

	<p>ADAPT <i>IST-2001-37173</i> <i>Artificial Development Approach to Presence Technologies</i></p>
---	---

Deliverable Item 2.3

A common psycho-physical vocabulary

Delivery Date: September 1st, 2005

Classification: Public

Responsible Person: Riccardo Manzotti and Giorgio Metta – UGDIST

Partners Contributed: ALL

Short Description: The development of robotics and psychology (as well as other related disciplines like neuroscience, philosophy of mind, neurophilosophy) will lead to the development of a common language. Historically the development of the study of the mind made use of a different language than the so called physical sciences. Historically the first challenge has been provided by psycho-physics and neuroscience which tried to translate the mental domain in a physical domain. Going in the opposite direction but sharing the same long term goal, robotics and AI are unique since they try to bridge the gap between the physical description of the world and the mental one. A few concepts are crucial and need to be analyzed: intentionality, representation, phenomenal experience, motivation, goal, presence. This delivery shows an attempt at finding a common vocabulary between traditionally separate disciplines by using the project results.

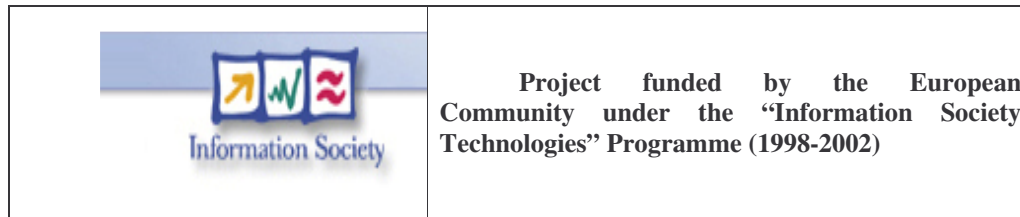


Table of contents

TABLE OF CONTENTS	2
1. INTRODUCTION.....	3
2. PRESENCE AND CONSCIOUSNESS	6
2.1.1 Two caveats: content and mental	9
3. REQUIREMENTS FOR A PSYCHO-PHYSICAL VOCABULARY	11
3.1 Ontological economy (Ockam’s razor)	11
3.2 Direct experience	12
3.3 Explicative power and predicting capability	13
3.4 Experiential adequacy	14
3.5 The compatibility of empirical science	15
3.6 Everyday experience compatibility	16
4. PSYCHO-PHYSICAL VOCABULARY.....	18
5. REFERENCES.....	21

1. Introduction

Historically there has been a separation between the study of the mind and the physical sciences. As a result a different terminology developed. The reasons of this fact are several and we can only offer a rough summary of them. The mind in his modern form has been inherited by the philosophical approach due to Descartes, Galileo and Locke. According to them the mental domain was ontologically separate from the physical domain. The reason for this assumption was due mainly to those aspects of the mind which cannot be reduced to the known properties of the physical world. The two most important ones were the teleological aspect of organism and the phenomenal aspect of experience. Both aspects were apparently lacking in all the known physical phenomena.

Further development of the study of the mind, under the guidance of philosophers and psychologists, led to a more refined definition of the properties of the mind which was none the less incompatible with the physical picture of the world. This baffling situation is well testified by recurring statement in the history of philosophy of mind. For instance, in 1866 Thomas Huxley wrote that

How it is that 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. (Huxley 1866)

Classic authors like Bertrand Russell had no problem in admitting their difficulties even with the more simple case of direct 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 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 points, like a stretched rope when it snaps (Russell 1954)

The same situation is reflected in text book for neurosciences, like the Goldstein's handbook on sensation and perception

Even the most brilliant scientist could not tell how electrical signals in the brain become perceptions (Goldstein 1996)

Or on the same topic, more recently, the philosopher Jaegwon Kim wrote that

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" (Kim 1998)

And other authors like Michael Tye echoed

Look at the neurons for as long as you like, and you still will not find phenomenal consciousness (Tye 1996)

The difference between the mental domain and the physical domain is paralleled by the difference 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. There are three correlated concepts: being a person, being a human being, and being a subject. The first concept is juridical 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 (Khun 1962).
- Several species *mutates mutandis* (cats, dogs, dolphins, monkeys) could deserve the status of subjects (Allen 1997).
- 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 her consciousness, she is 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 (Chalmers 1996; Kim 1998; Edelman and Tononi 2000), 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 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 (Levine 1983). And someone even argued that the relation between the two will be forever unknown to men: a modern *ignoramus et ignorabimus* (McGinn 1989). The problem of the nature of phenomenal consciousness has become so obsessively difficult that has become known as *hard problem* (Chalmers 1996).

Notwithstanding these difficulties there is a growing belief that new techniques or new approaches will allow a scientific understandings of the nature of consciousness and, inasmuch, of presence, of subjectivity, of intentionality, of phenomenal experience. As the editor of Nature Neuroscience wrote that

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. (Jennings 2000)

In a similar fashion, Gerard Edelman claimed that

To understand the mental we may have to invent further ways of looking at brains. We may even have to synthesize artifacts resembling brains connected to bodily functions in order fully to understand those processes. Although the day when we shall be able to create such conscious artifacts is far off we may have to make them before we deeply understand the processes of thought itself. (Edelman and Tononi 2000)

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 (Aleksander 1994; McCarthy 1995; Aleksander 1996; Martinoli, Holland et al. 2000)(Nilsson and Ojala 1995; Steels 1995; Schlagel 1999; Buttazzo 2000).

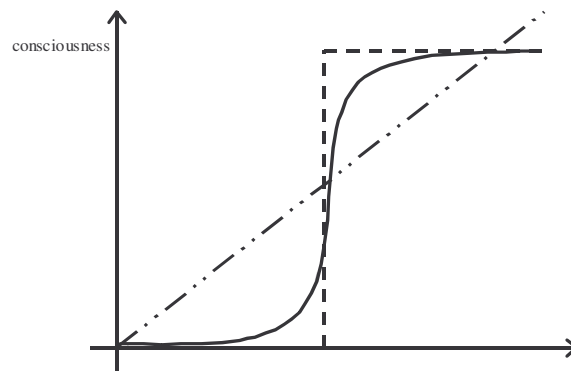


Figure 1 Up to now, it is not clear what the essential element that produces phenomenal consciousness is. It is not clear if consciousness arises gradually or abruptly. In the figure there are a few possible curves. For example the dotted line corresponds to *on-off* theories of consciousness. It must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

It is surprising to observe that the playground on which the mind is studied resembles so much the framework suggested in the XVIIth century. Although nowadays the role of the soul is taken by the brain plus the body, the way in which the problem is framed is still the same (for instance in (Dennett 1991; Chalmers 1996; Kim 1996; Edelman and Tononi 2000)). So far, this view led to many unsolved problems. The most important of which are:

- if secondary properties (or qualities) are separate and different from primary properties, how can secondary properties arise out of primary ones (*hard problem*)?
- if secondary properties (or qualities) are separate and different from primary properties, how can secondary properties produce effects on them (*mental causation*)?
- if secondary properties (or qualities) are separate and different from primary properties, how can secondary properties represent any properties of the external world (*representation and intentionality*)?

2. Presence and consciousness

The development of a real psycho-physical vocabulary is necessarily linked with the development of a theory of consciousness and the mind. Such a theory is not yet available although this project is among the attempts at providing a successful one.

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:

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. (Chalmers 1996)

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 (Davidson 1980; Churchland 1989; McGinn 1989; Kim 1993).

What could be the structure of a successful vocabulary capable of crossing the bridge between the mind and the physical world? Before suggesting a possible solution let us briefly analyze the available solution to this problem. 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. An infinite literature is concerned with the status that must be given to phenomenal entities (Galilei 1623; Descartes 1641; Locke 1690; Leibniz 1714; Eddington 1929; Nagel 1974; Kripke 1980; McGinn 1989; Shoemaker 1990; Shoemaker 1994; Strawson 1994; Russell 1995; Chalmers 1996; Stubenberg 1998; Block 1999). 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 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 noumenally 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 (Churchland 1985; Churchland 1989; Churchland 1990; Crick and Koch 1990). 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. (Crick 1994)

Comment [m1]: Controllare citazione precisa dal libro

Not differently Christoph Koch stated the same principle in his (Koch 2004). 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? How can they refer to something that is in the external environment? 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. A related approach is given by the so-called representational theory that presupposes some innate representational medium in the brain (Fodor 1987; Pulvermuller 1999). 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 (Armstrong 1968). However, they are internalist regarding where to locate the physical medium for representation.

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 (Putnam 1975). 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 Dretske and Tye (Dretske 1993; Dretske 1995; Tye 1996). 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 (Putnam 1975; Dennett 1996). 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. For example, an interesting survey is provided by (Block 1999) or by (Tye 1991). The theory 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 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?»

2.1.1 *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 (Kim 1998). 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. Secondly, 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 (O'Brien and Opie 1999), 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 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. A similar position was maintained by Franz Brentano and, more recently, by Joh Searle. According to Brentano, there are no unconscious psychical phenomena.

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 H₂O 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 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.

3. Requirements for a psycho-physical vocabulary

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. (Chalmers 1996)

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. (Clark 1997)

Before entering into the details of the proposed vocabulary, 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 uphold 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 judgment, 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 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.

3.1 *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.

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 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) 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 (Nagel 1974).

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.

3.2 *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 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. A classical example of this kind of subjective

judgment is given by the paradox of phenomenal judgment (Chalmers 1996), p. 150. Our 'objective' judgments 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).

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? 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 behavior. 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.

3.3 *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 skepticism.

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.

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

The optimum would be to find a crucial experiment, as has happened for most of the scientific theories that effectively have revolutionized 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 (for example dealing with the when and the how of conscious phenomenal experience) and in which only one theory succeeds in predicting it.

3.4 *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 (**Error! Reference source not found.**). 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 latter. Objective physicalistic metaphysics has failed to achieve its ambitious aim: so much the worse.

The idea that subjective facts are real has gained wider and wider acceptance. 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. (Stubenberg 1998),

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

3.5 *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 chapter in the last volume of a neurosciences encyclopædia. 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).

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.

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. (Chalmers 1996)

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 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 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 our 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» (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). 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.

3.6 *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.

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 (Fodor 1987), p. xii.

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 (Di Francesco 1996), p. 18.

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

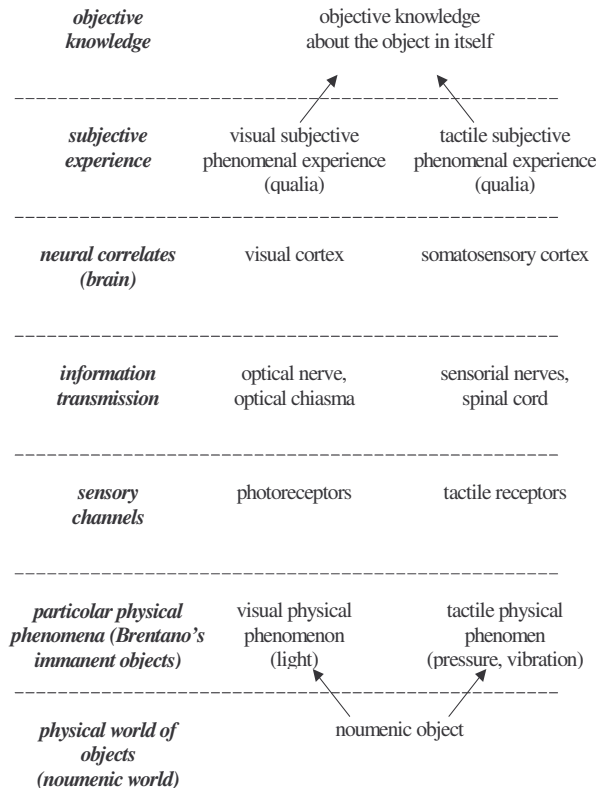


Figure 1 A possible diagram of the phases of perception. The physical world is at the bottom while objective knowledge is at the top. Although this series of stages is more or less accepted there are several dishomogeneities lurking between the levels. If there were an efficient theory of mind, such dishomogeneities ought to disappear.

4. Psycho-Physical vocabulary

A psycho-physical vocabulary must be based on some a theory of mind. In this case we will base our attempt on the theory presented in the deliverable 2.1; the theory labeled Theory of the Enlarged Mind whose fundamental ontological claim is the onphene.

As we have argued there are three fundamental pillars of this theory: externalism, process view, realism.

Process or onphene. This is the basic constituent of reality. According to the process view advocated here, reality is made of processes. This position has been already defended, in different ways, by many different authors (Russell 1924; Whitehead 1929/1978; Davidson 1980; Stapp 1998; Mc Henry 1999; Skrbina 2005). In short it suggests that at the rock bottom of reality there are only processes and not static objects. It is a view of reality which is perfectly compatible with physicalism and with the data offered by contemporary science.

Event. A recognizable chunk of reality whose presence or absence does make a difference. As we have argued the concept of event is secondary with respect to the concept of process. A useful analogy is offered by the north and south pole of a magnetic field. They do not really exist as two separate parts of a magnetic field; they are two conventional and very useful way to refer to the same phenomenon: the magnetic field. Similarly, here the events are seen as conventional ways to refer to processes. By applying Davidson's principle according to which events can be identified by means of other events (their effects), it is possible to argue that an isolated event is an abstraction. Events take place as part of a network of processes. An isolated event is as impossible as an isolated magnetic pole.

Cause. Whenever we have a process we can refer to its beginning as its cause. Therefore we can isolate a part of the process (which would be otherwise not existent) as the event cause of the process. Yet, the cause takes place only by means of the effect. This is to say that the cause and the effect (a pair of events) are not isolated and autonomous. They are two perspectives on the same process.

Effect. The event which is identified as the end of a process. Due to the fact that there are no isolated events, the effect of another event can be conceived only as the cause of a further process. In this way we cannot be aware of the final part of the processes which constitute us until they became causes of further processes. This is confirmed by the fact that we are aware of the content of our thoughts and not of our thoughts as thoughts unless our thoughts become the content of further thoughts.

Content. According to the presented theory, no onphene is without content; content being the same as the occurrence of an onphene. Since each onphene is a process, each onphene carves out a part of reality. Such part of reality can be considered the content of the onphene. However assuming that the content and the onphene are different is just a terminological mismatch. There is no need to assume any difference between them and we could as well drop the use of the word "content". On the contrary, the use of some kind of content is necessary for those theories of the mind that assumes the existence of some kind of a priori epistemic structure (dualism, cognitivism, functionalism share this need).

Representations. As there is no need of content, so there is no need of representations. The idea of representation is deeply based on a dualistic view of reality or, at least, on the assumption on the separation between the subject and the object. As we have previously argued, a re-presentation assumes a difference between what is represented and its representation. According the view presented here, there is no more need to distinguish between the two. An onphene represents something since it is that something. For instance, in order to represent a table under a certain perspective, an onphene must be constituted by that table; that table being the event causing the onphene. As we said earlier, this does not mean that the event is autonomous with respect to the process. The event is just one way to describe the taking place of the process. All the brain structure considered a support for representing the external world are the physical support that allow to a collection of processes to take place. According to this view the real “representation” are not the structure in the brain, but the physical processes extending in time and space from the external environment to the inside of the brain. We can call them representations since they play the same role of classic representation. However, it would be more correct to talk only of “presentation”. In short, we could say that the presented theory advocates an identity-theory of perception. What about conventional representation like words, signs, characters and information? They are the support of processes which are the real presentation insofar are causally connected and thus a unity with what they should represent. For instance, a card with my name on it represents myself insofar it is a causal extension of myself.

Sensations. Sensation corresponds to the simplest level of onphenes involved in a subject. No relation between them is contemplated and thus it is made of processes which are, in a way, absolute. They do not share any kind of relation.

Perceptions. Presentations of higher order. They are based on the lowest level of onphenes and are, of course, just onphenes connecting more complex combinations of other onphenes. They are the foundation of the feeling of being there. A perception (or a sensation) is just a presentation embodied by an onphene. What about the subjective side of sensations and perceptions? According to this theory there is no need to distinguish between a subjective rainbow and a objective one. Both take place, both are private, both have the same relevant properties (being coloured, being an arch, being a private physical phenomenon, having “perspectivalness”).

Memory. Perception delayed in time or, which is the same, presentations with a longer than usual time span.

Dream. Incoherent perception delayed in time.

Illusion. Incoherent perception without time delay.

Aboutness or intentionality in the Brentano’s sense. According to a dualistic view, mental states possess intentionality or aboutness in the sense of an arrow pointing from the mental state to its content (whatever this is, a physical reference or a immanent object). Conceived as such, intentionality has been both baffling and elusive. No satisfying account has been presented until now. According to our theory, there is no more need of this kind of

intentionality. Since the fundamental constituent of reality is a process, it can be seen as originally intentional in nature. The onphene, as a process, links different parts of reality. The mind does not need to “be about its content”, since the mind and its content are the same: they are an onphene. This is the end of intentionality in the classic Brentano’s sense.

Goal. A goal is an event which by taking place is going to increase the probability of taking place again. Reproduction is such an event. If an agent acquires the capability to reproduce, the act of reproduction is going to take place more frequently. Other event have the same structure. For instance, recognition of something would lead to further recognition of the same something. Learning a skill would lead to the exploitation of the same skill. Nothing like playing a sport, to make more likely to play again the same sport. A goal, which is usually conceived as “representation of a desirable future state of things” is a past event with the aforementioned property of increasing its own repeatability. This kind of events are constitutive of living organisms and intelligent agents. Goals therefore come from the past. A goal can be instantiated by an organism or an agent without it having any kind of representation of it.

Mind. The mind is a collection of presentations or onphene (or causal processes) whose unity is achieved by means of a common goal. Since the mind physically spans time and space comprehending those events that constitute its content (old terminology), we use the expression “the enlarged mind”. The mind is not inside the boundaries of the skull or of the skin. The mind is no longer an emergent property of neural processes. The brain is no sufficient of the mind. The mind is a collection of processes, taking place thanks to the brain, the body and the external world, extending in time and space and including all that is experienced by the subject herself.

Subject. A mind in his historically development.

Intentionality as volition. Intentionality has another meaning which corresponds to being motivated to accomplish a certain event or state of affairs. According to our theory an agent is a collection of onphenes which gets its unity by being aimed at a common goal. Therefore the unity of the subject is due to the fact that the all the processes taking place share the same goal. This is different from other criteria (for instance being instantiated by the same body or brain). Thus, intentionality as volition is the having of goals. There cannot be a subject without intentionality: it would have no unity and thus it would not be.

Motivation. A goal which is at the same time a representation for the agent: a perceived goal.

Presence or sense of being there. At a given instant in time, presence is the collection of all the event causes of the processes constituting the mind at the same instant. Presence is an integrated collection of processes that merge together different and otherwise scattered events in the environment. Presence usually requires the integration of several sensor modalities which carve out different domains of events (visual, tactile, acoustic, etc.). Presence is conceivable only in a subject.

In conclusion, we made a series of hypothesis here. Let us summarize them:

- presence is due to a series of phenomenal mental events that are contentful and integrated;
- the intentional and phenomenal status of mental events is due to their identity with physical processes that include the external target of these events;
- these processes have a role in shaping both the environment and the subjective experience; for this reason they have been named onphenes;
- a collection of these processes (or onphenes) constitutes a moment of presence
- the unity between separate onphenes is due to the progressive entanglement of causal processes in order to achieve a goal;
- the final unity of separate onphenes (which possess intentionality in the philosophical sense of aboutness) is eventually achieved by their cooperation to reach a given goal thus obtaining intentionality in the psychological sense.

5. References

- Buttazzo, G. (2000). Can a machine ever become self-aware? Artificial Humans. R. Aurich, W. Jacobsen and G. Jatho. Los Angeles, Goethe Institute: 49-49.
- Chalmers, D. J. (1996). The Conscious Mind: In Search of a Fundamental Theory. New York, Oxford University Press.
- Crick, F. (1994). The Astonishing Hypothesis: the Scientific Search for the Soul. New York, Touchstone.
- Davidson, D. (1980). Essays on Actions and Events. Oxford, Clarendon Press.
- Dennett, D. C. (1991). Consciousness explained. Boston, Little Brown and Co.
- Edelman, G. M. and G. Tononi (2000). A Universe of Consciousness. How Matter Becomes Imagination. London, Allen Lane.
- Goldstein, E. B. (1996). Sensation and Perception, Brooks/Cole Publishing Company.
- Huxley, T. H. (1866). Aphorisms and reflections. New York, McMillan.
- Jennings, C. (2000). "In Search of Consciousness." Nature Neuroscience 3(8): 1.
- Kim, J. (1996). Philosophy of Mind. Boulder, Westview Press.
- Kim, J. (1998). Mind in a Physical World. Cambridge (Mass), MIT Press.
- Koch, C. (2004). The Quest for Consciousness: A Neurobiological Approach. Englewood (Colorado), Roberts & Company Publishers.
- Mc Henry, L. B. (1999). "Revisionary and Descriptives Theories of Events." Process Studies.
- McGinn, C. (1989). "Can we Solve the Mind Body Problem?" Mind 98(891): 349-366.
- Nilsson, M. and J. Ojala (1995). Toward Conscious Robots: "Self-awareness" Speeds Learning. Robotikdagarna, Linkoping (sweden).
- Russell, B. (1924). The Analysis of mind. London, New York, G. Allen & Unwin; Macmillan.
- Russell, B. (1954). The analysis of matter.
- Schlagel, R. H. (1999). "Why not Artificial Consciousness or Thought?" Minds and Machines 9: 3-28.
- Skrbina, D. (2005). Panpsychism in the West. Cambridge (Mass), MIT Press.
- Stapp, H. P. (1998). Whiteheadian Process and Quantum Theory of Mind. Silver Anniversary International Conference, Claremont (Cal).

Steels, L. (1995). Is artificial consciousness possible? Consciousness: Distinction and Reflection. G. Trautteur. Napoli, Bibliopolis.

Whitehead, A. N. (1929/1978). Process and Reality. London, Free Press.