

CRASH-TESTING THE SENSORIMOTOR HYPOTHESIS: THE CASE OF SPATIAL COMPETENCE

Dario Taraborelli
Institut Jean Nicod
CNRS EHESS ENS
Paris
taraborelli@ens.fr

Matteo Mossio
LPPA
Collège de France
Paris
matteo.mossio@college-de-france

Abstract

The study of capabilities a perceptual system can acquire from systematic exposure to sensorimotor correlations, i.e. the regular co-occurrence of motor patterns and sensory states, has become a major research trend in current cognitive science. Defendants of the sensorimotor approach have claimed that the study of sensorimotor learning can provide a coherent and parsimonious framework for understanding how specific perceptual capabilities are acquired and used. Yet, lacking a solid characterization of key theoretical notions like ‘sensorimotor invariant’, these alternative approaches can hardly provide more accurate explanations than those of mainstream perceptual research. In this paper, we fix a number of general constraints that any sensorimotor theory must meet in order to provide a valid alternative to traditional models. By focusing on the case-study of spatial competence, i.e. what knowledge a system must possess in order to behave spatially, we address the issue of what distinguishes sensorimotor theories from traditional approaches. We show how, in the case of spatial cognition, the sensorimotor approach can provide a viable and promising scientific programme for explaining the acquisition and use of spatial capabilities without making use of internal representations of external space. We finally suggest possible research directions related to the study of spatial competence in terms of fixation of sensorimotor invariants.

Keywords

Sensorimotor invariants – enaction – perception and action – spatial representations – frames of reference – space.

INTRODUCTION

1. Explaining spatial competence

In current cognitive literature, a number of very different kinds of capabilities are grouped together under the label of “spatial competence”. A survey of the literature on spatial cognition shows that tasks such as: *grasping an object by visuomotor co-ordination between the eyes and the arm; repeating a previously performed path; evaluating the distance and position of a sound source; estimating the absolute size of a visually presented object* – are all studied as particular instances of a more general capacity displayed by cognitive systems to manage with spatial properties of the environment.

Cognitive scientists engaged in understanding similar cases of spatial competence should provide an explanation of the requirements a system has to meet in order to behave spatially. Explaining spatial competence should give an answer to questions like:

- How do cognitive systems succeed in realizing spatial tasks?
- What kind of knowledge is needed in order to perform spatial tasks in a correct and efficient way?
- How do cognitive systems acquire this knowledge?

The classical strategy adopted in cognitive science to provide an explanation for spatial behaviour is the appeal to the appropriate *representational resources* mediating the correct accomplishment of the task. Our analysis starts by addressing the problem of how to constrain the notion of spatial representation evoked in mainstream explanations of spatial competence.

The concept of “spatial representation” is in most cases a very broad (and consequently under-constrained) theoretical notion: almost every kind of internal state invoked to explain spatial competence can be considered as a *prima facie* representational resource, insofar as it can be said to *correlate* with some spatial properties of the environment and carry information about these properties in virtue of this correlation. “*A representational system can be analysed as a homomorphism: a mapping from objects in one domain to objects in another domain such that relations among objects in the first domain are mirrored by corresponding relations among corresponding objects in the representation*” – adapted from Palmer (1999).

In order to understand how representational resources can explain spatial capabilities, we need to analyse at a finer grain what is implied, in current cognitive science, by the notion of spatial representation. We will argue that the “traditional view” endorses a specific idea of spatial representations and their alleged explanatory role. According to the “traditional view”, as we will see, any explanation of spatial capabilities must account for:

- What different *internal* representations of *external* spatial properties a system must possess;
- What their mutual relations must be (their global architecture).

In what follows, we will try to contrast this traditional view with a different approach (the sensorimotor paradigm) that, although compatible with the general idea that some representational resources are needed for explaining spatial knowledge, challenges the fundamental tenet of the traditional approach, i.e. the idea according to which explaining spatial competence means describing how *internal* representations refer to *external* space

In SECTION I of this paper, we will focus on some methodological and conceptual

consequences of this challenge. In particular we will try to show that the traditional and the sensorimotor account critically diverge on the answers they give to each of the following questions:

- What kind of structures underlie spatial competence?
- How complex must be the architecture of a spatially competent system?
- What has to be explained by a model of spatial competence?

The outcome of this analysis will support the view, outlined in SECTION II, that the sensorimotor paradigm allows for a more *general* characterisation of spatial competence that avoids the appeal to internal representations of external space. We will finally argue (CODA) that a sound scientific explanation of spatial competence should not rely on an intuitive notion of space: the kind of models provided by the sensorimotor paradigm seems to avoid the appeal to such an intuitive notion, that lie at the core of traditional approaches.

2. Terminological note

In this article, we will refer to two notions that echo the classical distinction between competence and performance in the linguistic domain.

By '*spatial competence*' we mean the possession of items of knowledge (or internal structures) that are required by a system in order to correctly perform any spatial task.

By '*spatial performance*' we mean whatever kind of skilled observable interaction between a subject and the world that the community of spatial agents (those who share the same spatial competence) will judge efficient.

Principles described by a theory of competence are such that they are necessary conditions for any theory of performance. The object of our investigation will be to debate two alternative characterisations of spatial competence: our aim is not to provide models of how spatial knowledge can be actually used to produce spatial behaviour (which is a problem for theories of spatial performance) but to investigate under which conditions a system can be considered spatially competent.

SECTION I - CONTRASTING THE TWO APPROACHES

1. WHAT KIND OF STRUCTURES UNDERLIE SPATIAL COMPETENCE?

The first class of problems that we address in this section concern alternative requirements for the explanation of spatial competence that are postulated by traditional theories and sensorimotor theories. We will try to characterize what it means for both approaches *mastering* spatial capabilities.

1.1. Traditional requirements for explaining spatial competence

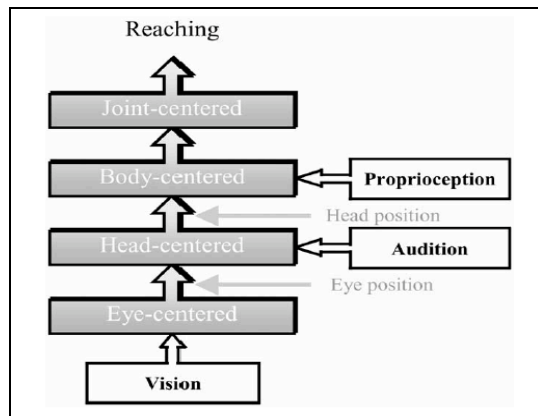
By way of introduction, let us consider two common examples of spatial capabilities, i.e. the ability to perform a VISUOMOTOR COORDINATION and a NAVIGATION task.

The traditional explanation of the way in which a subject succeeds at grasping a visually presented object runs as follows:

1. Relevant spatial properties of the object (size, position, distance) must be extracted from the visual field.
2. These properties must be projected, through appropriate transformational rules, onto the different spatial frames of reference that are involved in the act of grasping (possible internal representations that are required include eye-centred, head-centred, body-centred, arm-centred, shoulder-centred, elbow-centred, wrist-centred, hand-centred, finger-centred frames of reference).
3. On the basis of the initial conditions specified by 2., the most appropriate motor scheme is selected and executed.
4. The subject needs to update 1., 2. and 3. across time until the goal is achieved.

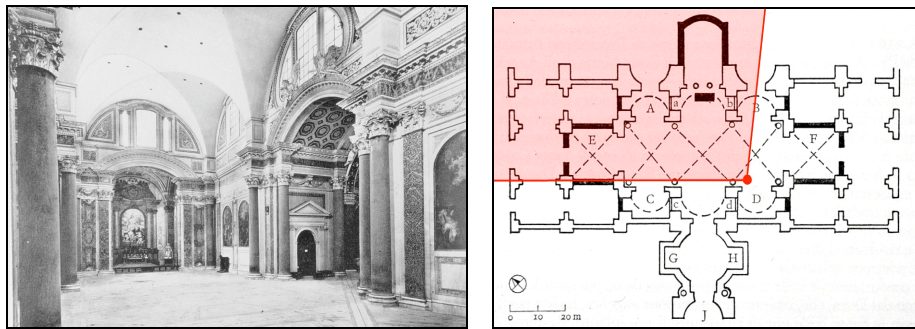
As Pouget et al. (2002) recently put it: “*In order to reach for an object currently in view, our brain must compute the set of joint angles of our shoulder, arm and hand (a.k.a. the joint coordinates) that bring the fingers to the location of the target. This involves combining the retinal coordinates of the object – provided by the visual system – with posture signals such as the position of the eyes in the orbit and the position of the head with respect to the trunk. As illustrated in Fig. 1, this process can be broken down into several intermediate transformations in which the position of the object is successively recoded into a series of intermediate frames of reference*”

Fig. 1. Coordinate transforms for multisensory motor transformations.



A second typical example of how spatial abilities are explained through the idea of mutually coordinated internal representations is NAVIGATION. How are subjects able to orient themselves through space and return to a previously visited place? It is usually assumed that body-centred (or “egocentric”) representations are insufficient to explain navigation tasks: what is required for correctly performing a navigation task is the ability to access a superordinate level of spatial representations, commonly called “cognitive maps” or “allocentric representations”, that are not anchored on a subject’s body but on external landmarks – Tolman (1948); O’Keefe & Nadel (1978); Klatzky (1998). *Egocentric representations have the disadvantage that in order to remain valid over the long term, they must be actively updated to reflect changes in the subject’s location and heading. Unless corrected by new sensory information, any errors in this updating process will be cumulative, so that egocentric representations of locations are unreliable for long-term storage. In contrast, processes demanding long-term memory of a location should make use of representations that relate location to each other and to landmarks in the environment, rather than to the subject. [...] A set of locations represented in an allocentric framework can be thought of as a ‘cognitive map’*” – Hurlley & Burgess (2002).

Fig. 2. Egocentric vs. allocentric spatial representations.



A general problem for mainstream models of spatial competence is then to understand how egocentric representations are tied to allocentric representations and, vice versa, how allocentric representations can be articulated into body-centred representations. “*How can map-like representations be abstracted from the egocentric information available to sensory systems? What form does the allocentric cognitive map take, and how could it support navigation?*” – Hurlley & Burgess, *cit.*

The above examples (the visuomotor coordination and the navigation tasks) allow us to focus on what we take as the basic requirements that a cognitive system has to satisfy, according to the traditional view, in order to be described as spatially competent. We maintain that these requirements consist in:

1. The possession of a set of correct internal representations, anchored to specific frames of reference, encoding relevant spatial properties of the external world. ‘Relevant properties’ are those properties that are minimally required for the accomplishment of the considered task¹.
2. The capacity to mutually co-ordinate such representations and keep their co-

¹ See 1.3.2 for a detailed discussion of what we mean by “relevant properties”.

ordination updated across time.

Generally speaking, this view can be seen as an example of classical representational explanations, according to which the possession of internal representations of external properties explains one's ability to interact with the external world. *"A representation occurs only as a part of a larger representational system that includes two related but distinct worlds: the represented world outside the information processing system (usually called the external world or environment), and the representing world within the information processing system (usually called the internal representation or simply the representation. What enables an internal world to represent an external world? One possibility is that the internal representation preserves information about the structure of the external world by virtue of having a similar structure. For this to happen, the structure of the two worlds must be the same to some extent"* – Palmer, *cit.*. According to the traditional view, then, a homomorphic mapping must exist between properties of external world and properties of the internal representation: *"if such a mapping exists, then the internal objects can function as a representation of certain aspects of the external environment"* – Palmer, *cit.*

The idea of a homomorphic relation between external space and internal representations entails the fact that representations are intrinsically *spatial*.

"Physical spatial structures are represented by [internal] spatial structures. Unlike a mere sign-relation, both the representing and the represented domains exhibit spatial structure; the representing domain represents physical spatial structures via transformed but similar [internal] spatial structures" – adapted from Hatfield (2003).

The homomorphism preserves the spatial structure from the represented to the representing domain: this similarity in structure is the tenet of classical representational explanations. Explaining spatial behaviour by relying on internalisation of external spatial properties amounts, then, to describing a *global* spatial competence in terms of a set of *local* spatial competences that are mutually co-ordinated: traditional approaches are committed to the idea that invoking spatial knowledge at subordinate levels is a *good explanation* of a subject's global spatial competence. We will see in the next paragraph that sensorimotor theories impugn the fact that postulating a homomorphism between internal and external structures provide good explanations of spatial competence and, hence, that internal representations of external space should be necessary for understanding spatial competence.

Incidentally, it should be emphasized that in characterising the traditional view, we are not endorsing any specific hypothesis about the way in which a subject succeeds in *acquiring* internal representations of external space: *how* these representations are built is not relevant for the present discussion. Accordingly, under the label 'traditional view', we class any approach committed to the assumption that spatial properties of the external world must be correctly represented by spatially competent cognitive system².

1.2. Alternative requirements for explaining spatial competence: the sensorimotor approach

A major alternative to the traditional framework, in the explanation of spatial competence, is what we call the sensorimotor approach.

The great philosopher and scientist Hermann von Helmholtz wrote: *"When we*

² See Appendix 1 for a more detailed discussion of different paradigms that comply with what we call "traditional view".

perceive before us the objects distributed in space, this perception is the acknowledgement of a lawlike connection between our movements and the therewith occurring sensations [...]. What we perceive directly is only this law” – Helmholtz (1878/1977).

The sensorimotor approach tries to take this idea seriously. Its fundamental claim is that the explanation of spatial competence does not need to rely on internal representations of external space: the sensorimotor paradigm holds, on the contrary, that the necessary requirement for a spatially competent system is the possession of the appropriate set of *sensorimotor invariants*.

What are sensorimotor invariants? And what does it mean that sensorimotor invariants are not internal representations of space? A *prima facie* definition of this notion can be found in O'Regan & Noë (2001), where they introduce the concept of “sensorimotor contingencies” as *“the rules governing the sensory changes produced by various motor actions”*. Sensorimotor contingencies are presented (although their characterization is merely sketched) as statistical rules associating specific motor patterns to specific properties of the sensory feedback: each time that a movement is triggered and a specific sensory pattern is produced, the rule that links the latter to the former is reinforced. The main goal of the authors in this paper is to show, then, that studying the way in which cognitive systems learn sensorimotor regularities can provide an alternative and promising account *“of visual consciousness, and of the differences in the perceived quality of sensory experience in the different sensory modalities”*.

More recently, Philipona et al. (2003) applied the sensorimotor paradigm to the specific case of spatial competence. They defend the idea that *“what biological organisms perceive as being the limits of their bodies, as well as the geometry and dimensionality of space outside them, are deducible, without any a priori knowledge, from the laws linking the brain’s inputs and outputs. The approach we are taking derives from the basic idea that the basis of sensory experience consists in extracting and exercising laws of sensorimotor dependencies”*. The aim of this paper was to see whether spatial capabilities might be acquired by a robot and explained without referring to internal representations.

Conceiving space as the result of the fixation of sensorimotor regularities, they argue, has several interesting and counterintuitive consequences. In SECTION II we will address specific foundational issues raised by a sensorimotor account of space. In this section we seek to put this hypothesis to a rigorous epistemological test and to probe whether sensorimotor approaches to spatial competence represent a tenable alternative to traditional explanatory strategies. The fundamental question that has to be answered here is: in which sense the sensorimotor approach differs from the traditional view? Most criticisms of O'Regan & Noë’s proposal have focused on the lack of a clear-cut distinction between their view and mainstream approaches. It has been argued that, lacking a precise theoretical characterisation, the alleged explanatory role of sensorimotor contingencies is either too vague or, worse, looks like a rehash of traditional representational approaches. *“It is not clear that this [sensorimotor] model truly lacks internal representations and memory. The repeated appeal to “knowledge” of sensorimotor contingencies seems little different from an internal memory or representation of an object. The only difference from a traditional object representation is that the “knowledge” in this case is of dynamic rather than static information*” – Scholl & Simons (2001).

We believe that much of these criticisms are grounded: the notion of sensorimotor learning needs to be theoretically constrained in order to compete (and be contrasted) with traditional approaches and lay claim to autonomous explanatory power. A theoretical foundation for any coherent sensorimotor explanation of spatial competence (and more generally of other kinds of perceptual and cognitive abilities) has still to be provided. The present work aims at contributing to such a theoretical foundation. Our

strategy consists in putting some constraints on the core concepts of sensorimotor theories and contrast them with the theoretical notions used in traditional models, i.e. internal representations. We will suggest that the notion of “sensorimotor contingency” be abandoned, because of its vagueness, in favour of that of “sensorimotor invariant”, a notion whose precise theoretical characterisation will be the goal of the next paragraphs: we proceed to formulate two main background hypotheses with which, we submit, any “sensorimotor” approach must comply.

1.2.1. The Genetic Hypothesis: basic conditions for sensorimotor learning

The first fundamental hypothesis concerns *what can be learnt* by perceptual systems given their internal constraints and initial resources. The genetic hypothesis claims that:

1. Perceptual systems come equipped with sensitivity to systematic correlations between motor sequences and sensory patterns³;
2. Sensitivity to invariant properties of co-occurring motor and sensory patterns can bootstrap a learning process;
3. This learning process results in the fixation of sensorimotor invariants.

The sensorimotor approach assumes thus that perceptual mechanisms can be modelled (to a large extent) as *extractors of sensorimotor invariants* (SMI), where by SMI we mean formal descriptions of regular properties of sensorimotor correlations. We state that such invariants describe lawlike relations, which we note $R(S,M)$, between properties of sensory patterns (S) and properties of motor patterns (M). In the most general case, the system has no *a priori* knowledge of this relations, and must *estimate* them.

Two aspects of our characterization of SMI must be pinned down.

First, the relation R that determines a SMI has a *statistical* character: SMI describe *statistically regular* correlations between properties of co-occurring sensory and motor patterns; we call them “invariants” simply to emphasize their tendency to stabilise over time.

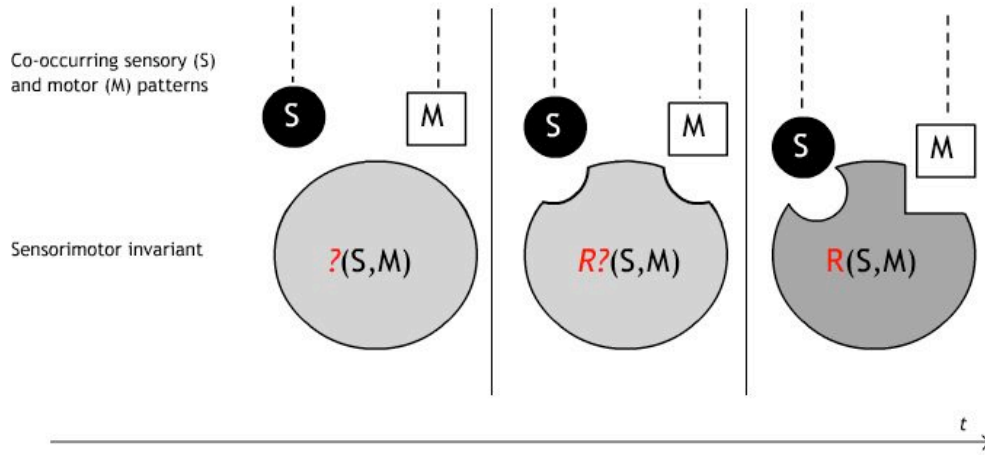
Second, our definition does not entail – as O’Regan & Noe (2001) suggest - that SMI should encode the resulting sensory changes *produced by motor actions*. Even if it can well be the case that active control of motion is playing a major role in fixating SMI, in our definition motor patterns have no particular priority over sensory patterns. The system is not necessarily discovering how sensory information changes through self-motion. Our Genetic Hypothesis only assumes that perceptual systems must be able to extract and fixate regularities from the *co-occurrence* of (S,M) patterns. In this sense, our formulation explicitly intends to avoid any *oriented* causal interpretation of the $R(S,M)$ relation according to which, for instance, the system is supposed to learn from testing what modifications are produced in a sensory feedback by specific kinds of motor programs⁴.

³ What are the specific requirements for being sensitive to such correlations will be extensively discussed later: *how* a perceptual system actually manages to extract, discriminate and identify SMI is an extremely delicate issue, involving a number of theoretical problems that we will address in section II.

⁴ The indispensable role of active motion control for the extraction of SMI and consequently the idea that SMI can be characterized as descriptors of sensory changes *produced* by self-motion is an interesting issue that is not implied by our characterization of SMI. Some aspects of this divergence (in particular, a distinction between two possible interpretations of sensorimotoricity) will be discussed in more details in Appendix 2.

The picture below can help visualize how systematic exposure to co-occurrent sensory and motor patterns over ontogenetic development can result in the fixation of a SMI. We will say that sensorimotor learning progressively *shapes* a system's knowledge by fixating the appropriate $R(S,M)$ laws and, conversely, that a system's sensorimotor knowledge at a certain point of its development corresponds to its *shape*.

Fig. 3. Systematic exposure to sensorimotor couplings results in the fixation of a SMI



We can illustrate the kind of sensorimotor learning defined by the Genetic Hypothesis by imagining a monocular system with no prior knowledge about its own structure or the environment in which it is embedded. The Genetic Hypothesis assumes that, by executing different classes of motor schemes, this system will be exposed to regular associations between its movements and co-occurring sensory patterns. For example, each time the system performs a rotation of the eye (A.), a specific kind of optical flow will be generated; in contrast, each time the systems performs a lateral translation without changing its point of fixation (B.), a different kind of optical flow will co-occur.

Fig. 4. Distinct sensorimotor correlations yield distinct SMI.

- A. Optical flow produced by eye rotation
Eye rotation \Leftrightarrow Resulting optical flow
- B. Optical flow produced by eye translation
Eye translation \Leftrightarrow Resulting optical flow

[Credits: M. Wexler, LPPA]

Through exposure to these two distinct correlations between patterns of visual stimulation and self-motion, according to the Genetic Hypothesis the system will learn to *discriminate* between them, i.e. acquire two different SMI describing distinct (S,M) couplings.

1.2.2. The Functional Hypothesis: the role of sensorimotor invariants

The second hypothesis concerns the *use* of sensorimotor invariants for performing specific spatial tasks. The Genetic Hypothesis introduced above claims that, through the experience of systematic (S,M) correlations, a cognitive system stores a library of sensorimotor invariants. The Functional Hypothesis, in turn, assumes that:

1. The set of available SMI describe the spatial knowledge a system possesses (at a specific stage of its development) as a result of its previous sensorimotor learning.
2. A system's spatial knowledge is embodied by operators (we will call them *sensorimotor comparators*) delimiting the set of possible R (S, M) compatible with the system's knowledge (given its previous learning). Comparators work as *filters*: they monitor ongoing motor schemes and co-occurring sensory patterns, evaluating their match to acquired SMI⁵.

The Functional Hypothesis refers to two different levels of description of a skilled system. Whereas (1.) concerns the level of formal description of a system's knowledge, (2.) concerns, in turn, the level of instantiation of SMI. To rephrase (2.), we can say that the set of stored SMI *shapes the internal structure* of sensorimotor comparators, taken as the fundamental unit of sensorimotor systems: in order for a system to be able to store and make use of SMI, we need to postulate that such systems possess devices working as sensorimotor comparators.

In the example above, the sensorimotor invariant describing the dynamic properties of the optical flow during eye rotation can be instantiated as a sensorimotor filter delimiting the set of possible ways in which the actual optical flow and the co-occurring rotation are associated in the specific environmental context in which the system is embedded.

1.2.3. Mastery of spatial competence as compliance with sensorimotor invariants

A skilled sensorimotor system, according to O'Regan & Noë (2001), is a system that "*masters the use of sensorimotor invariants*". This notion of '*mastery*' has been applied by the authors to develop original experimental protocols and support the explanation of interesting perceptual phenomena, such as cases of *sensory substitution* or *perceptual learning*⁶. Nevertheless, one can object that (as in the case of some other key concepts of their proposal) lacking a precise theoretical characterisation, the explanatory role of this '*mastery*' remains too vague. The notion of mastery needs to be constrained in order to play an explanatory role. The two background hypotheses we have formulated above enable us to fix a technical definition of '*mastery*', which can be fruitfully put to work in order to contrast the sensorimotor paradigm and the traditional one. We can account for the idea of sensorimotor mastery (and for its counterpart, self-correction), by appealing to the interplay of two operators, namely *comparators* (as previously described) and *correctors*.

1.2.3.1. Mastery of spatial competence as possession of reliable filters

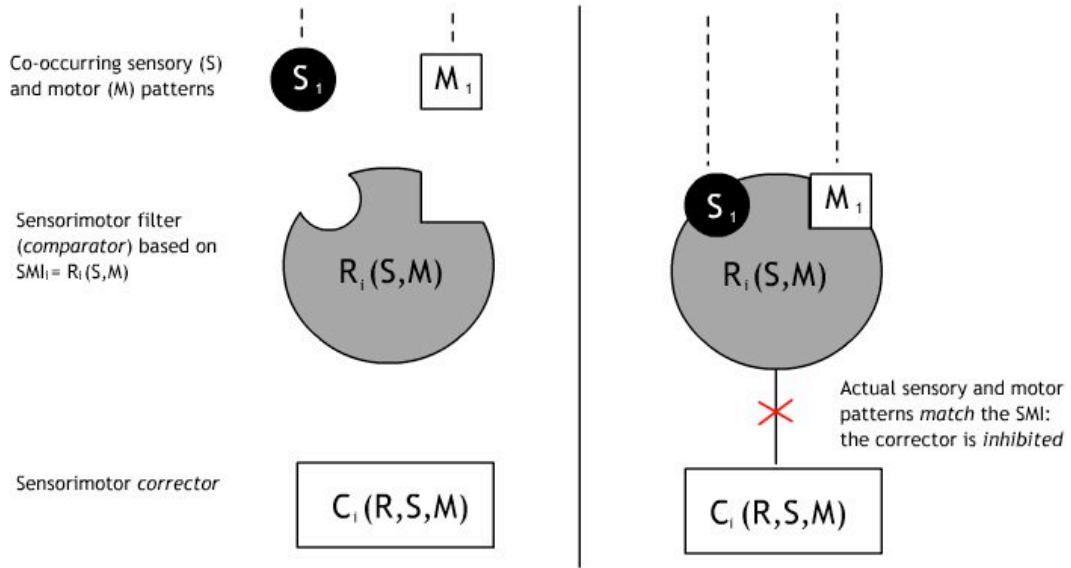
We will say that, for a given task, a system *masters* a class of SMI as soon as it uses a

⁵ Again, our characterization of sensorimotor operators as *filters* delimiting the class of possible (S,M) is different (although partially compatible) with current literature postulating sensorimotor operators as *predictive devices*. See Appendix 2.

⁶ See for instance Bach-y-Rita (1972) and (1984).

comparator as a sufficiently reliable filter of the relation between actual sensory and motor states in real-world situations. Using a comparator as a “reliable” filter means that the comparator is not able to evaluate any significant difference between, on the one hand, past (S,M) couplings the filter has become attuned to and, on the other hand, actual, ongoing (S,M) relation. To illustrate this idea, let us consider the following scheme:

Fig. 5. *Reliable filters match actual sensorimotor couplings*



As a result of previous learning, a filter is shaped by a specific class of SMI. Thanks to its shape, a filter based on a class of SMI restricts the set of possible relations R between properties of sensory patterns S and motor patterns M . Such properties will be, for instance, motion vectors of single points in the optical flow and in motor programmes. As soon as the difference between possible and actual properties of (S,M) couplings is so small that the system is not able to discriminate between them, we will say that the comparator is a reliable filter for actual co-occurring properties. This *indiscriminability* between relations allowed by the filter and the actual (S,M) relation is what we call a system’s ‘*mastery*’ of a class of SMI.

In this sense, mastery of SMI is what makes the difference between a system which is undergoing a learning process and a *skilled* system, i.e. a system that does not need further learning in order to possess the sensorimotor knowledge required to correctly perform spatial tasks.

It is important to remark, though, that *being skilled* for a system does not imply that the system is not susceptible of undergoing any further learning process. Further learning is required as soon as the comparator is no more working as a reliable filter, i.e. as soon as it can make a difference (discriminate) between the set of possible (S,M) correlations (as fixed by R) and the actual (S,M). In other words, further sensorimotor learning is needed every time sensorimotor invariants fail to account for actual (S,M) couplings: typically, as soon as the system moves to a new context in which the regularities that worked in the former one do not hold any more, or when the system changes its physical structure.

To illustrate this point, let us imagine a creature that has never been exposed to transparent surfaces. We will say that this creature is spatially competent (i.e. a *spatially skilled system*) as far as it possesses sufficiently reliable sensorimotor filters tailored to a world made up of opaque surfaces. As soon as this creature moves to a different environment (in which surfaces can either be opaque or transparent), its available spatial knowledge (*the set of stored SMI*) will not fit the new domain: for instance, it will not be able to perform the task of avoiding an obstacle, in the case of a transparent surface. In this case, we will say that since the library of former invariants is not valid any more, a new learning process has to be undertaken in order to make the system's filtering operators reliable again.

This interplay between the notion of mastery and that of self-correction (i.e., further sensorimotor learning) can help solve, we submit, the seeming puzzle of the incompatibility between mastering a class of SMI while still being susceptible of learning: a system can be said to have mastery of SMI only in respect to the specific environmental context it has been exposed to. If the environment (including the physical structure of the agent) does not change and the system has properly attuned its sensorimotor invariants to this environment, no further sensorimotor learning will be needed. We assume that this is the case for adults that we describe as spatially competent.

Given this characterization of spatial competence as mastery of sensorimotor invariants, what does it mean for a system to be *skilled* to perform a particular task? A sensorimotor explanation for the visuomotor coordination task that we mentioned in §1.1 will run as follows: grasping an object requires a subject to perform a movement that results in a motor sequence and a pattern of sensory states *compatible* with the known SMI. Explaining this kind of spatial task requires, then, a description of the sensorimotor invariants that the subject must have internalised and the way in which she puts them to work for accomplishing the task.

1.2.3.2. *Further learning for new environmental contexts*

We have suggested, in the previous paragraph, that a sensorimotor comparator is not a reliable filter whenever properties of the actual sensorimotor correlation do not match the set of possible $R(S,M)$ representing the system's knowledge. When this happens, it is necessary to account for the calibration mechanisms thanks to which a system adjusts or acquires new SMI fitting the new context (as, for example, a world with transparent surfaces). An insufficient knowledge will result in a sensible difference between the actual (S,M) and the estimated $R(S,M)$ law: we assume that the system uses this difference to improve its discriminatory abilities, thus adapting its sensorimotor knowledge. In what follows, we will try to outline the nature of this self-correction process. Let us call *sensorimotor discrepancy* (SMD) any discriminable difference a system can detect between the estimated law $R(S,M)$ embodied by a sensorimotor filter and the actual (S,M) relation. Schematically speaking, any estimation made by a filter can fall in either of the two following scenarios:

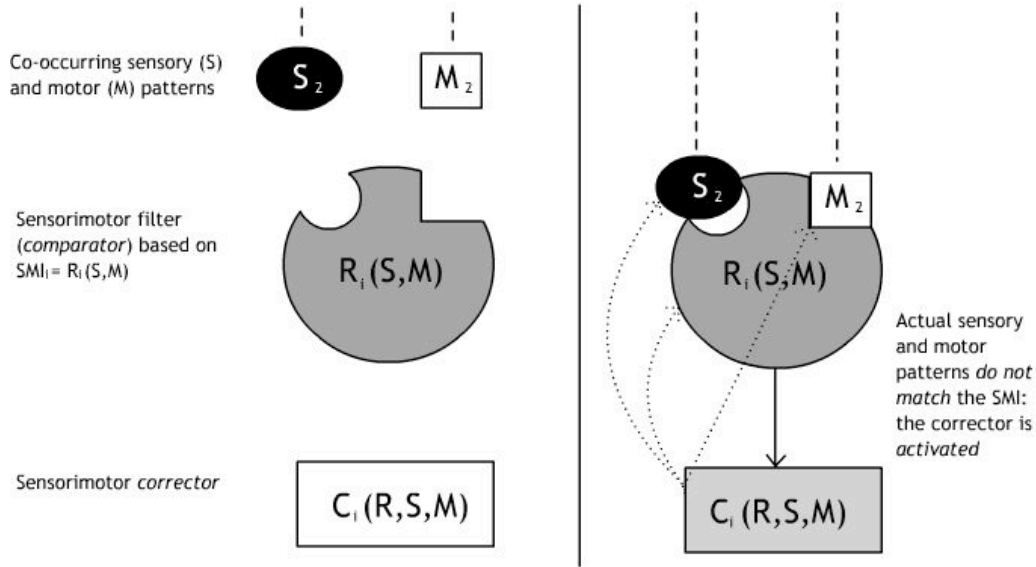
- A. The system is not able to detect a SMD, i.e. the difference between estimated and actual (S,M) relations is below the threshold of detectability of the system: the actual (S,M) coupling matches (*is compatible with*) the operator's estimation $R(S,M)$. This means that the filters are reliable: the system masters the appropriate SMI⁷.

⁷ It should be noted that given our definition of mastery, online correction mechanisms (like real-time correction of an arm's trajectory in reaching for an object) belong to a skilled system's competence: they qualify as part of a system's mastery and not as cases of self-correction

- B. The system detects a SMD: a process of calibration of the comparator is triggered.

The operator in charge of this self-correction process is what we call a *sensorimotor corrector*. A corrector detects any sensorimotor correlation that does not match the filter's SMI.

Fig. 6. Incompatible sensorimotor correlations trigger a self-correction process



The comparator and the corrector work, then, in a complementary fashion: *either* the (S,M) sensorimotor correlation is possible (i.e., compatible with a system's available knowledge), in which case the comparator inhibits the activation of the corrector (see Fig. 5), *or* it is incompatible, in which case the corrector is activated by the comparator (Fig. 6). The outcome of a calibration process is the improvement of a system's spatial knowledge, i.e. the extension of the set of possible SMI allowed by sensorimotor filters. We can rephrase this idea by saying that self-correction mechanisms reshape a system's discriminatory capabilities.

By way of illustration, consider the following example of self-correction. Imagine a child who has learnt how to ride a bicycle. In sensorimotor terms, we will say that she has stored a specific set of SMI that give her mastery of this ability. She knows, for instance, that, pushing a pedal by applying a certain force while riding on a flat road will be correlated with a sensory pattern corresponding to a forward displacement. Now, suppose the child comes to ride on a slightly downhill road and suppose she has not learnt so far to discriminate between this and the former context. As soon as she performs the motor scheme that she has learned so far, the co-occurring sensory patterns will be sensibly different from those estimated through her previous sensorimotor experience: they will yield, for instance, an unexpected acceleration of the optical flow causing the child to fall off the bicycle. We will say that the child's comparators have detected a sensorimotor

procedures. They do not threaten a system's available sensorimotor knowledge: on the contrary, they represent a high degree of mastery of SMI, namely the ability to cope with complex interferences.

discrepancy between the actual (S,M) coupling and the filter’s estimation. This, in turn, will trigger a learning process increasing the child’s knowledge that will allow her to cope with both flat and sloped surfaces.

Providing a full-fledged account of self-correction processes in real-world agents demands a complex analysis of the mutual interaction between several possible factors. This goes beyond the aim of the present work. However, it is worth mentioning some points in order to clarify the problem. The child’s falling suggests that she did not have sufficiently accurate spatial knowledge, i.e. she did not master the appropriate kind of sensorimotor invariants. A calibration is needed. *Why* was the child’s available knowledge unreliable? *How* do her sensorimotor filters have to be corrected? The child will need to learn what has to be calibrated in order to acquire reliable sensorimotor comparators. This adjustment can be rather difficult to achieve. It might be the case, for instance, that it is the sensory component (S) of the invariant that must be refined (so as to distinguish, for example, the difference produced on the optical flow by a flat and by a downhill road) and, consequently, store two different SMI involving different sensory patterns coupled to the same motor pattern. But, obviously, the calibration process can also affect the motor component (M) or even the R(S,M) relation *itself* associating a sensory and a motor pattern. What we mean to suggest is that a sensorimotor discrepancy is per se *neutral* towards the calibration process it generates: the specific kind of correction required by the system is underdetermined by a SMD. It is, though, an interesting empirical question to investigate, if the sensorimotor hypothesis makes sense, how real-world systems manage to solve this problem⁸.



Let us briefly summarize the conclusions of this first paragraph. The requirements that traditional approaches invoke for explaining spatial competence (the possession of internal representations and the ability to mutually coordinate them) are significantly different from those specified by sensorimotor theories. Whereas the former assume that spatial abilities are explained by appealing to a system’s internalization of structures that bear a homomorphic relation to spatial properties of the external world, the latter deplore that such internal representations are not required. What is needed, in contrast, is a number of internal structures encoding lawlike connections between self-motion and co-occurring sensations. We have tried to characterize how both the traditional and the sensorimotor approach account for the explanation of spatial tasks by appealing to two distinct ideas of mastery. It is, though, arguable that internal spatial representations and sensorimotor invariants should be plainly incompatible, as the defendants of sensorimotor theories have claimed. In the next paragraph we will address the delicate issue of understanding whether and in which sense sensorimotor invariants can be considered as a kind of representational device and what distinguishes them, still, from traditional internal representations.

⁸ See Section II for a number of working questions related to the problem of sensorimotor calibration.

1.3. How to distinguish sensorimotor invariants from internal representations

1.3.1. The compatibility claim: SMI as representational devices

Mastery of spatial capabilities is universally supposed to involve representations. Are sensorimotor approaches really doing without internal representations? “*According to many representational theories, the perceiver [...] sees external objects, but her ability to do so is underlain by neural features that function as representations of the scene. Thus, we need to be clear about what something must do to “function as a representation” in the relevant context. Are representations in that sense compatible, incompatible, or perhaps even entailed or strongly implied by the sensorimotor model of vision?*” – Van Gulick (2001).

Many other authors – see Scholl & Simons (2001), Tatler (2001) – have similarly claimed that, in spite of the original purpose of O’Regan & Noë, *cit.*, whose aim was to sell the sensorimotor paradigm as an alternative to mainstream representational approaches, sensorimotor regularities can (and cannot but) be described as a specific kind of representational devices. We will call this objection the “Compatibility Claim” (CC).

According to the CC, whatever system is able to:

1. extract invariant properties from sensory stimulation and
2. make such information available for action control, perceptual judgment or reasoning

can be described as a representational system.

There seem to be two possible interpretations of the CC: a weak and a strong version.

If the CC is simply taken to mean that representations are internal structures whose properties bear a homomorphic relationship to properties of what is represented (weak interpretation, see Palmer’s first quotation, p. 3), then a system storing SMI can be legitimately considered as an instance of a representational system. In fact, since we assumed that SMI fix some regular properties of sensorimotor correlations, SMI can be described as *representations* of these properties. Therefore, under a weak version of the CC, we argue that the sensorimotor paradigm is clearly compatible with a general representational framework.

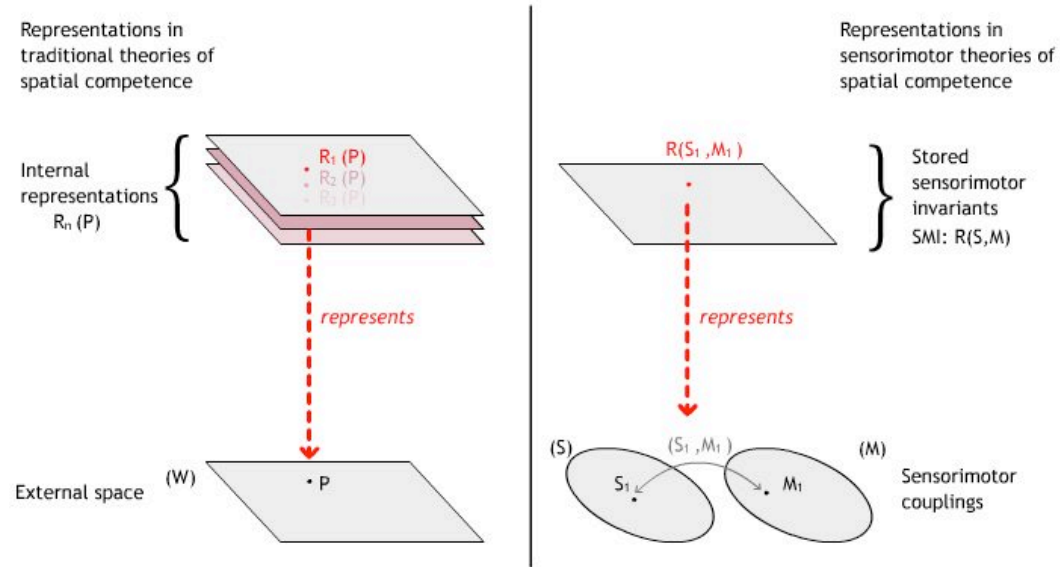
On the other hand, we claim that the sensorimotor view is incompatible with the traditional view if the CC is taken as implying a further engagement about *what must be represented* in a SMI (strong interpretation, see Palmer’s second quotation, p. 3-4): the incompatibility concerns in other words the representational content of a SMI. Let us see why.

According to traditional approaches, what is represented by an internal representation is generically some *external* spatial property P. Spatial properties of the external world project onto internal frames of reference (see 1.1): the representational link a system has to establish, in order to become spatially competent, is the mapping between the internally projected (*representing*) property P’ and the external (*represented*) property P it refers to. According to the sensorimotor paradigm (see 1.2.1), instead, the relevant representational link a system has to learn is the one between classes of co-occurring motor and sensory pattern (S,M) – the *represented* structures – on the one hand, and internally stored R(S,M) laws – the *representing* structures. SMI can be said to represent the coupling between classes of motor schemes and properties of sensory patterns or, conversely, that the

representational content of SMI is made of sensorimotor correlations.

SMI are representations of properties of (S,M) couplings and *not* of properties of the external world. Hence, if the strong interpretation of the CC implies that representations must be correlations between internal and external properties, SMI *cannot* be considered as representational devices⁹.

Fig. 7. Different representational links in traditional and sensorimotor theories.



1.3.2. What represents a SMI?

The distinction between the two distinct representational links has crucial consequences on the selection of *relevant properties* that perceptual systems are supposed to use in acquiring spatial competence. Relevant properties, according to traditional approaches, are paradigmatically restricted to spatial features of proximal sensory stimulation that are geometrical projections of distal spatial properties. Typically, such properties are those that allow the solution of an *inverse problem*¹⁰. In contrast, relevant properties according to the sensorimotor paradigm are dynamic properties of the sensory stimulation that are susceptible of co-occurring with motor schemes.

Let us illustrate this difference in the selection of relevant properties through an example.

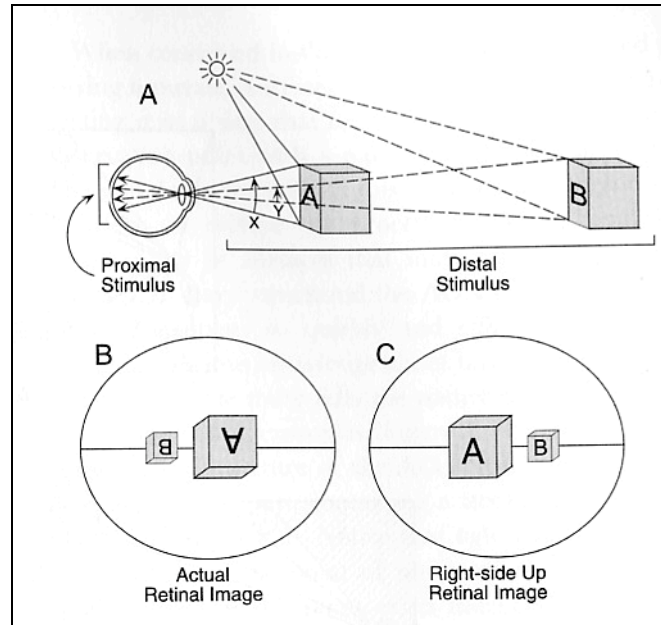
⁹ It is worth mentioning a curious consequence of our criterion of demarcation between traditional and sensorimotor approaches. In current literature about spatial cognition, the adjective “sensorimotor” has been widely applied to a number of different theories based on the idea that sensorimotor loops are involved in the extraction of the spatial structure of the environment – see for instance Wolpert (1995); Soechting & Flanders (1989) and (1992). Paradoxically, most of such theories, following our criterion, belong nevertheless to the “traditional view” and not to what we call “sensorimotor paradigm”, in that they still assume (either explicitly or implicitly) that spatial competent agents must have internalized properties of the external space.

¹⁰ A basic formulation of the inverse problem for visual perception can be found in Palmer, *cit.*, p.23.

Be $D_S(P)$ the *relative distance of P from a subject S* and $D_S(Q)$ the *relative distance of Q from a subject S*, where P and Q are objects of equal size. What properties are used by a perceptual system for estimating if $D_S(Q) < D_S(P)$?

Traditional approaches, on the one hand, take difference in size of the retinal projection of the objects as an example of relevant variable that perceptual systems must extract. This assumption derives from basic considerations of projective geometry: the size of the projection of two segments on the retina is inversely proportional to their relative distance. Since the relative distance is the (external) property that has to be estimated, then a difference in size of retinal projections is taken as a useful variable for solving this specific spatial problem. Relevant properties are those that a geometrical mapping associates with external spatial properties.

Fig. 8. The relation between distance and relative retinal size



Following a sensorimotor approach, on the other hand, relevant properties for estimating the distance of objects are properties of dynamic sensory patterns associated with a specific class of motor schemes. For example, a perceiver's lateral translations will be regularly associated with optical flows in which two different angular velocities are detectable: the closer an object is to the perceiver, the bigger the angular velocity of its retinal projection *when the perceiver performs lateral translations*. Relevant properties are such that, *insofar as they are coupled with specific classes of motor schemes*, they allow the system to discriminate between two different (S,M) correlations, namely to fix different SMI. What enables a perceiver to make a distinction between a close and a distant object is the ability to discriminate the sensory pattern they produce when the subject performs specific classes of motor schemes.

1.3.3. Sensorimotor vs. ecological invariants

One might be tempted to say that sensorimotor invariants, as they have been characterized in the present work, are equivalent to the well-known concept of ecological invariants – Gibson (1979). As Scholl & Simons, *cit.*, remark: “*The content of the sensorimotor theory is highly reminiscent of Gibson’s work on direct perception [...]. The sensorimotor framework would be greatly clarified by considering in detail which parts of the theory are substantive departures from earlier Gibsonian arguments*” - Ecological invariants – it might be argued – can qualify as a particular case of sensorimotor invariants, in that they seem, *prima facie*, to perfectly fit our formal description of a SMI (we will call this the “ecological objection”). Ecological invariants are, in fact, dynamic properties of the sensory stimulation that are elicited by specific classes of motor schemes.

The example of the previous paragraph, for instance, could be reformulated in ecological terms as follows: when estimating the relative distance of two objects, the observer is extracting the value of the parallax produced through lateral translation on the retina by the two objects. An estimation of the relative distance is accordingly supposed to be made by comparing the two values of parallax. Parallax can then be considered as an ecological invariant in virtue of the fact that it is a geometrical relation between points on the retina that remains constant across the optical flow produced by movement. The ecological literature provides many examples of similar invariants, including among others: the *vanishing point* produced by forward and backward translation – Gibson, *cit.*, or the *cross-ratio* between co-linear points on the retinal projection of a rigidly moving object – Cutting (1986).

The claim that ecological invariants and SMI are different description of the same kind of structures relies, we hold, on an incomplete understanding of the difference between the two notions: we argue that an ecological invariant can fit our characterization of a SMI only under two specific conditions.

These conditions are respectively:

- A. The assumption that alleged sensitivity to ecological invariants is compatible with the internal constraints of the system.
- B. The hypothesis that the system is able to couple ecological invariants with specific classes of motor schemes.

As for (A.), according to the laws of ecological optics, there is a nomological relation between specific kinds of geometrical transformation in the world (say, the motion of a rigid object) and the corresponding projection on the retinal image (the invariant cross-ratio on the retinal projection of the moving object). It is because of this nomological relation that invariants can be said to carry reliable information¹¹ about spatial properties of the world. Now, the ecological hypothesis as a theory of perception affirms that perceptual systems actually *exploit* these informative relations in order to acquire spatial knowledge of the external world. Yet, from the fact that there is a reliable informative relation between these invariants and spatial properties, it does not follow that perceptual systems should *actually* exploit this correlation, but only that they might *possibly* use it.

Turning to (B.), we may say that constraining the use of invariants in a perceptual system

¹¹ It is not relevant for the present discussion the debate about the reliability of ecological invariants for reconstructing world properties. Appendix 1 will briefly review the main points of this debate.

means formulating explicit hypotheses about the internal setup of a system that allows it to actually *make use* of these invariants. Traditionally, the ecological approach has privileged the study of informational reliability of invariants, without being especially concerned with hypotheses on internal constraints of perceptual systems¹². In contrast, the sensorimotor paradigm assumes that perceptual systems must be equipped with sensitivity to regular co-occurrence of invariant properties of sensory stimulation associated to particular motor schemes (internal constraints of a system are those specified by the Genetic and Functional Hypotheses, see §1.2.1-1.2.2). Ecological invariants can then enter in the formal description of SMI under the condition that the system be able to bind them to specific motor patterns.

What is required for redescribing ecological invariants as SMI is an account of the way in which perceptual systems actually use them and a description of the specific motor schemes in which they occur. Lacking these elements, ecological invariants are just abstract descriptions of invariant properties in dynamic sensory patterns. In conclusion, our reply to the “ecological objection” is the following: ecological invariants *can* be equivalent to SMI only under the restrictions imposed by the above conditions.



Having distinguished in which sense traditional and sensorimotor models can be said “representational” and having circumscribed the notion of SMI from neighbour concepts like that of ecological invariant, we can now turn to a number of side issues related to the distinction between the two main approaches.

2. THE ARCHITECTURAL COMPLEXITY OF SPATIALLY COMPETENT SYSTEMS

A major issue related to the distinction between the two approaches concerns the complexity of the architecture of a *spatially competent* system.

We argue that, under the hypotheses formulated above, the sensorimotor approach can provide a unitary framework to explain the requirements of spatial capacities without postulating a complex architecture. More precisely, the sensorimotor paradigm predicts that a system, under the hypothesis that it be endowed with the capacity to fix regular properties of its sensorimotor experience, will be able to perform a number of tasks that are traditionally explained by appealing to a set of dedicated mechanisms and distinct representational levels.

SMI-based operators can be seen as general-purpose devices that do not require the modeller to postulate heavy constraints on the architecture of the system, related to each of the spatial routines that call for an explanation. Any kind of spatial routines can be accounted for by appealing to the very same kind of operators: we call this the “architectural minimalism” of the sensorimotor approach. Traditional approaches, on the other hand, postulate a much richer architecture of dedicated operators.

Let us consider, for example, the requirements that a system has to meet for extracting the dimensionality of its environment. Traditional approaches assume that extracting the

¹² We maintain that models lacking an explicit account of internal constraints of perceptual systems are better described as theories of abstract ecological optics than as proper theories of perception.

third dimension from a bidimensional retinal array is a problem requiring an explanation. As Palmer, *cit.*, puts it: “ *The early stages of visual perception can be viewed as trying to solve what is often called the inverse problem: how to get from optical images back to knowledge of the object that gave rise to them. From this perspective, the most obvious solution is for vision to try to invert the process of image formation by undoing the optical transformation that happens during image formation*”.

Since *three-dimensionality is a property of the external space* that is not preserved in the sensory projection, it is necessary to postulate dedicated computational levels for *reconstructing* (i.e., correctly representing) the original dimensionality of the scene¹³.

What is meant by *dedicated levels*? We assume that dedicated levels can either be individuated by

- the specific kind of algorithms they implement;
- the specific kind of properties they are sensitive to (the specific domain they refer to);
- the specific representational format they use;

Postulating dedicated operators increases architectural complexity: spatial tasks of alleged different complexity will demand distinct computational levels of processing. For instance, comparing bidimensional properties, extracting depth or estimating rigidity are tasks that are supposed to be accomplished by operators that process properties of the stimuli at different computational levels.

In a sensorimotor perspective, on the contrary, it is unnecessary to postulate the existence of a dedicated level for, say, the extraction of three-dimensional properties of the visual scene. Perceiving objects as three-dimensional rather than bi-dimensional amounts to being able to discriminate between two different classes of sensorimotor invariants: fine-grained explanations of perceptual and motor capabilities on 3D objects might benefit from the architectural parsimony of the sensorimotor framework.

3. WHAT HAS TO BE EXPLAINED BY A THEORY OF SPATIAL COMPETENCE?

We have contrasted so far two different approaches to the explanation of spatial competence and stressed what distinguishes them and what makes them mutually compatible. We have shown to what extent the sensorimotor and the traditional framework can provide competing scientific programmes for the explanation of the same kind of spatial capabilities. Yet, a more fundamental question concerning the epistemic status of these paradigms has still to be addressed. *What does it mean to explain spatial competence?* We will try to show that, even if both approaches are meant to provide an analytic description of the internal knowledge a system must possess in order to perform spatial tasks, the explanatory role of this internal knowledge is critically different.

According to the traditional view, explaining spatial competence has to be understood as *formulating constraints* on the competence of a well-formed system, given the existence of internal spatial representations: the object of the explanation is to show how the spatial capabilities of a system can be derived, by *appealing* to internal representations it is supposed to possess. Internal representations of external space are considered as essential prerequisites for a system to behave spatially. To put it differently, the most economic and

¹³ Marr (1982) is probably the most famous example of architectural levels of visual spatial representation.

accurate way to account for a system's spatial capabilities is to assume its possession of a class of mutually linked internal representations of external space. How a system comes to acquire these internal representations is generally considered by the traditional approach a problem for developmental theories, not for a theory of spatial competence. Understanding spatial competence, according to this view, means describing how the use of these representations in a skilled system can account for observable spatial behaviour. We will say that in this kind of approach the system can be seen as *informationally closed* i.e. as already provided with knowledge required to account for its spatial competence. The availability of such knowledge is justified, in general, by appealing to evolutionary or ontogenetic considerations¹⁴.

Following a different strategy, the sensorimotor approach tries to explain spatial competence from a *genetic* point of view: the object of the explanation are the minimal requirements thanks to which a system, exposed to an environment with a specific structure, *becomes* spatially competent, and not how its competence can be deduced from allegedly available internal spatial knowledge. What distinguishes the sensorimotor approach in this respect is the fact the object of the explanation is at the same time the acquisition and the use of spatial competence. The system is not supposed to possess internal representations that allow it to correctly interact with the world: on the contrary, knowledge of the structure of space must be extracted from the exercise of sensorimotoricity. We can say that in this kind of approach, the perceptual system is considered *informationally open*, i.e. knowledge allowing it to perform spatial tasks is defined (and studied) as the emerging result of the interaction between the system and the specific environment it is embedded in. The characterisation of this interaction and the fixation of rules that allow efficient interactions become thus the main object of the explanation.

To summarize the above distinctions, we can say that the sensorimotor and the traditional view bear a different epistemological stance towards their *explanandum*, i.e. spatial competence, and their *explanatory strategy*.

- They diverge on the *focus of the explanation*, insofar as they give a different weight to the problem of the acquisition of spatial competence, which is considered as integral part of the explanandum in the sensorimotor approach and is squeezed off the scope of the explanation in traditional theories.
- They diverge as to the *explanatory strategy* to be adopted: if, on the one hand, for traditional theories explaining spatial competence means describing the possession and mutual coordination of internal representations of external space, i.e., how to deploy representational devices that encode the spatial structure of the external world, on the other hand for sensorimotor theories explaining spatial competence means describing the use of a set of operators that determine the properties of space for a specific creature in a specific environmental setting. Explaining spatial competence does not require internal representations of external space whose structure must be encoded; rather spatial competence is a matter of parameter fixation, in which the relevant parameters are correlations of (S,M) couplings.



¹⁴ See O'Keefe & Nadel (1978); Kubovy (2001); Shepard (2001); Todorovic (2001).

Thinking of space as the result of a fixation of sensorimotor parameters provides a number of interesting working questions that we will discuss in more detail in the next section of this work. In particular, we will push forward the analogy between learning the syntax of a language and learning the spatial structure of the environment. We will consider some of the counterintuitive consequences of this analogy: especially, the possibility of conceiving different spatial abilities for agents having fixed different classes of SMI (because of significant differences in world regularities or in the setup of sensory devices). What remains to be addressed, before moving to these more speculative issues, is an assessment of the scope and limits of sensorimotor explanations.

4. SCOPE AND LIMITS OF SENSORIMOTOR EXPLANATIONS

Two kinds of general criticisms have been addressed against the theoretical framework proposed by O'Regan & Noë, *cit.* On the one hand, it is arguable that a sensorimotor approach can actually do without internal representations and explain problems that are traditionally accounted for by appealing to representational devices. On the other hand, it is legitimate to wonder whether sensorimotor explanations, as characterized by the Genetic and Functional Hypotheses, can actually account for the whole range of complex phenomena we refer to when talking of "spatial competence. In particular, it has to be shown how complex kinds of spatial tasks can be modelled in a purely sensorimotor fashion.

The first kind of criticism has often made reference to the case of perceptual recognition. *"How might we explain an infant's putative ability to hold in mind both the existence and location of an occluded object, or more impressively, to recognize when that object should obstruct the movement of another occluded object, within the formal framework of sensorimotor contingencies?"* – Schlesinger (2001).

"No theory of human perception (as a whole) can avoid the fact that perception includes object recognition, and that recognition involves categorization. Categories are pivot points that determine choices between potentially large sets of diverse actions, and they affect action at every level from eye movements on upward. The broad shift of behaviour that comes from realizing, for example, that a store is closed or that an animal is stuffed demonstrates that the individual knows what these facts mean. And the stored information about what something means can be considered to be a representation, whether or not the function of that information is to generate behaviour. Even if a theory of object recognition is devoted to explaining human action, claiming that the theory has no representations may be little more than an aesthetic decision regarding labels" – Pani (2001).

Much of these examples have been raised, in particular, to counter the claim that we do not need internal representations of the world because *"the external world works as its own external memory"* – O'Regan (1992); O'Regan & Noë, *cit.* As we have shown in §1.3.1, this conceptual problem vanishes, since we explicitly agree that sensorimotor systems make use of internal representations. The only difference between the traditional and the sensorimotor paradigm, in the specific case of spatial competence, concerns the nature of the *representational content* and not the idea of representation in itself.

The second kind of criticisms is more interesting, and deserves a more articulated analysis. It concerns the scope of sensorimotor explanations and, in particular, the possibility to account for any kind of spatial performance by appealing to sensorimotor-based devices. Such criticisms can be grouped in two distinct kinds of argument:

1. *Action-based-capabilities argument*: the sensorimotor hypothesis can maybe account for a specific set of spatial capabilities, namely those in which the contribution of

action is necessary. Nevertheless, there are many spatial capabilities that can hardly be characterized in virtue of any sensorimotor relation in a meaningful way. “*While it is true that we often use our visual system to determine our actions, we also use it to find out what is in the world simply because we want to know. As George Miller once put it, we are basically informavores: we seek to know even if we have no possibility of acting towards what we see – as we do when we watch television or visit an art gallery or read a book. Most things we see are things we cannot act upon directly, such as the words in the target article*” – Pylyshyn (2001). As a corollary of this argument, it is arguable that for a number of high-level spatial capabilities a purely sensorimotor account might be provided. Typical examples include conceptual and linguistic spatial reasoning and abstract geometrical skills.

2. *Representational-format argument*: from the idea that a system is sensitive to sensorimotor correlations it does not follow that information is stored only in sensorimotor terms (embedded by sensorimotor operators). It can well be the case that a variety of different representational formats can be developed by cognitive systems to encode knowledge acquired through sensorimotor experience. “*As organisms move away from sensorimotor links that subserve particular behaviours, they develop structures that carry information in a form accessible to an open-ended range of applications. Rather than acquiring a set of interactive contingencies of the sort needed to guide a specific behaviour, they store information in a format that can be applied in a diversity of ways, should the situations arise. I am hard pressed to see why taking the sensorimotor aspect of vision seriously should lead one to reject such plausible representationalist models*” – Van Gulick, *cit.*

We have already suggested that many arguments of this kind have been raised because of the overstatement in the expected explanatory power of the sensorimotor approach, as defended by its original advocates. We agree that it seems hardly the case that a sensorimotor explanation can fruitfully apply to *any* kind of space-related capabilities. Moreover, since the sensorimotor programme *de facto* has not provided so far any analytical explanation of specific spatial routines, it is legitimate to wonder if sensorimotor approaches might provide a tenable alternative to traditional models. On the other hand, we maintain that arguments about the *de jure* impossibility of explaining certain aspects of spatial competence on the basis of sensorimotor are not receivable unless they are spelled out as explicit challenges or crucial tests for sensorimotor explanations.

Even if the sensorimotor approach is not incompatible with other explanatory strategies for spatial competence, we defend the idea that it is not possible to determine *a priori* which aspects or level of spatial competence can be explained by sensorimotor approaches, and which cannot. For instance, it seems unjustified to claim that no aspect of high-level spatial capabilities might be underwritten (and hence accounted for) by some kind of sensorimotor learning. Serious confutations of the sensorimotor hypothesis require either of the following scenarios: the formulation of articulated *de jure* arguments showing that *specific* sensorimotor models cannot account for *specific* cases of spatial capabilities; the failure of a structured application of a sensorimotor programme to obtain interesting results in explaining spatial competence at different levels.

If, on the one hand, most of the above criticisms seem at least intuitively plausible, on the other hand they look premature in their conclusions: they might be tenable, if grounded, as criticisms to an established research programme (which, so far, does not yet exist); they look too hasty if addressed to an emerging theoretical hypothesis.

SECTION II: TOWARDS A SENSORIMOTOR ACCOUNT OF SPATIAL COMPETENCE

1. FOUNDATIONAL ISSUES ABOUT THE NATURE OF SPATIAL COMPETENCE

(will compare the acquisition of syntax as parameterization of universal grammar to specific linguistic environments to the acquisition of spatial competence: how does statistical learning match internal constraints of the system)

The fundamental tenet of what we call the *traditional view* is the idea that, in order to display a correct spatial behaviour, a cognitive system has to faithfully internalise the relevant spatial properties of the physical world.

Recalling what we said in §1.1, the traditional view assumes that:

1. it exists a definite set of spatial properties belonging to the physical world;
2. the acquisition of a spatial competence demands the internalisation of these properties.

This approach relies upon the presupposition, discussed in section I, that the set of spatial properties ascribed to the physical world can be univocally defined: in other words, traditional theories are afflicted by normative constraints about what has to be taken as a *good* spatial property.

The general goal of this section is to show that the projection of these properties on the physical world is inappropriate, since it attributes to the latter properties that don't belong to it *necessarily*. Consequently, choosing a specific set of constraints (according to the traditional view of the explanation) will lead to the characterisation of a “local” spatial competence, i.e. a competence whose hypothesis are valid only under certain conditions.

The peculiarity of the sensorimotor approach is that it offers, by contrast, the possibility to conceive a theoretical framework in which what characterizes a spatial competence is a set of properties more general than those commonly invoked. In particular, it allows getting a notion of “objectivity of space” that doesn't have to be identified with the respect of the constraints invoked by the representational approaches nor with the representation of immutable properties of the physical world.

Following Philipona et al., *cit.*, we assume that “spatial” has to be defined as “subject to sensorimotor invariants”: we claim that the mastery of rules associating the whole set of a system's motor schemes to specific sequences of sensory stimulation is a necessary and sufficient condition for yielding a spatial competence.

The central question is: under which conditions a cognitive system can extract a sensorimotor invariant, i.e. in which conditions it is capable to establish that a specific regularity governs the association between a class of movements and some classes of sensory patterns?

We propose two conditions:

- a. *a parte obiecti*, (S,M) relations have to manifest a minimal degree of regularity. A world reacting to the subject's movements in a totally random way would preclude, to the latter, any possibility to develop a spatial competence;
- b. *a parte subiecti*, the cognitive system has to be minimally equipped with a capacity to find regularities between sensory patterns and self-motion. A subject who couldn't identify the equivalence between some sensory patterns properties as co-occurring with analogous motor schemes couldn't get any spatial competence.

Moreover, as we explicitly claimed in § 1.2.1., sensorimotor correlation can be the object of a learning process insofar as they are statistically regular co-occurrences of sensory and motor patterns: it is because of their regular association (co-occurrence) that (S,M) relations are good candidates for becoming the object of a learning process. From this claim two parallel consequences have to be drawn: first, the sensorimotor approach assumes that in many cases spatial competence can only be properly accounted by appealing to the interaction between specific classes of system's endogenous activity (what we call "motor patterns") and specific classes of exogenous perturbations (what we call "sensory patterns"); second, our approach does not exclude that some other kinds of statistically regular correlations (not necessarily sensory-*motor*) can shape the acquisition of spatial competence¹⁵, insofar as these patterns could affect the system's sensitivity. Simply, we maintain that the influence of other kinds of correlations is less salient in that it is statistically unlikely that spatial learning processes take place without any contribution of co-occurring *motor* patterns.

2. EXPERIMENTAL RESEARCH DIRECTIONS

Is it empirically possible to show that spatial competence can be characterized in a much more general way than that of the traditional view? We argue that, (by) modifying underlying (S,M) relations between the agent and its environment, its spatial competence might be manipulated. Such evidence might shed some light on the fact that, since spatial learning is the result of the fixation of specific SMI (that can be dramatically change as a function of different environmental contexts), traditional explanatory strategies might turn out to be intrinsically contextual, i.e. framed to a specific (one among many possible others) instance of spatial regularities.

We suggested that, in a sensorimotor perspective, spatial competence consists in the acquisition of invariants associating motor schemes to sensory patterns. This point of view leads to the (testable) idea that, by modifying systematically the relation between motor commands and co-occurring sensory patterns, a cognitive system might develop a coherent and alternative form of spatial competence.

This suggests that characterizing spatial competence by formulating constraints on

¹⁵ There exists in fact a large philosophical literature concerning the "minimal requirements" for spatiality and objectivity. In many cases, metaphysical theories of space do not take action as an essential condition for defining "minimal" notions of space. See Strawson (1959), Evans (1985) and Proust (1997).

internal representations turns the explanation into a contextual one.

The growing technical possibilities offered by devices for the manipulation of environment's properties (like Virtual Reality) represent a concrete opportunity to test this hypothesis.

3. WELL-POSED WORKING QUESTIONS.

To what extent problems raised by the sensorimotor approach are susceptible to orient research in a qualitatively different direction? Here's a number of problems and working questions that might interest the study of acquisition of spatial competence from a sensorimotor perspective:

- The problem of the “identity” of two sensorimotor invariants;
- The problem of the threshold of sensitivity for mechanism monitoring (S,M) couplings.
- The problem of granularity, discrimination, difference between classes of SMI.
- The problem of a causal relation between action and sensory patterns and of their spatio-temporal correlation (could one learn to associate motor schemes with sensory events that *precede* the realisation of movement?)
- The (empirical) problem of plasticity and adaptability of the human perceptual system to different SM conditions.

CODA

(will contain conclusive remarks on the notion of space assumed in mainstream research)

In this article, we referred to some general concepts, which are usually considered as intuitive notions upon which there is implicit agreement. We assume that a scientific theory of spatial competence cannot rely upon an intuitive characterisation of the explanandum, much in the same way as a scientific theory of linguistic competence cannot refer to an intuitive characterisation of what grammar is.

From the viewpoint of a general characterisation of what could count as *spatial* competence we start with no preliminary hypotheses about what space *is*: Relying on judgments of spatial competence by the community of spatial agents amounts to considering spatial competence in analogy to the way in which grammaticality is judged by speakers of a specific linguistic community. Spatial competence – we argue – is the result the fixation of some parameters that subjects learn by interacting with the world.

REFERENCES

1. Bach-y-Rita, P. (1972) *Brain mechanisms in sensory substitution*. Academic Press
2. Bach-y-Rita, P. (1984) The relationship between motor processes and cognition in tactile vision substitution. In: *Cognition and motor processes*, ed. A. F. Sanders & W. Prinz. Springer.
3. Berthoz, A. (1997) *Le sens du mouvement*, Odile Jacob.
4. Berthoz, A. (1991) Reference frames for the perception and control of movement. In: Paillard J (ed) *Brain and Space*. Oxford Science Publications, Oxford, 81-111.
5. Cutting J.E. (1986), *Perception With an Eye for Motion*, MIT Press.
6. Eilan, N., Brewer, W., McCarthy, R. (1993) *Spatial Representation*, Blackwell, Oxford.
7. Evans, G. (1985) *Collected Papers*, Oxford Clarendon Press.
8. Gibson, J.J. (1979) *The ecological approach to visual perception*, Houghton Mifflin.
9. Hartley T., Burgess N. (2002) *Models of spatial cognition*. Encyclopaedia of Cognitive Science, MacMillan.
10. Hatfield, G. (2003) Representation and constraints: the inverse problem and the structure of visual space, *Acta Psychologica* , 355-378.
11. Hecht, H. (2001) Regularities of the Physical World and the Absence of their Internalization, *Behavioral and Brain Sciences* 24 (3).
12. Helmholtz, H. (1878/1977) *Epistemological writings*, The P. Hertz/M. Schlick Centenary Edition of 1921, English translation by M.F. Lowe, Reidel Publishing Company.
13. Johansson, G. (1964). Perception of motion and changing form. *Scandinavian Journal of Psychology*, 5, 181-208.
14. Klatzky R. L.,(1998) Allocentric and Egocentric Spatial Representations: Definitions, Distinctions, and Interconnections, in: In C. Freksa, C. Habel, & K. F. Wender (Eds.), *Spatial cognition - An interdisciplinary approach to representation and processing of spatial knowledge* (Lecture Notes in Artificial Intelligence 1404), Berlin: Springer-Verlag.
15. Kubovy, M. (2001), Internalization: A metaphor we can live without, *Behavioral and Brain Sciences*, 24 (3).
16. O'Keefe, J. & Nadel, L. (1978), *The Hippocampus as a Cognitive Map*. Oxford: Oxford University Press.
17. O'Regan, J.K. & Noë, A. (2001) A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* 24 (5), 939-1011
18. O'Regan, J.K. (1992) Solving the 'real' mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, 46, 461-488.
19. Palmer, S. E. (1999), *Science Vision: photons to phenomenology*, The MIT Press, Cambridge, Massachusetts.
20. Palmer, S. E. (1978), Fundamental aspects of cognitive representation, in: E. Rosch & B. Lloyd (Eds.), *Cognition and Categorisation*, (pp. 261-304). Hillsdale, NJ: Erlbaum.

21. Pani, J. R. (2001) Perceptual theories that emphasize action are necessary but not sufficient, in: O'Regan, J.K. & Noë, A. (2001) A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* 24 (5), Open Commentaries.
22. Philipona, D., O'Regan, J.K., Nadal, J.-P. (2003) Is there something out there? Inferring space from sensorimotor dependencies. *Neural Computation* 15 (9).
23. Pylyshyn, Z. (2001), Seeing, acting, and knowing, in: O'Regan, J.K. & Noë, A. (2001) A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* 24 (5), Open Commentaries.
24. Pouget, A., Ducom J.C., Torri J., Bavelier, D. (2002) Multisensory spatial representations in eye-centered coordinates for reaching, *Cognition*, 83(1), B1-11.
25. Proust, J. (1997) Espace sens et objectivité, in : J.Proust (Ed.) *Perception et Intermodalité*, PUF.
26. Rossetti, Y. (1997) Des modalités sensorielles aux représentations spatiales en action : représentations multiples d'un espace unique. In *Perception et Intermodalité. Approches actuelles du problème de Molyneux* (Proust, J., ed.), PUF.
27. Schlesinger, M. (2001) Reexamining visual cognition in human infants: On the necessity of representation, in: O'Regan, J.K. & Noë, A. (2001) A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* 24 (5), Open Commentaries.
28. Scholl, B.J. & Simons D.J.(2001), Change blindness, Gibson, and the sensorimotor theory of vision, in: O'Regan, J.K. & Noë, A. (2001) A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* 24 (5), Open Commentaries, 1004-1006.
29. Shepard, R.N. (2001), Perceptual-Cognitive Universals as Reflections of the World, *Behavioral and Brain Sciences*, 24 (3).
30. Soechting, J.F., & Flanders, M. (1992). Moving in three-dimensional space: frames of reference, vectors, and coordinate systems. *Annual Review in Neuroscience*, 15, 167–191
31. Soechting, J.F., & Flanders M. (1989) Sensorimotor representations for pointing to targets in three-dimensional space. *J. Neurophysiol.* 62: 582-594.
32. Strawson, P. F. (1959) *Individuals*, London, Methuen and Co.
33. Todorovic, M. (2001), Is kinematic geometry an internalized regularity?, *Behavioral and Brain Sciences*, 24 (3).
34. Tolman, E.C. (1948) Cognitive maps in rats and men, *The Psychological Review*, 55(4), 189-20
35. Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press
36. Van Gulick, R. (2001), Still room for representations, in: O'Regan, J.K. & Noë, A. (2001) A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* 24 (5), Open Commentaries.
37. Wexler M., Lamouret I., Droulez J.,(2001) The stationary hypothesis : an allocentric criterion in visual perception, *Vision Research* (41), 3023-3037.
38. Wolpert, D.M., Ghahramani, Z., Jordan, M.I. (1995) An internal model for sensorimotor integration. *Science* 269 (5232), 1880-1882.

APPENDIX 1: INTERNAL ARTICULATION OF THE “TRADITIONAL” VIEW

The original purpose of the ecological psychologists can be described as the idea to offer an alternative approach founded on the refusal of the axiom claiming the discrete and instantaneous character of sensory input, and the consequent necessity of a cognitive elaboration in order to get a proper representation of the external world. Gibson and his successors proposed to consider perception as a temporally extended process: the cognitive system doesn't work with a succession of instantaneous stimuli, but rather with a continuous flow of stimulation, in which it can make a difference between what varies and what does not vary during the dynamic evolution of sensory configurations. Namely, what do not vary are the perceptual invariants, specifying the spatial layout of the environment. As James Cutting explains clearly in his *Perception with an eye for motion* (1986), the ecological approach holds that invariants specify spatial properties of the environment: by picking up an invariant we perceive the correlated environmental spatial properties.

This kind of specification between the invariant and the spatial property is an example of what we call “the traditional view” about representational relations. According to the Ecologism, spatial properties are intrinsic to the environment, and they can be described independently from the activity of the subject. (In other words, Gibson is, epistemologically, a realist): the cognitive system extracts the properties, but it does not contribute in any way to their constitution.

What distinguishes Gibson's approach is its characterisation of the homomorphic mapping between the perceptual invariant and the correlated property of the environment: according to him, this mapping is constant and one-to-one. Gibson argues that if the flow of stimulation is rich enough (and in ecological setting this is the case), then it conveys univocal, non-ambiguous and sufficient information about a specific environmental spatial property. That means that the cognitive system does not need to elaborate or construct the correct representation: the sensory stimulation provides sufficient information about the external world. However, the (explicit) assumption is that there is an external and objective space that is specified (represented) by the perceptual invariants: we have the two essential properties of the “traditional view”.

These assumptions have critical and concrete consequences concerning the perceptive role of a subject's movement. It is well known that Gibson has strongly stressed the importance, for the subject, to move around in his environment in order to obtain a sufficiently rich sensory stimulation. In particular, the continuous optic flow provoked by the movement allows the subject to distinguish between what varies and what does not vary in stimulation: movement reveals invariants; and invariants specify spatial layout of the environment.

Gibson's theoretical influence on posterior generations has been strong. Today, almost everyone accepts the general theoretical framework he proposed about perception processes. Nevertheless, at least one central aspect has been criticized and reformulated. We would show that, even if scientists have challenged Gibson's approach on this crucial point, their approaches can be still considered, in our terms, “traditional”.

The crucial point concerns the nature of the mapping between the invariants and the external spatial properties. Where Gibson thought that the relation between them is univocal and specific (one-to-one mapping), later cognitive scientists have argued that, in the general case, information is ambiguous or not sufficient (or both), so that the sensory stimulation does not specify clearly the corresponding property of the external space. If the relation is ambiguous, the literature speaks of a “one-to-many” mapping, which constitutes a problem (given its not symmetric nature) that the cognitive system is supposed to solve: this is what traditionally is referred to as the “underspecification

problem” (Hecht, 2001).

According to some researchers, subject movement can be sufficient, in some cases, to solve the problem and disambiguate the information or to provide enough information (as Gibson claimed), so that the system can choose a unique “interpretation” about what it perceives. Nevertheless, cognitive scientists agree in claiming that in the general case dynamic stimulation provoked by movement is still insufficient to provide clear information.

The most important hypothesis scientists use in order to solve this serious problem is what is called “the internal model” or “internalisation” hypothesis. Very broadly, appealing to an internal model means to suppose that the cognitive system has and uses a certain kind of knowledge about what has to be perceived allowing it to constrain the number of possible interpretations of the sensory stimulation. As Hecht (2001) points out, “According to this claim, the mind has internalised universal principles (regularities) that allow it to disambiguate situations that would otherwise be unsolvable. Provided the world is not changing, such universal principles are very efficient. For the visual system, this explains why we can make sense of stimuli that by themselves do not suffice to specify our perceptions”.

As Hecht recalls, this kind of hypothesis appeals to the idea that, along evolution, the cognitive system has internalised some external regularities [Shepard, 1984] characteristic of the physical environment. Specific hypothesis about internalised spatial knowledge are: Chasles’s theorem of kinematic geometry [Shepard, 1984], the “rigidity principle” [Johansson, 1964; Ullmann, 1979] or, more recently, the “stability principle” [Wexler, 2001]. For instance, as Todorovic (2001) stresses, “one way to explain the origin of this principle [rigidity] is to propose that overexposure to rigid body motions in the environment somehow has induced the perceptual apparatus, over the course of evolution, to prefer the rigid interpretation over the infinity of geometrically equally appropriate non-rigid interpretations of stimuli”.

What is important for our discussion is to underline that, even if all these approaches admit explicitly the fundamental role of movement (and, more in general, of action-perception loops) in extracting and perceiving the spatial properties of the environment, they accept the basic “traditional” idea according to which 1) what has to be represented are a set of properties that the physical world intrinsically possesses and 2) the cognitive system have to represent these properties as correctly as possible. In this sense, this set of theories and approaches are essentially different from what we call, in this paper, a sensorimotor approach.

APPENDIX 2: THE “CONTROL PARADIGM” VS. THE “ENACTIVE” INTERPRETATIONS OF SENSORIMOTRICITY.

(Why the enactive approach do not require a full-fledged theory of action)

Sensorimotor comparators can, and normally do, work in an oriented way, i.e. the system’s control on the motor patterns generates a hierarchy between the components of a SMI.

Sensorimotor comparators, as far as they instantiate SMI, can be described in a symmetric manner, i.e. there’s no an a priori causal interpretation of (S,M) relation (“enactive” interpretation). Nevertheless, the same abstract description of sensorimotor comparators is compatible with different concrete realisations in physical systems. In particular, the comparators’ physical realisation can be subject to specific constraints or properties intrinsic to the system: in the case of humans or animals, for instance, the status of the motor component is different from the sensory component, as far as perceptual systems seem able to exercise a control on motor pattern (which are often called “motor commands” or “motor” programs”) unknown for sensory patterns. Consequently, for physical perceptual systems the symmetry of sensorimotor correlation is often broken and oriented in a specific way.

In this appendix, we compare two systems that both involve a sensorimotor operator. The first one belongs to what we call the “control paradigm”. Classically, in order to reach a goal, the agent must narrow the gap between a target “behavioural value” and the current one. In a more sophisticated way, many systems predict the future sensory state given a motor scheme and then compute its difference with the target. The motor action that minimizes the gap is selected. Note that this process remains as asymmetric as the one we discussed earlier [O’Regan & Noë, 2001], because the sensor value *triggers* the action, which *modulates* the sensor value and so on.

The second one belongs to what we can label as an “enactive” interpretation of sensorimotricity. It describes acquisition and mastery of SMI as based on filters of (S,M) relations (as defined by our functional hypothesis). While the systems evolves on its own, according to an R (S,M) law unknown to the system but that it may estimate, the filter estimates a set of possible relations. No a priori causal interpretations of R(S,M) are involved, as in the case of the “control paradigm” interpretation of sensorimotricity.