stronger-than).
34. The content of logical events corresponds to logical propositions that are relations among other relations (entails that, true, false, twice as).
35. Every phenomenal representation and every meaning entails a real event – something that exists (the *Principle of conservation of representation and meaning*).
36. Every onphene unifies that part of reality, which is its content (the *Principle of unification of reality*).
37. Every onphene has, as content, an event – or a group of events – that is defined as a *critical event*.
38. The *critical event* of an onphene is that event – or group of events – whose existence has been necessary and sufficient for that onphene.

39. The critical event is the content of its onphene.
40. The subject is a unified collection of representations.
41. A collection of representations is unified when the collection is the critical event of another representation (according to the Principle of unification of reality).
42. The subject is a collection of representations unified by their being the critical event of an onphene.
43. The onphene that unifies a subject is called *principle of the ego* or *ego*.
44. When the onphene *ego* is the content of another onphene, it is termed *self*.
45. The *ego* seen as an object is the *self*.
46. The self and the ego are two real unities, so they exist.
47. Consciousness is referred to as the fact of being a subject – which is a unified collection of onphenes.
48. The mind is always a conscious mind, which is a conscious subject.
49. *Self-consciousness* means that the subject has, among its contents, the self.
50. The mind is part of reality: having an experience entails enlarging the part of reality that constitutes ourselves as subjects.
51. Every content of an individual mind is part of that subject because a new onphene is unified in that subject collection of onphenes.
52. There is only one kind of act through which a mind gets its content: by enlarging the collection of onphenes that constitutes it; a subject that enlarges itself by including a new onphene.
53. Having new experiences entails a transformation of the subject.
54. All mental events are (directly or indirectly) conscious events: subjective experiences, dreams, objective observations, knowledge, beliefs, thoughts, goals, motivations, feelings).
55. The phenomenal subjective experiences correspond to first order onphenes.
56. The empirical objective knowledge (observations) corresponds to second order onphenes.
57. The *a priori* knowledge corresponds to onphenes of higher orders.
58. Perception is representation: there is no perception without content.
59. Traditionally perception is referred to as first and second order onphenes, while sensation is usually confined to first order onphenes.
60. Having a perception means that the subject enlarges itself to include a new onphene.
61. Perceptual content is the critical event of the included onphene.

62. Sensation is perception.
63. *Memory* is perception whose critical event occurred some time before. Memory is usually produced voluntary.
64. Hallucinations and fosphenes are perceptions whose critical events occurred some time before. They are usually involuntary.
65. Non veridical perception is a perception whose first order contents are associated with second order contents that are unusual.
66. Every conscious event is something; therefore it must correspond to an onphene and it must have content.
67. Every perception has content.
68. There is only one kind of mental act: the belonging of an onphene to a subject or, that is exactly the same, the enlarging of a subject to a new onphene.
69. All mental acts (experiencing, feeling, perceiving, believing, getting by intuition, understanding, knowing) correspond to the same event (onphenes becoming part of a subject) but can be differentiated on the basis of the kind of content (first order, second order, and so on).
70. A *thought* is an onphene with a content of second or further order.
71. *Thinking* means perceiving one's own thoughts.
72. *Grasping a thought* means to enlarge ourselves to a new onphene of second or further order.
73. *Language* is a collection of relations among contents.
74. A *concept* is a collection of relations that identifies content.
75. A concept either corresponds to an existing onphene or could be a network without a real content. In the former case it is a *real* concept while in the latter case, it is a *conventional* concept.

References

Abbott, B. (1997). "A note on the nature of water." *Mind* **106**(422 April): 311-319.

Adolphs, R., D. Tranel, et al. (1998). "The human amygdala in social judgement." *Nature* **393**: 470-474.

Agazzi, E. (1981). "Intentionality and Artificial Intelligence." *Epistemologia* **IV**: 195-228.

Aleksander, I. (1994). *Towards a Neural Model of Consciousness*. ICANN 94.

Aleksander, I. (1996). *Impossible Minds: My Neurons, My Consciousness*, Imperial College Press.

Allen, C. (1997). Animal cognition and animal minds. *Philosophy and the Sciences of the Mind: Pittsburgh-Konstanz Series in the Philosophy and History of Science vol. 4*. Pittsburgh University Press, P. Machamer & M. Carrier Pittsburgh. and Konstanz: 227-243.

Arbib, M. A., Ed. (1998). *The Handbook of Brain Theory and Neural Networks*. Cambridge (Mass.), The MIT Press.

Arditi, A., J. D. Holtzman, et al. (1988). "Mental imagery and sensory experience in congenital blindness." *Neuropsychologia* **26**(1): 1-12.

Armstrong, D. (1981). What is Consciousness. *The Nature of Mind*. Ithaca, Cornell University Press: 55-77.

Armstrong, D. M. (1968). *A Materialist Theory of Mind*. London, Routledge & Kegan Paul.

Armstrong, D. M. (1988). Can a Naturalist Believe in Universals? *Ulmann-Margalit*: 103-15.

Baars, B. (1988). *A Cognitive Theory of Consciousness*. Cambridge, Cambridge University Press.

Barbieri, M. (1988). "La creazione dell'informazione." *Epistemologia* **XI**: 283-292.

Basti, G. (1996). *Neural Images and Neural Coding: The Semantic Problem in Cognitive Neuroscience*. Downward Processes in the Perception Representation Mechanism, Ischia (NA) Italy, World Scientific.

Bateson, G. (1972). *Steps to an Ecology of Mind*. New York, Ballantyne.

Bateson, G. (1979). *Mind and Nature: A Necessary Unity*. London, Wildhood house.

Bechtel, W. (1988). *Philosophy of Mind*. Hillsdale (N.J.), Lawrence Erlbaum Associates Inc.

Becker-Colonna, A. L. (1966). *An introduction to the study of Egyptian hieroglyphs and symbols*. El Cerrito (Cal.), College Press Western Baptist Bible College.

Bermùdez, J. L. (1998). *The Paradox of Self-Consciousness*. Cambridge (Mass.), The MIT Press.

Binder, E. (1939). I, Robot. *Amazing Stories*. **13**.

Bizzi, E. (1979). "Strategy of Eye-Head Coordination." *Progress in Brain Research* **50**: 795-803.

Bizzi, E. (1981). Eye-Head Coordination. *American Physiology Society*, American Physiology Society**:** 1321-1336.

Blackmore, S. (1998). *The Meme Machine*. New York, Basic Books.

Block, N. (1980). "Are Absent Qualia Impossible?" *Philosophical Review* **89**(2): pp. 257-74.

Block, N. (1988). What Narrow Content is Not. *Meaning in Mind: Fodor and his Critics*. B. Loewer and G. Rey. Oxford, Blackwell.

Block, N. (1990). "Anti-Reductionism Strikes Back." *Philosophical Perspectives*.

Block, N. (1990). "Inverted Earth." *Philosophical Perspectives* **4**: 52-79.

Block, N. (1995). The Mind as the Software of the Brain. *An Invitation to Cognitive Science*. D. Osherson, L. Gleitman, S. Kosslyn, E. Smith and S. S. Cambridge (Mass), The MIT Press.

Block, N. (1999). Mental Paint. *forthcoming in a book of essays on Tyler Burge*. M. Hahn and B. Ramberg, The MIT Press.

Block, N., O. Flanagan, et al. (1996). *The Nature of Consciousness*. Cambridge (Mass.), The MIT Press.

Bohm, D. (1990). "A new theory of the relationship of mind and matter." *Philosophical Psychology* **3**(2): 271-286.

Boltuc, P. (1998). "Reductionism and Qualia." *Epistemologia* **XXI**: 111-130.

Boncinelli, E. (1999). *Il cervello, la mente e l'anima.* Milano, Mondadori.

Brentano, F. (1973). *Psychology From an Empirical Standpoint*. London, Routledge & Kegan paul.

Brooks, R. (1986). "A Robust Layered Control System for a Mobile Robot." *IEEE Journal of Robotics & Automation* **2**: 237-244.

Brooks, R. A. (1986). Achieving Artificial Inteligence through Building Robots, MIT AI-Lab,.

Brooks, R. A. (1990). "Elephants Don't Play Chess." *Robotics and Autonomous Systems* **6**: 3-15.

Brooks, R. A. (1991). "Intelligence Without Representations." *Artificial Intelligence Journal* **47**: 139-159.

Brooks, R. A. (1991). "New Approaches to Robotics." *Science* **253**(September): 1227-1232.

Brooks, R. A., C. Breazeal, et al. (1998). *Alternate Essences of Intelligence*. AAAI 98.

Brooks, R. A., C. Breazeal, et al. (2000). The Cog Project: Building a Humanoid Robot. T. a. i. a. S.-V. L. N. i. C. S. Volume.

Burge, T. (1992). "Frege on Knowing the Third Realm." *Mind* **101**(404): 633-650.

Burgess, A. (1994). *A dead man in Deptford*. London, Vintage.

Caracciolo, A. (1962). *La struttura dell'essere nel mondo e il modo del Besorgen in Sein und Zeit di Martin Heidegger*. Genova, Libreria Mario Bozzi.

Carnap, R. and M. Gardner (1995). *An introduction to the philosophy of science*. New York, Dover.

Carpenter, R. H. S. (1988). *Eye Movements*. London, Pion.

Carpenter, R. H. S. (1988). *Movements of the Eyes*. London, Pion Limited.

Carpenter, R. H. S. (1991). *Eye Movements*, The Macmillan Press.

Casalegno, P. (1997). *Filosofia del Linguaggio*. Roma, La Nuova Italia Scientifica.

Chalmers, D. (1996). "The Components of Content." *submitted to Mind*.

Chalmers, D. and A. Clark (1999). "The Extended Mind." (submitted).

Chalmers, D. J. (1996). *The Conscious Mind: in Search of a Fundamental Theory*. New York, Oxford University Press.

Chisholm, R. M. (1986). *Brentano and Intrinsic Value*. Cambridge (Mass.), Cambridge University Press.

Chomsky, N. (1968). *Language and Mind*. New York, Harcourt Brace.

Chomsky, N. (1988). *Language and Problem of Knowledge, The Managua Lecturers*. Cambridge (Mass.), The MIT Press.

Churchland, P. (1985). "Reduction, Qualia, and the Direct Inspection of Brain States." *Journal of Philosophy* **82**: pp.8-28.

Churchland, P. (1989). *A Neurocomputational Perspective*. Cambridge (Mass.), The MIT Press.

Churchland, P. (1990). *On the Nature of Theories: A Neurocomputational perspective*. Scientific Theories: Minnesota Studies in the Philosophy of Science.

Churchland, P. S. and T. J. Sejnowski (1992). *The Computational Brain*. Cambridge (Mass.), The MIT Press.

Cimatti, F. (2000). *La scimmia che si parla. Linguaggio autocoscienza e libertà nell'animale umano*. Torino, Bollati Boringhieri.

Cioni, G., M. Favilla, et al. (1984). "Development of the Dynamic Characteristics of the Horizontal Vestibulo-Ocular Reflex in Infancy." *Neuropediatrics* **15**: 125-130.

Clark, A. (1997). *Being there: putting brain, body and world together again*. Cambridge (Mass.), The MIT Press.

Clark, A. (1998). "Twisted Tales: Causal Complexity and Cognitive Scientific Explanation." *Minds and Machines* **8**: 79-99.

Coffa, A. J. (1998). *La tradizione semantica da Kant a Carnap*. Bologna, Il Mulino.

Cohen, L. G., P. Celnik, et al. (1997). "Functional relevance of cross modal plasticity in blind humans." *Nature* **389**(11 September): 180-182.

Conee, E. (1999). "Metaphysics and the morality of abortion." *Mind* **108**(432 October): 619-646.

Cramer, J. G. (1986). "The Transactional Interpretation of Quantum Mechanics." *Reviews of Modern Physics* **58**(July): 647-688.

Cramer, J. G. (1988). "An Overview of the Transactional Interpretation of Quantum Mechanics." *International Journal of Theoretical Physics* **27**(227).

Crick, F. (1994). *The Astonishing Hypothesis: the Scientific Search for the Soul*. New York, Touchstone.

Crick, F. and C. Koch (1990). *Toward a Neurobiological Theory of Consciousness*. Seminars in Neuroscience.

Crowley, J. L. (1985). *Navigation For An Intelligent Mobile Robot*. IEEE Journal on Robotics and Automation,.

Cui, Y., D. L. Swets, et al. (1995). *Learning-Based Hand Sign Recognition using SHOSLIF-M*. International Conference on Computer Vision (ICCV 95).

Curchland, P. S. and T. J. Sejnowski (1992). *The Computational Brain*. Cambridge (Mass.), The MIT Press.

D'Agostini, F. (1997). *Analitici e continentali*. Milano, Raffaello Cortina Editore.

Damasio, A. (1994). *Descartes' Error; Emotion, Reason, and the Human Brain*. New York, Avon Books.

Damasio, A. R. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York, Harcourt Brace & Company.

Damiani, A. (1994). "Carenze epistemologiche del connessionismo e limiti dei modelli connessionistici." *Sistemi Intelligenti* **1**(April): pp. 95-125.

Davidson, D. (1980). *Essays on Actions and Events*. Oxford, Oxford University Press.

Dawkins, R. (1990). *The Selfish gene*, Oxford University Press.

Del Re, G. (1991). *Il rapporto di Napoli sul problema mente-corpo*. Napoli, I.P.E.

Dennett, D. (1988). Quining Qualia. *Consciousness in Contemporary Science*. A. Marcel and E. Bisiach. Oxford, Oxford University Press.

Dennett, D. C. (1969). *Content and consciousness*. London, Routledge & Kegan Paul.

Dennett, D. C. (1978). *Brainstorms: philosophical essays on mind and psychology*. Montgomery, Bradford Books.

Dennett, D. C. (1987). *The intentional stance*. Cambridge (Mass.), The MIT Press.

Dennett, D. C. (1988). Quining Qualia. *Consciousness in Contemporary Science*. A. Marcel and E. Bisiach. Oxford, Oxford University Press.

Dennett, D. C. (1991). *Consciousness explained*. Boston, Little Brown and Co.

Dennett, D. C. (1995). *Darwin's dangerous idea: evolution and the meanings of life*. New York, Simon & Schuster.

Dennett, D. C. (1996). *Kinds of minds: toward an understanding of consciousness*. New York, Basic Books.

Dennett, D. C. (1998). *Brainchildren: essays on designing minds*. Cambridge (Mass.), The MIT Press.

Descartes, R. (1641). *Meditazioni Metafisiche*. Bologna, Patron.

Dewey, J. (1900). *Experience and Nature*.

Di Francesco, M. (1996). *Introduzione alla filosofia della mente*. Urbino, La Nuova Italia Scientifica.

Di Francesco, M. (1996). *L'io e i suoi sè.* Milano, Raffaello Cortina Editore.

Di Francesco, M. (2000). *La coscienza*. Bari, Laterza.

Drestke, F. (2000). *Perception, Knowledge and Belief*. Cambridge, Cambridge University Press.

Dretske, F. (1993). "Conscious Experience." *Mind* **102**(406): 263-283.

Dretske, F. (1995). *Naturalizing the Mind*. Cambridge (Mass.), The MIT Press.

Dretske, F. I. (1981). *Knowledge & the flow of information*. Cambridge (Mass.), The MIT Press.

Dretske, F. I. (1988). *Explaining behaviour: reasons in a world of causes*. Cambridge (Mass.), The MIT Press.

Dummett, M. (1998). *Origins of Analytical Philosophy*. Cambridge (Mass.), Harvard University Press.

Dummett, M. A. E. (1978). *Truth and other enigmas*. London, Duckworth.

Eckardt, B. v. (1993). *What is Cognitive Science?* Cambridge (Mass.), The MIT Press.

Eco, U. (1997). *Kant e l'ornitorinco*. Milano, Bompiani.

Eco, U. and T. A. Sebeok (1983). *Il Segno dei Tre*. MIlano, Bompiani.

Eddington, A. S. (1929). *The nature of the physical world*. New York, Macmillan.

Edelman, G. M. (1987). *Neural Darwinism. The Theory of Neuronal Group Selection*. New York, Basic Books.

Edelman, G. M. (1992). *Bright air, brilliant fire: on the matter of the mind*. New York, Basic Books.

Edelman, G. M. and G. Tononi (2000). *A Universe of Consciousness. How Matter Becomes Imagination*. London, Allen Lane.

Egan, F. (1990). "Individualism, Computation, and Perceptual Content.".

Eric, L. S. (1994). Computational Studies of the Spatial Architecture of the Primary Visual Cortex, Plenum Press,**:** 359--411.

Fayrabend, P. (1990). *Addio alla ragione*. Roma, Armando Editore.

Ferretti, G., G. Cioni, et al. (1998). "Visual information processing in infants with focal brain lesions." *Exp Brain Res* **123**(1-2): 95-101.

Findlay, J. N. (1963). *Meinong's theory of objects and values*. Oxford, Clarendon Press.

Fodor, J. (1998). *Concepts - Where Cognitive Science went Wrong*. Oxford, Oxford University Press.

Fodor, J. A. (1975). *The language of thought*. New York, Crowell.

Fodor, J. A. (1981). *Representations: philosophical essays on the foundations of cognitive science*. Cambridge (Mass.), The MIT Press.

Fodor, J. A. (1987). *Psychosemantics: the problem of meaning in the philosophy of mind*. Cambridge (Mass.), The MIT Press.

Frege, G. (1892). "Uber Sinn und Bedeutung." *Zeitschrift fur Philosophie und philosophische Kritik*(100).

Fukui, I. (1981). *TV image processing to determine the position of a robot vehicle*. Pattern Recognition,.

Fukushima, K. (1975). "Cognitron, A Self-Organizing Multilayered Neural Network Model." *Biological Cybernetics* **20**(121-136).

Fukushima, K., S. Miyake, et al. (1983). "Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition." *IEEE Transactions on Systems, Man, and Cybernetics* **13**: 826-834.

Fukushima, K., M. Okada, et al. (1994). "Neocognitron with Dual C-Cell Layers." *Neural Networks* **7**(1): 41-47.

Galilei, G. (1623). *Il Saggiatore*.

Gallistel, C. R. (1990). *The Organization of Learning*. Cambridge (Mass.), The MIT Press.

Gandolfo, F., G. Sandini, et al. (1996). *A field-based approach to visuo-motor coordination*. Workshop on Sensorimotor Coordination: Amphibians, Models, and Comparative Studies, Sedona, Arizona USA.

Gazzaniga, M. (1998). *The Mind's Past*. S.Francisco, University of California Press.

Geman, S., E. Bienenstock, et al. (1992). "Neural networks and the bias/variance dilemma." *Neural Computation* **4**: 1-58.

Gibson, W. (1998). *IDORU*. London, Grafton.

Gibson, W. and B. Sterling (1991). *The Difference Engine*. London, Bantam Spectra Book.

Giorello, G. and P. Strata (1991). *L'automa spirituale*. Bari, Laterza.

Girosi, F., M. Jones, et al. (1995). "Regularization Theory and Neural Networks Architectures." *Neural Networks* **7**: 219-269.

Goldberg, S. (1998). *Consciousness, Information and Meaning, The Origin of the Mind*. Miami, MedMaster.

Goldstein, E. B. (1996). *Sensation and Perception*, Brooks/Cole Publishing Company.

Goodman, N. (1978). *Of Mind and Other Matters*. Cambridge (Mass.), Harvard University Press.

Goodman, N. (1978). *Ways of Worldmaking*, The Harvester Press.

Goodman, N. (1979). *Fact, Fiction and Forecast*, The Harvester Press.

Griffin, D. R. (1998). *Unsnarling the world-knot: consciousness, freedom, and the mind-body problem*. Berkeley, University of California Press.

Hacking, I. (1975). *What does language matter to philosophy?* Cambridge, Cambridge University Press.

Haier, R. J., B. V. J. Siegel, et al. (1992). "Regional glucose metabolic changes after learning a complex visuospatial/motor task: a positron emission tomographic study." *Brain Research* **570**(1-2): 134-43.

Hameroff, S. H. (1994). "Quantum coherence in microtubules: A neural basis for an emergent consciousness?" *Journal of Consciousness Studies* **1**: 91-118.

Hameroff, S. H. (1998). Did Consciousness Cause the Cambrian Evolutionary Explosion? *Toward a Science of Consciousness II: The 1996 Tucson Discussions and Debates*. S. Hameroff, A. Kaszniak and A. Scott. Cambridge (Mass.), The MIT Press**:** 421-437.

Hameroff, S. H. (1998). "'Funda-mentality': is the conscious mind subtly linked to a basic level of the universe?" *Trends in Cognitive Neurosciences* **2**(4): 119-127.

Hameroff, S. H. (1998). "Quantum computation in brain microtubules? The Penrose-Hameroff 'Orch OR' model of consciousness." *Phil. Trans. R. Soc. London* **356**: 1-28.

Hameroff, S. R. (1994). "Quantum coherence in microtubules: A neural basis for an emergent consciousness?" *Journal of Consciousness Studies* **1**: 91-118.

Harman, G. (1989). "The Intrinsic Quality of Experience." *Philosophical Perspectives* **4**: 31-52.

Haugeland, J. (1997). *Mind Design II*. Cambridge (Mass.), The MIT Press.

Heidegger, M. (1988). *The basic problems of phenomenology*. Indianapolis, Indiana University Press.

Heisenberg, W. (1958). *Physics and Philosophy*. New York, Harper.

Hirai, K., M. Hirose, et al. (1998). *The development of Honda humanoid robot*. IEEE International Conference on Robotics and Automation, Leuven, Belgium.

Hofstadter, D. R. (1979). *Gödel, Escher, Bach: an eternal golden braid*. New York, Basic Books.

Hofstadter, D. R. (1985). *Metamagical themas: questing for the essence of mind and pattern*. New York, Basic Books.

Hofstadter, D. R. and D. C. Dennett (1981). *The mind's I: fantasies and reflections on self and soul*. New York, Basic Books.

Holt, J. (1999). "Blindsight in Debates about Qualia." *Journal od Consciousness* **6**(5): 54-71.

Hornik, K., M. Stinchcombe, et al. (1989). "Multilayer Feedforward Networks are Universal Approximators." *Neural Networks* **2**: 359-366.

Hughes, C. (1999). "Bundle Theory From A To B." *Mind* **108**(429 January): 149-156.

Humayun, M. S. and E. J. de Juan (1998). "Artificial vision." *Eye* **12**(3b): 605-7.

Humayun, M. S., E. J. de Juan, et al. (1996). "Visual perception elicited by electrical stimulation of retina in blind humans." *Arch Ophthalmol* **114**(1): 40-6.

Hume, D. (1956). *An enquiry concerning human understanding*. Chicago, Gateway Editions.

Humprey, N. (1992). *History of the Mind*. London, Chatto & Windus.

Husserl, E. (1969). *Ideas; general introduction to pure phenomenology*. London, Allen & Unwin.

Huxley, T. H. (1866). *Aphorisms and reflections*. New York, McMillan.

Inoue, H., H. Mizoguchi, et al. (1985). *A Robot Vision System with Flexible Multiple Attention Capability*. Proc. of ICAR,.

Ittyerah, M. and M. Goyal (1997). "Fantasy and reality distinction of congenitally blind children." *Percept Mot Skills* **85**((3 Pt 1)): 897-8.

Jackson, F. (1986). "What Mary didn't know." *Journal of Philosophy* **83**(5): 291-295.

Jackson, F. (1996). "Mental Causation." *Mind* **105**(419 July): 377-413.

Jacobson, L. and H. Wechsler (1985). *Joint spatial/spatial-frequency representations for image processing*. SPIE Int. Conf. on Intelligent Robots and Computer Vision,, Boston (MA).

James, W. (1890). *The Principles of Psychology*.

James, W. (1904). "Does 'consciousness' exist?" *Journal of Philosophy, Psychology and Scientific Methods* **1**: 477-491.

James, W. (1908). *A Pluralistic Universe*.

James, W. and B. Kuklick (1981). *Pragmatism*. Indianapolis, Hackett Pub. Co.

Jaynes, J. (1976). *The Origin of Consciousness in the Breakdown of the Bicameral Mind*.

Jones, B. (1975). "Spatial perception in the blind." *Br J Psychol* **66**(4): 461-72.

Kandel, E. R., J. H. Schwartz, et al. (1991). *Principles of Neuroscience*, Elsevier.

Kant, I. (1783). *Wiener Logik*.

Kant, I. (1958). *Critique of pure reason*. New York, Modern Library.

Katz, J. J. (1992). "The New Intensionalism." *Mind* **101**(404): 689-719.

Kentridge, R. W. and C. A. Heywood (1999). "The Status of Blindsight." *Journal od Consciousness* **6**(5): 3-11.

Khun, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago, The University of Chicago.

Kim, J. (1993). *Supervenience and Mind*. Cambridge, Cambridge University Press.

Kim, J. (1998). *Mind in a Physical World*. Cambridge (Mass.), The MIT Press.

Köhler, W. and H. Wallach (1944). "Figural after-effects." *Proceeding of the American Philosophical Society* **88**: 269-357.

Kosslyn, S. M. (1988). "Aspects of a Cognitive Neuroscience of Mental Imagery." *Science* **240**(17 Jun): 1621-1626.

Kosslyn, S. M., W. L. Thompson, et al. (1995). "Topographical representations of mental images in primary visual cortex." *Nature* **378**(30 november): 496-498.

Kreiman, G., C. Koch, et al. (2000). "Imagery neurons in the human brain." *Nature* **408**: 357-361.

Kripke, S. A. (1980). *Naming and necessity*. Cambridge (Mass.), Harvard University Press.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, University of Chicago Press.

Laeng, B., J. Shah, et al. (1999). "Identifying objects in conventional and contorted poses: contributions of hemisphere-specific mechanisms." *Cognition* **70**(1): 53-85.

Lanfredini, R. (1994). *Husserl, La teoria dell'intenzionalità*. Bari, Laterza.

Ledoux, J. (1996). In search of an emotional system in the brain: leaping from fear to emotion and consciousness. *The Cognitive Neuroscience*. M. Gazzaniga. Cambridge (Mass.), The MIT Press.

Ledoux, J. (1997). *The emotional brain*.

Leibneiz (1714). *Monadologia*, [cercare].

Levine, J. (1983). "Materialism and qualia: The explanatory gap." *Philosophical Quarterly* **64**: 354-61.

Libet, B. (1985). "Unconscious Cerebral Initiative and the Role of Conscious Will in Volutary Action." *Behavioural and Brain Sciences* **VIII**: 529-566.

Locke, J. (1690). Essay concerning Human Reasoning. Paris.

Lupacchini, R. (1997). "The Emergence of Phisical Meaning." *Epistemologia* **XX**: 33-66.

Lycan, W. G. (1990). "Consciousness as Internal Monitoring." *Philosophical Perspectives* **9**: 1-14.

Lyons, W. (1995). *Approaches to Intentionality*. Oxford, Harvard University Press.

Manzotti, R., G. Metta, et al. (1998). *Emotions and learning in a developing robot*. Emotion, Consciousness and Qualia, Naples and Ischia (Italy),.

Manzotti, R., G. Sandini, et al. (2001). "Disparity in log polar images." *Computer Vision and Image Understanding*.

Manzotti, R. E. S., G. Metta, et al. (1999). Make New Sense: Extending Consciousness by Developing New Meanings from External World. Genova, Lira Lab, DIST, University of Genova.

Manzotti, R. E. S., G. Metta, et al. (1999). "On building a conscious being." (submitted).

Manzotti, R. E. S. and G. Sandini (1999). "Intentionalizing Nature." (submitted).

Marr, D. (1982). *Vision*. S.Francisco, Freeman and Company.

Marr, D. (1991). *From Retina to Neocortex, Selected Paper*. Birkhauser, The MIT Press.

Martinoli, A., O. Holland, et al. (2000). Internal representations and Artificial Conscious Architectures, California Institute of Technology.

Marzi, C. (1999). "Why is Blindsight Blind?" *Journal of Consciousness* **6**(5): 12-18.

Massone, L. and E. Bizzi (1989). *A Neural Network Model for Limb Trajectory Formation*. NATO ARW on Robots and Biological Systems,, Il Ciocco, Tuscany, Italy,, Springer-Verlag.

Matsushita, T., S. Sakane, et al. (1990). *A System for Autonomous Execution of Hand-Eye Tasks -An Approach to Cooperative Hand-Eye Programming and Active Vision Sensing*. IEEE Intl. Workshop on Intelligent Robots and Systems,.

Maturana, H. R. and V. F.R. (1980). *Autopoiesis and cognition: the relization of the living*. Boston, Reidel.

Maturana, H. R., U. G., et al. (1968). "A theory of relativistic colour coding in the primate retina." *Arch. Biologica y Med. Exp.* **1**: 1-30.

Mc Dowell, J. (1994). *Mind and World*. Cambridge (Mass.), Harvard University Press.

McCarthy, J. (1995). "Making Robot Conscious of their Mental States." *http://www-formal.stanford.edu.jmc*.

McGinn, C. (1989). "Can we Solve the Mind Body Problem?" *Mind* **98**(891): 349-366.

McKennitt, L. (1997). The Book of Secret. London, Quinlan Road Limited.

Menzies, P. (1988). "Against Causal Reductionism." *Mind* **xcvii**(388): 551-574.

Messeri, M. (1997). *Verità*. Firenze, La Nuova Italia.

Metzinger, T. (1995). *Conscious Experience*. Schoningh, Imprint Academic.

Millikan, R. G. (1984). *Language , Thought , and other Biological Categories: New Foundations for Realism*. Cambridge (Mass.), The MIT Press.

Milner, D. A. and M. A. Goodale (1995). *The Visual Brain in Action*. New York, Oxford University Press.

Minsky, M. (1985). *The Society of Mind*. New York, Simon & Schuster.

Minsky, M. and S. Papert (1969). *Perceptrons: An Introduction to Computational Geometry*. Cambridge (Mass), The MIT Press.

Modenato, F. (1980). *Coscienza ed essere in Franz Brentano*. Bologna, Pàtron.

Moravia, S. (1986). *L'enigma della mente*. Roma, Laterza.

Morris J.S., A. O. a. R. J. D. D. (1998). "Conscious and unconscious emotional learning in the human amygdala." *Nature* **393**: 467-470.

Nagel, T. (1974). "What is it like to be a Bat?" *Philosophical Review* **4**: 435-450.

Nagel, T. (1995). *Other Minds: Critical Essays 1969-1994*. New York, Oxford University Press.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge (Mass.), Harvard University Press.

Newman, A. (1988). "The Causal Relation and its Terms." *Mind* **xcvii**(388): 529-550.

Nida-Rumelin, M. (1996). "Pseudonormal Vision: An actual case of qualia inversion?" *Philosophical Studies* **82**: 145-57.

Nijhawan, R. (1994). "Motion extrapolation in catching." *Nature* **370**: 256-7.

Nilsson, N., J. (1965). Learning Machines. New York.

Noonan, H. W. (1999). "Microphysical Supervenience and Consciousness." *Mind* **108**(432 October): 755-759.

Normann, R. A., E. M. Maynard, et al. (1999). "A neural interface for a cortical vision prosthesis." *Vision Research* **39**(15): 2577-87.

Oatley, K. (1978). *Perception and Representation*. London, Methuen.

O'Brien, g. and J. Opie (1999). "A connectionist theory of phenomenal experience." *Behavioural and Brain Sciences* **21**: 127-196.

Olgiati, F. (1953). *I fondamenti della filosofia classica*. Milano, Scotti Editore.

Oliver, A. (1996). "The Metaphysics of Properties." *Mind* **105**(417): 1-80.

O'Regan, K. J. (1992). "Solving the real misteries of visual perception: the world as an ourside memory." *Canadian Journal of Psychology* **46**(3): 461-488.

Panerai, F. and G. Sandini (1998). "Oculo-Motor Stabilization Reflexes: Integration of Inertial and Visual Information." *Neural Networks* **11**: 1191-1204.

Parisi, D. (1999). *La mente*. Bologna, Il Mulino.

Peirce, C. S. (1868). "Some Consequences of Four Incapacities." *Journal of Speculative Philosophy* **2**: 140-157.

Peirce, C. S. (1869). "How to Make Our Ideas Clear." *Popular Science Monthly* **12**(January): 286-302.

Penrose, R. (1989). *The Emperor's New Mind*. Oxford, Oxford University Press.

Penrose, R. (1994). *Shadows of the Mind*. Oxford, Oxford University Press.

Pessoa, L., E. Thompson, et al. (1999). "Finding out about filling-in: A guide to perceptual completion for visual science and the philosphy of perception." *Behavioural and Brain Sciences* **21**: 723-802.

Pirsig, R. M. (1974). *Zen and the art of motorcycle maintenance: an inquiry into values*. New York, Morrow.

Pirsig, R. M. (1991). *Lila: an enquiry into moral*. New York, Morrow.

Poggio, T. and V. Torre (1990). Ill-Posed Problems and Regularization Analysis in Early Vision, MIT A.I. Laboratory,.

Poggio, T., V. Torre, et al. (1985). "Computational Vision and Regularization Theory." *Nature* **317**: 314-319.

Pons, T. (1996). "Novel sensations in the congenitally blind." *Nature* **380**(11 April): 479-481.

Popper, K. R. (1959). *The Logic of Scientific Discovery*. New York, Basic Books.

Popper, K. R. and J. C. Eccles (1977). *The self and its brain*. New York, Springer International.

Popper, K. R. and P. A. Schilpp (1974). *The Philosophy of Karl Popper*. La Salle (Ill.), Open Court.

Porter, N. (1869). *The Human Intellect*. New York, Charles Scribner & Company.

Pulverlmuller, F. (1999). "Words in the brain's language." *Behavioural and Brain Sciences* **22**: 253-336.

Putnam, H. (1975). *Mind, language, and reality*. Cambridge, Cambridge University Press.

Putnam, H. (1983). *Realism and reason*. Cambridge (Mass.), Cambridge University Press.

Putnam, H. (1988). *Representation and reality*. Cambridge (Mass.), The MIT Press.

Putnam, H. (1992). *Renewing Philosophy*. Harvard, President and Fellows of Harwad College.

Quartz, S. R. and T. J. Sejnowski (1997). "The neural basis of cognitive development: a constructivist manifesto." *Behavioural and Brain Sciences*: 537-596.

Quartz, S. R. and T. J. Sejnowski (1997). "The neural basis of cognitive development: A constructivist manifesto." *Behavioural and Brain Sciences*(20): 537-596.

Quine, W. V. (1960). *Word and object*. Cambridge (Mass.), Technology Press of the Massachusetts Institute of Technology.

Quine, W. V. (1969). *Ontological relativity, and other essays*. New York, Columbia University Press.

Quine, W. V. (1973). *The roots of reference*. La Salle (Ill.), Open Court.

Quine, W. V. (1980). *From a logical point of view: 9 logico-philosophical essays*. Cambridge (Mass.), Harvard University Press.

Quine, W. V. (1996). *From a Logical Point of View*. Cambridge (Mass.), Harvard University Press.

Quine, W. V. O. (1951). "Two Dogmas of Empiricism." *Philosophical review*(60).

Ramachandran (1998). "Phantom limb." *Journal of Consciousness Studies*.

Ramachandran, V. S. and W. Hirstein (1997). "Three laws of qualia: what neurology tells us about the biological functions of consciousness." *Journal of Consciousness Studies* **4**(5/6): 429-57.

Ramachandran, V. S. and W. Hirstein (1998). "The perception of phantom limb." *Brain* **121**: 1603-30.

Rebaglia, A. (1990). "L'universo e la sua descrizione." *Epistemologia* **XIII**: 251-278.

Revonsuo, A. (1995). "Prospects for a cognitive neurosciences of consciousness." *Behavioural and Brain Sciences* **18**: 694-695.

Revonsuo, A. (1996). *Toward a multidisciplinary science of consciousness*. Downward Processes in the Perception Representation Mechanism, Ischia (NA) Italy, World Scientific.

Rodriguez, V. (1990). "Information and Its Meaning." *Epistemologia* **XIII**: 77-100.

Rogers and Howard (1995). *Binocular vision and stereopsis*, Oxford Clarendon Press.

Rojas, R. (1996). *Neural Networks: a Systematic Introduction*. New York, Springer.

Rosenberg, G. H. (1997). *A Place for Consciousness: Probing the Deep Structure of the Natural World*.

Russell, B. (1927). *The Analysis of Matter*. London, Routledge & Kegan Paul.

Russell, B. (1954). *The analysis of matter*.

Russell, B. (1979 (1927)). *An outline of Philosophy*. London, Allen & Unwin.

Russell, B. (1995). *The analysis of mind*. London, Routledge & Kegan Paul.

Ryle, G. (1984). *The concept of mind*. Chicago, University of Chicago Pres.

Sacks, O. (1985). *The man who mistook his wife for a hat*. New York, Knops.

Sadato, N., A. Pascual-Leone, et al. (1996). "Activation of the primary visual cortex by Braille reading in blind subjects." *Nature* **380**(11 April): 526-528.

Sandini, G., A. Alaerts, et al. (1998). *The Project SVAVISCA: a Space-Variant Colour CMOS Sensor*. AFPAEC'98, Zurich, SPIE.

Sandini, G., G. Metta, et al. (1997). *Human Sensori-motor Development and Artificial Systems*. AIR&IHAS '97.

Sandini, G., P. Questa, et al. (2000). *A Retina-like CMOS Sensor and its Applications*. SAM-2000, Cambridge, USA, IEEE.

Sandini, G. and V. Tagliasco (1980). "An Anthropomorphic Retina-like Structure for Scene Analysis." *Computer Vision Graphics and Image Processing* **14**: 365-372.

Sandini, G. and V. Tagliasco (1980). "An Anthropomorphic Retina-like Structure for Scene Analysis." *Computer Vision, Graphics and Image Processing* **14**(3): 365-372.

Santos-Victor, J. and G. Sandini (1997). "Embedded Visual Behaviours for Navigation." *Robotics and Autonomous Systems* **19**(3-4): 299-313.

Scassellati, B. (2000). *Theory of Mind for a Humanoid Robot*. First IEEE/RSJ International Conference on Humanoid Robotics.

Schiffman, H. R. (1996). *Sensation and Perception: an Integrated Approach*. New York, John Wiley & sons.

Schlick, M. (1938). Form and content: An introduction to philosophical thinking. *Philosophical Papers*. Dordrecht.

Schmid, R. and D. Zambarbieri (1991). Plasticita' e apprendimento nel controllo oculomotorio, Patron Editore,.

Schneider, H. (1947). *A History of American Philosophy*. New York, Columbia University Press.

Schwartz, E. L. (1977). "Spatial Mapping in the Primate Sensory Projection: Analytic Structure and Relevance to Perception." *Biological Cybernetics* **25**: 181-194.

Scruton, R. (1994). *Modern Philosophy*. London, Sinclair & Stevenson.

Searle, J. R. (1980). "Minds, Brains, and Programs." *Behavioural and Brain Sciences* **1**: 417-424.

Searle, J. R. (1983). *Intentionality, an essay in the philosophy of mind*. Cambridge (Mass.), Cambridge University Press.

Searle, J. R. (1984). *Minds, brains, and science*. Cambridge (Mass.), Harvard University Press.

Searle, J. R. (1985). *Intenzionalità*. Milano, Bompiani.

Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge (Mass.), The MIT Press.

Searle, J. R., D. C. Dennett, et al. (1997). *The mystery of consciousness*. New York, New York Review of Books.

Seitz, R. J., E. Roland, et al. (1990). "Motor learning in man: a positron emission tomographic study." *Neuroreport* **1**(1): 57-60.

Sellars, W. (1997). *Empiricism and the Philosophy of Mind*. Cambridge (Mass.), Harvard University Press.

Senden, V. (1932). *Raum und Gestaltauffassung bei operierten Blindgeborenen vor und nach der Operation (conception of space and gestalt in congenital blind children before and after surgery)*. Leipzig, Verlag Jd Ambrosus Barth.

Severino, E. (1990). *Legge e caso*. Milano, Adelphi.

Shannon, C. E. (1948). "A Mathematical Theory of Communication." *The Bell System Technical Journal* **27**(July): 379-423, 623-656.

Shannon, C. E. and W. Weaver (1949). *The Mathematical Theory of Communication*. Urbana, University of Illinois Press.

Shapiro, S. (1993). "Modality and Ontology." *Mind* **102**(407): 455-481.

Shepard, R. N. and N. Metzler (1971). "Mental Rotation of Three Dimensional Figures." *Science* **171**: 701-703.

Sheperd, G. M. (1988). *Neurobiology*. New York, Oxford University Press.

Shoemaker, S. (1982). "The Inverted Spectrum." *Journal of Philosophy* **81**(7): 357-381.

Shoemaker, S. (1990). "First Person Access." *Philosophical Perspectives* **4**: 187-214.

Shoemaker, S. (1990). "Qualities and Qualia: What's in the Mind?" *Philosophy and Phenomenological Research* **50**: 109-131.

Shoemaker, S. (1994). *The First Person Perspective*. APA Proceedings.

Smith, B. C. (1996). The Foundations of Computing, Smith, Brian Cantwell. **1996**.

Smith, B. C. (1998). *On the Origins of Objects*. Cambridge (Mass.), The MIT Press.

Somenzi, V. and R. Cordeschi (1986). La filosofia degli automi. Torino, Bollati Boringhieri.

Specht, D. F. (1990). "Probabilistic Neural Networks." *Neural Networks*(3): 109-118.

Springer, S. P. (1981). *Left Brain, Right Brain*. S.Francisco, Freeman and Company.

Srinivasan, M. V. and S. Venkatesh (1997). *From living eyes to seeing machines*. London, Oxford University Press.

Stapp, H. P. (1998). *Whiteheadian Process and Quantum Theory of Mind*. Silver Anniversary International Conference, Claremont (Cal.).

Statton, G. M. (1897). "Upright Vision and the Retinal Image." *Psychological Review* **4**: 182-187.

Stein, B. E. and M. A. Meredith (1999). *The merging of the senses*. Cambridge (Mass.), The MIT Press.

Strawson, G. (1994). *Mental reality*. Cambridge (Mass.), The MIT Press.

Stubenberg, L. (1998). *Consciousness and qualia*. Amsterdam, J. Benjamins Pub.

Sturgeon, S. (1998). "Physicalism and Overdetermination." *Mind* **107**(426 April): 411-432.

Sugita, Y. (1996). "Global plasticity in adult visual cortex following reversal of visual input." *Nature* **380**(11 April): 523-526.

Sutton, R. S. and A. Barto (1998). *Reinforcement Learning: an Introduction*. Cambridge, MIT Press.

Sutton, R. S. and A. G. Barto (1998). *Reinforcement Learning*. Cambridge (Mass.), The MIT Press.

Tagliasco, V. (1999). *Dizionario degli esseri umani fantastici e artificiali*. Milano, Mondadori.

Tratteur, G. (1991). *Io, Anima e Software*. Il Rapporto di Napoli sul problema mente-corpo, Napoli, I.P.E.

Trinkaus, E. and S. P. (1992). *The neanderthals: changing the image of mankind*. New York, Alfred E. Knopf.

Trundle, R. C. (1990). "Existentialism and Phenomenology: the Overlooked Bases of Scientific realism." *Epistemologia* **XIII**: 279-302.

Turing, A. (1950). "Computing Machinery and Intelligence." *Mind* **59**: 433-460.

Tye, M. (1990). "Representational Theory of Pains and their Phenomenal Character." *Philosophical Perspectives* **9**: 223-239.

Tye, M. (1991). *The Imagery Debate*. Cambridge (Mass.), The MIT Press.

Tye, M. (1996). *Ten Problem of Consciousness*. Cambridge (Mass.), The MIT Press.

Tye, M. (1999). "Phenomenal Consciousness: The Explanatory Gap as a Cognitive Illusion." *Mind* **108**(432 October): 705-725.

Uexküll, J. v. (1909). *Umwelt und Innenwelt der Tiere*. Berlin, J. Springer.

Uexküll, J. v. (1934). *A Stroll through the Worlds on Animals and Men: A Picture Book of Invisible Worlds*. New York, International University Press.

Vaiana, L. (2000). *La nuova sfida del meccanicismo*. Roma, Armando Editore.

van Gulick, R. (1993). Understanding the phenomenal mind: Are we all just armadillos? *Consciousness: A Mind and Language Reader*. M. Davies and G. Humphreys. Oxford, Blackwell.

Varela, F. (2000). *Neurophenomenology*. Tucson 2000, Tucson.

Varela, F. J. and J. Shear, Eds. (1999). *The view from within. First-person approaches to the study of consciousness*, Imprint Academic.

Von Eckardt, B. (1993). *What is cognitive science?* Cambridge (Mass.), The MIT Press.

Weng, J. J. (1996). Cresceptron and SHOSLIF: Toward comprehensive visual learning. *Early visual learning*. S. K. Nayar and T. Poggio. New York, Oxford University Press.

Weng, J. J. (1998). *The Developmental Approach to Intelligent Robots*. 1998 AAAI Spring Symposium Series, Integrating Robotic Research: Taking The Next Leap, Stanford University.

Whitehead, A. (1928). *Symbolism, its Meaning and Effects*. London, Cambridge University Press.

Whitehead, A. N. (1925). *Science and the modern world*. New York, Free Press.

Whitehead, A. N. (1927). *Process and Reality*. London, The Free Press.

Whitehead, A. N. (1929). *Process and reality, an essay in cosmology; Gifford lectures delivered in the University of Edinburgh during the session 1927-28*. New York, The Macmillan Company.

Whitehead, A. N. (1933). *Adventures of ideas*. New York, Free Press.

Whitehead, A. N. (1938). *Modes of thought*. New York, The Macmillan company.

Whitehead, A. N. (1958). *Symbolism: its meaning and effect*. New York, Macmillan.

Whitehead, A. N. (1978). *Process and Reality*. London, The Free Press.

Wiener, N. (1961). *Cybernetics, or Control and Communication in the Animal and the Machine*. Cambridge (Mass), The MIT Press.

Wilson, R. A. and F. C. Keil (1999). *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge (Mass.), The MIT Press.

Wilson, S. W. (1983). "On the retino-cortical mapping." *International Journal of Man-Machine Studies* **18**: 361-389.

Wittgenstein, L. (1974). *Tractatus logico-philosophicus*. London, Routledge & Kegan Paul.

Wittgenstein, L. (1995). *Ricerche filosofiche*. Torino, Einaudi.

Zeki, S. and A. Bartels (1998). "The asynchrony of consciousness." *Proc R Soc Lond B Biol Sci* **265**: 1583-5.

Zimler, J. and J. M. Keenan (1983). "Imagery in the congenitally blind: how visual are visual images?" *J Exp Psychol Learn Mem Cogn* **9**(2): 269-82.

Zohar, D. and I. Marshall (1993). *The Quantum Society: Mind, Physics and a New Social Vision*. London, Bloomsbury.

Index