THE PSYCHOLOGICAL REALITY OF COGNITIVE THEORIES: CONCEPTUAL ROOM FOR THE BRAIN AS A FUNCTIONAL *BRICOLEUR*

Dissertation presented by

Oscar Vilarroya Oliver

and directed by

Daniel Quesada

Universitat Autònoma de Barcelona
Bellaterra, September 1998

Chapter 4

Empirical foundations

"Can you do Addition?" the White Queen asked. "What's one and one and one and one and one and one and one and one and one and one?"
"I don't know," said Alice. "I lost count."
"She can't do Addition," the Red Queen interrupted.
Through the Looking Glass

The last two chapters have been devoted to examining in detail Davies' and Peacocke's proposals. We have concluded that they provide two promising, though still insufficient, accounts of psychological reality. On the one hand, Davies' proposal seems to be, prima facie, a robust conception. His account gives content to the notion of psychological reality: A theory is psychologically real if it describes the causal structure in the mind of the cognizer that accounts for the competence to be explained. I have presented nevertheless objections concerning the development of such conception in a specific criterion proposed to legitimize the psychological reality of a specific theory, i.e., what Davies labels the *Mirror Constraint*: Theory *T* is real if its derivational structure matches the causal-explanatory structure in the mind of the cognizer. Such a criterion has been found to be too weak to account for the psychological reality of a theory since the Mirror Constraint purports to compare two incomparable elements, namely, a mental element with its mode of characterization. I have argued that it is epistemologically impossible to separate these elements and, *a fortiori*, to compare them. Additionally, I have found the basic notion to be too strong insofar as it cannot accommodate some genuine possibilities of knowledge attribution.

On the other hand, we have seen that, according to Peacocke, a theory is psychologically real if it complies with an explanation of what he reckons as level 1.5. This level identifies the information "drawn upon" by an algorithm, that is, the information that the states of the algorithm carry and which causally influences the algorithm. My analysis has developed some objections to the account. The bottom line of these objections is the fact that a *pure* informational account will come short of providing a pervasive notion. We have seen some examples that lend support to the view that mechanisms that draw upon the same information, that is, equivalent with respect to the Informational Criterion, leave certain relevant cognitive differences unspecified. In essence the question is that there are relevant cognitive differences that cannot be specified in informational terms. I have argued that it is not a necessary condition that the information that theories draw upon is the *solely* causally significant description of the mechanism at the cognitive level. I have tried to show that these examples are not exceptions since many cognitive theories have a more or less important component directed at explaining *how the cognitive system handles information*.

A logical follow-up would be now to try modifying the two accounts. However, it is my opinion that both accounts are founded on an underlying assumption about functional ascriptions that would compromise any revision. This assumption could be succinctly put as the idea that functional attributions *must* be framed, in cognitive science theorising, within Grandpa's explanatory framework presented in Chapter 1. Specifically, functional attributions must accord with the explanatory "classical cascade", as well as with the functional analysis strategy. Yet, it will be my task in the rest of the dissertation to provide sufficient argumentation to the contrary, as well as to provide some modifications to the framework to give a more efficient account about the psychological reality of theories. My analysis is that there is a distinction which is orthogonal to those established in Grandpa's framework, and this distinction concerns two explanatory projects. One corresponds to the explanation of the way in which cognitive agents seem to comply with a class of environmental (and possibly evolutionary significant) tasks, what I will call *capacity*; the other is consistent with the project of explaining the processes within the system that account for the satisfaction of the task, what I will call the *action*. My hypothesis is that

Davies' proposal overlooks this distinction, whereas Peacocke's account concerns the *capacity*, and problems appear because the Informational Criterion conflates the *capacity* and the *action* attributions.

In this chapter my aim is to examine different empirical projects within cognitive science in order to see how Grandpa's explanatory framework can be adapted to provide a psychologically real account of cognitive theories. The choice has been made according to three criteria. The first condition is that the project to be examined should have been characterized with reference to a well-established *task* level description. This is not easy since, in general, cognitive experiments are meant to describe the *algorithmic* level. Second, the project should not pertain to a domain where the knowledge at the task level comprises theories with major discrepancies. For example, the knowledge to be attributed in grammatical competence is of a great diversity, ranging from functionalist theories of language, through different versions of Chomskyan theory all the way up to connectionist projects. Such great discrepancies can defuse some of the points I want to make. Third, the project has to be articulated: it cannot be only a compilation of evidence subsumed under a particular or domain-specific theory, nor can it be some clinical identification, or a coarse-grained functional theory.

I have chosen four projects. The first one is the work of Shimon Ullman (1996) in what is known as visual cognition. Specifically it is the presentation of his theory about the way in which the brain perceives spatial relations, essentially during object perception. The second is the work of Dehaene (1992, 1997) in arithmetical cognition, that is, the cognitive abilities related to calculation and mathematical thinking. The third corresponds to the integrated philosophical and empirical work on object perception and motion undertaken by Spelke and Van der Walle, Munger and Cooper, and Peacocke compiled in the book *Spatial representation* (1993). Finally, I present some work of Steels (1994) in what is known as Artificial Life.

The presentation of each project will be divided into two sort of descriptions. On the one hand, I will describe the sort of *task* that the theory wants to account for and, on the other, I will present the sort of *cognition* supposedly employed in its satisfaction.

My aim in this chapter is metatheoretical; it is meant to test Grandpa's framework rather than providing a discussion of empirical evidence. Therefore, I will not present the projects as a review of literature, giving evidence for each hypothesis; rather, I am going to review the projects *as if* the evidence is undisputed, which obviously it is not. Likewise, I will not go into digressions about internal problems or criticisms made about the interpretation of the theories.

4.1.Visual cognition

Shimon Ullman (1996) has proposed an integrated theory of visual perception. For Ullman we use certain sorts of information, such as shape and spatial information, to perform a wide variety of tasks, among them object recognition and classification, but also for manipulating objects, planning and executing movements in the environment, selecting and following a path and the like. The visual analysis of shape and spatial relations among parts corresponds to what he calls "visual cognition".

Human's visual system is indeed remarkably adept at establishing a variety of spatial relations among items in the visual input: the visual extraction of spatial information is remarkably flexible and efficient. This ability is supported by the fact that the perception of spatial properties and relations that are complex from a computational standpoint nevertheless often seems immediate and effortless. However, the immediateness and ease of perceiving spatial relations is only apparent. As Ullman argues, underlying such an effortless process there is in fact a complex array of processes that have evolved to establish certain spatial relations with considerable efficiency. Ullman recognises that the mechanisms are still ill-understood. We are far from understanding, for example, how we compare the length of two line segments. We don't know at the moment whether we apply a sort of "internal yardstick" to the segments, shift them internally to test their overlap, or whether we use some other resources.

Ullman's proposal is an approach to account for how such efficiencies might be performed. His model states that perception of shape properties and spatial relations is achieved by the application of what he calls "visual routines" to early visual representations. These visual routines are efficient sequences of basic operations that are "wired into" the visual system. Routines for different properties and relations are then composed from the same set of basic operations, using different sequences. Using a fixed set of basic operations the visual system can assemble different routines and in this manner extract an essentially unbounded variety of shape properties and spatial relations that the organism needs to analyze. Within this framework, to understand visual cognition in general, it will be required to identify the set of basic operations used by the visual system. An explanation of the task of determining a particular relation such as "above", "inside", "longer-than" or "touching" would require a specification of the visual routine used to extract the property in question.

4.1.1.Task

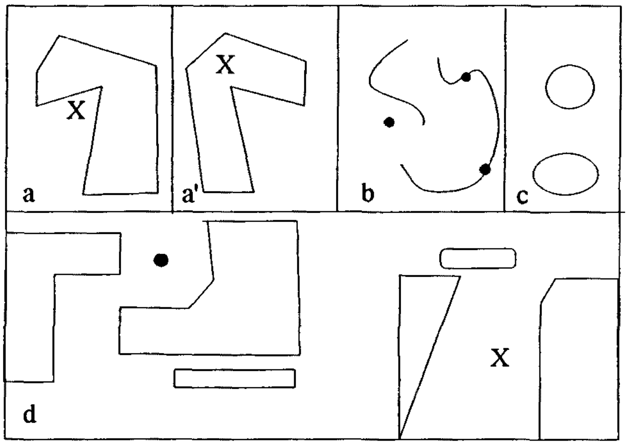


Figure 4.1. Examples of several "visual cognition" tasks involving the analysis of shape and spatial relations. (a,a') Inside/outside relation: for humans it is easy to determine whether the "X" lies inside or outside the closed figures. (b) The task is to determine whether two black dots lie on a common contour. (c) Elongation judgments, the task being to determine which is the most elongated figure. (d) The task is to determine whether the black spot can be moved to the location of the "X" without colliding with nearby shapes. (From Ullman 1996, p. 264)

Looking at an image such as figure 4.1, we can obtain almost immediately answers to a variety of questions regarding shape properties (which refers to a single item) and spatial relations (such as "above", "inside", and the like which refer to two or more items). It is easy for humans to determine only visually whether in figure 4.1.a the *X* lies inside or outside the closed curve. Ullman suggests that the answer just "pops out", and that we cannot give a full account of how the decision was reached. Such ability appears to be associated with relatively advanced visual systems. The pigeon, for example, shows a great capacity for figure classification and recognition but it is unable to perform these sort of tasks in a general manner. It can respond correctly only for simple figures, and appears to base its decision on simple local cues such as the convexity or concavity of the neighbouring contour. In figure 4.1.c the visual system has to detect elongations (which figure is the most elongated?), a task at which we are very efficient: for judgements of elongation can be made when the major axis of the ellipse is only 4-5-% longer that the minor axes. In figure 4.1.b.

the solution is obtained merely by looking at the picture. The task at 4.1.d. seems to be the more difficult, but it seems that we can use our visual capacities to simulate the motion and obtain the correct answer.

Ullman obviously acknowledges that these tasks are artificial, but argues nevertheless that similar visual cognition problems occur in natural settings, when we manipulate objects, plan actions, reason about objects in a scene, or navigate in the environment. It is precisely due to such basic visual abilities that we use visual aids such as diagrams, charts, sketches and maps. Such structures draw on the cognitive system's natural capacity to manipulate and analyze spatial information, and this ability can be used to help our reasoning and decision processes.

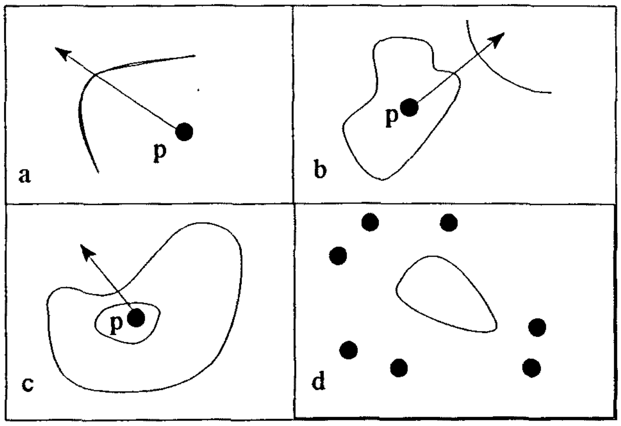
Let us suppose then that the task we want to examine is the determination of the tasks described in figure 4.1., such as "inside/outside", "above/under", "lying in/lying outside", and "elongation" relations. Say that we subsume such tasks under the label of "spatial-relation tasks", and let us finally suppose that these tasks should be considered to be *fundamental* to be able to perform an efficient engagement with the world.

Ullman considers that one possible reason for our proficiency at establishing inside/outside relations is their potential value in figure-ground segmentation: when a bounding contour of an object is identified, features inside the contour belong to the object and features outside it to belong the surroundings. The immediate perception of the specific inside/outside relation is subject to some limitations, though: when the bounding contour becomes highly convoluted, the distinction between inside and outside becomes more difficult. In any case, these limitations are fairly loose, however, and the computations performed by the visual system in distinguishing inside from outside exhibit considerable flexibility: the shapes can vary, and the positions of the *X* and the lines do not have to be known in advance.[18] To see the complication of computing such cases, the fact is that no

[18] In mathematics there is the "Jordan curve theorem" that states that a simple closed plane curve separates the plane into two disjoint regions, its inside and outside. This theorem requires an elaborate proof, contrary to the intuitiveness of the concepts of inside and outside.

algorithm has been found that can compute them with sufficient accuracy.[19]

4.1.2.Cognition



4.2. Inside/outside difficult tasks. (a) An open curve. (b, c) Embedding relations. (d) Parallel decisions. (From Ullman 1996, p. 271)

In order to explain the perception of inside/outside relations it is necessary, according to Ullman, to reveal the constituent operations that are actually employed by the visual system, and how they are used in different situations. For him the immediate perception of seemingly simple spatial relations requires complex computations that are difficult to unravel. His proposal is that there are a set of basic operations that the visual system can assemble in what he calls "visual routines" in order to extract from the visual representations shape properties and spatial relations. The idea is that visual routines are used by the cognitive system to perform a variety of tasks using a fixed set of elementary operations. The appropriate routine to be applied in a given situation depends on the *goal* of the computation, which is established by the requirements of the organism, and on various parameters of the configuration to be analyzed by the system. Therefore, it can happen that for a determinate *goal* the system uses different routines in different contexts.

Ullman draws a distinction between early visual representations and subsequent representations, what Ullman labels "incremental" representations. The earlier

[19] One algorithm that has been proposed as a way to establish whether a given point lies inside or outside a closed curve. The method, named ray-intersection, is based on drawing a ray emanating from the point in question, and extending it to "infinity". The number of intersections made by the ray with the curve is recorded. If the resulting intersection number is odd, the starting point of the ray lies inside the closed curve. If it is even (including zero), then it must be outside. This procedure has been implemented in computer programs and it has been successful. However, this method fails to account for the problems considered in figure 4.2. First, the algorithm must assume that the curve is closed, otherwise an odd number of intersections would not be indicative of an "inside" relation (as in 4.3.a). Second, it must be assumed that the curve is isolated: in figure 4.3.b and c, point *p* lies within the region bounded by the closed curve, but the number of intersections is even. These limitations are not shared, according to Ullman, by the human visual system: in all of the above examples the correct relation is easily established.

representations are produced prior to the application of visual routines. *They are produced in an unguided bottom-up manner, determined by the visual input and not by the goal of the processing.* Examples of early visual representations include the *extraction of edges and lines* from the image, the computation of motion, disparity, and colour. Ullman draws a sharp contrast between these early processes from the application of visual routines, and the properties and relations they extract, which are not determined by the input alone. For the same visual input different aspects will be made explicit at different times, depending *on the goals of the computation.* Unlike the base representations, the computations by visual routines are not *applied uniformly* over the visual field. In that sense, not all the possible inside/outside relations in the scene are computed, but only with regard to selected objects.

Another distinction between the two stages is that the construction of the early representations is essentially fixed and unchanging, while visual routines are open-ended and permit the extraction of newly defined properties and relations. For various visual tasks, the analysis of the visual information therefore divides naturally into two distinct successive stages: the creation of what Ullman calls "base representations", followed by the application of visual routines to these representations. The application of visual routines can define objects within the base representations and establish properties and spatial relations that cannot be established within the base representations.

I will now present Ullman's proposal on elementary operations. The exposition is only partial, that is, it does not include a complete set of possible operations. Additionally, Ullman justifies the use of examples based mainly on schematic drawings rather than natural scenes by arguing that simplified artificial figures allow more flexibility in adapting the pattern to the operation under investigation. He argues that, as long as we examine visual tasks for which our proficiency is difficult to account for, we are likely to be exploring useful basic operation even if we use simplified drawings rather than natural scenes. In fact, for Ullman, our ability to cope efficiently with artificially imposed visual tasks underscores two essential capacities in the computation of spatial relations. First, that the computation of spatial relations is flexible and open-ended: new relations can be defined and computed efficiently. Second, it demonstrates our capacity to accept non-visual specifications of a task and immediately produce a visual routine to meet these specifications.

4.1.2.1.Elementary Operations: Shifting the Processing Focus. In order to execute visual routines, it is necessary to control the location at which certain operations take place. In this sense, the operation of area activation will be of little use if the activation starts simultaneously everywhere. The fact is that it must start at a selected location, or along a selected contour. In other words, in applying visual routines it would be useful to have a "directing mechanism" that will allow the application of the same operation at different spatial locations. Directing the processing focus may be achieved partly by moving the eyes. However, this seems to be insufficient. Many shape properties and relations can be established without eye movements. A capacity to shift the processing focus internally is therefore required, and it has to be effective starting from early processing stages. Ullman believes that the focus of visual processing can be directed, either voluntarily or by manipulation of the visual stimulus, to different spatial locations in the visual input

Selecting a Location. According to Ullman, the mode of operation that seems to be used by the visual system in shifting the processing focus is based on the extraction of certain salient locations in the image, and then shifting the processing focus to one of these distinguished locations. Such salient locations are detected in parallel across the base representations, and can then serve as "anchor points" for the application of visual routines. As an example, he proposes to imagine that a page of printed text is to be inspected for the occurrence of the letter "A". In a background of similar letters, the "A" will not stand out, and considerable scanning will be required for its detection. If however, all the letters remain stationary with the exception of one which is jiggled, or if all the letters are red with the exception of one green letter, the odd-man-out will be immediately identified. The identification of the odd-man-out letter proceeds, in Ullman's theory in several stages. First, the odd-man-out location is detected on the basis of its unique motion or colour properties. Next, the processing focus is shifted to this odd-man-out location. As a result of this stage, visual routines can be applied to the figure. By applying the appropriate routines, the figure is identified. A similar process also played a role in the inside/outside example above. It was noted that one plausible strategy is to start processing at the location marked by the X figure. This raises a problem, since the location of the X and that of the closed curve are not known in advance. If the X is somehow sufficiently salient, it can serve to attract the processing focus, and the execution of the appropriate routine can start immediately at that location.

Defining a Location. Ullman's proposal is then that the focus of processing can be manipulated by moving it to a salient location in the scene. But then, what defines a distinguished location, that can be used for the purpose of shifting the processing focus and applying further operations? Ullman argues that certain odd-man-out locations which are sufficiently different from their surroundings can attract the processing focus directly, and eliminate the need for lengthy scanning. For example, differences in orientation and direction of motion can be used for this purpose while more complex distinctions, such as the occurrence of the letter "A" among similar letters, cannot define a distinguished location. These data are also in agreement with the physiological evidence. Properties such as motion, orientation, colour and binocular disparity are found to be extracted in parallel by units that cover the visual field. These units appear always to be active and unchanged whether the animal is awake, anaesthetized, or naturally sleeping. According to Ullman, these properties are suitable, on physiological grounds, for defining distinguished locations prior to the application of visual routines.

In sum, for Ullman the way of shifting the processing focus around in the course of applying visual routines is by first extracting a set of distinguished locations in the scene, and then shifting the processing focus towards one of these locations. In defining the distinguished locations, a small number of elementary properties, such as orientation, contrast, colour, motion, binocular disparity, and perhaps a few others, are computed in parallel across the early visual representations, prior to the application of visual routines. Simple differences in these properties can then be used to define distinguished locations. These locations can be used in visual routines by moving the processing focus directly to one of the distinguished locations, without the need of extensive search or systematic scan.

4.1.2.2.Elementary operations: "Colouring" and Incremental representations. The bounded activation, or "colouring" operation is a process that consists in the activation (the "colouring") of a given area. The method requires the spread of activation over a surface in the visual representation originating from a given location or contour, and stopping at discontinuity boundaries. Starting from a given point, the area around another point in the internal representation is activated. This activation spreads until a boundary is reached, but it is not allowed to cross the boundary. Depending on the starting point, either the inside or outside of the curve, but not both, will be activated. This can provide a basis for separating inside from outside. An additional stage is still required, however, to complete the procedure, and this additional stage will depend on the specific problem at hand in which we can apply different algorithms. According to Ullman, the colouring method has to be complemented to surmount certain difficulties to obtain, by itself, inside/outside relations generally.

For Ullman, the results of the colouring operation may be retained for further use by additional routines. Colouring provides in this manner one method for defining larger units in the initial visual representation: the "coloured" region becomes a unit to which routines can be applied selectively. Recognition routines could then concentrate on the activated region, ignoring the irrelevant contours. An example along this line is illustrated in figure 4.3.

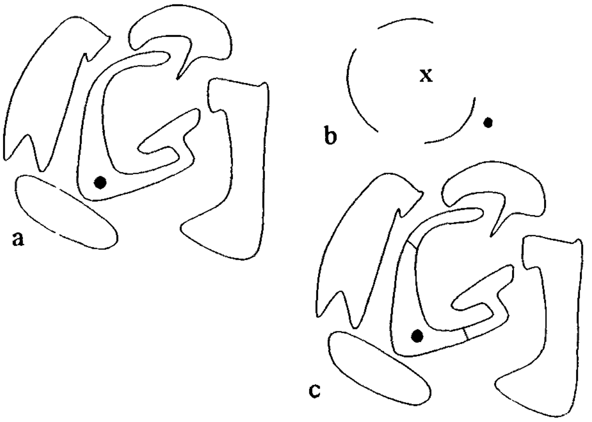


Figure 4.3. Examples of the colouring operation. In *a*, the visual task is to identify the subfigure containing the black dot. Such a figure, a letter G, can be recognized despite the presence of interfering figures. In *b* the boundaries are fragmented, but judgments of inside/outside can still be done. In *c*, additional lines are included which implies that the activation must spread across them to "colour" the G. (From Ullman 1996, p.295)

Discontinuity Boundaries for Colouring. One interesting question it is *what constitutes a discontinuity boundary* for the activation operation. The bounded activation, and in particular its interactions with different contours is not a simple process. It is possible that as far as the activation operation is concerned, *boundaries are not defined universally, but may be defined somewhat differently for different routines.*

4.1.2.3.Elementary operations: Boundary tracing. Ullman suggests a basic operation that could play a useful role in visual routines is the tracking of contours in an internal visual representation, above and beyond considering that contours and boundaries of different types are fundamental entities in visual perception. For instance, a case that benefits from the operation of contour tracing is the problem of determining whether a contour is open or closed. If the contour is isolated in the visual field, a solution can be obtained by detecting the presence or absence of places where contour terminates. This strategy would not apply, however, when other contours are present. The problem can be solved by tracing the contour and testing for the presence of termination points on the traced contour. Another example of the possible use of boundary tracing goes back to the inside/outside example discussed above. Tracking can be used in conjunction with the area activation operation to establish inside/outside relations, by moving along a boundary, colouring only one side. If the curve is closed, its inside and outside will be separated. Otherwise, the fact that the curve is open will be established by the colouring spread, and by reaching a termination point while tracking the boundary. One interesting case is that when we deal with coloured curves, we employ a different strategy. The task seems to be solved by using a simplified strategy, of checking whether two points lie on a curve of the same colour. The visual system seems to *be capable of using this sort of shortcut, and adjusts its global strategy accordingly*, without deliberate or conscious planning.

In sum, the tracing and activation of boundaries are useful operations in the analysis of shape and the establishment of spatial relations. This is a complicated operation since flexible, reliable tracing should be able to cope with breaks, crossings, and branchings all of which have *different resolution* requirements. Even if the examples are only schematic, we have to take into account that if boundary tracing is indeed a basic operation in establishing properties and spatial relations, it should be expected to be applicable not only to such contours, but also to the different types of contours and discontinuity boundaries in early representations, such as boundaries defined by discontinuity in depth, motion and texture.

4.1.2.4.Elementary operations: Marking. Ullman holds that in the course of applying a visual routine, the process shifts across base representations, from one location to another. Of course, to control and coordinate the routine, it would be useful to have the capability to keep a partial track of the locations already visited. Ullman proposes that a simple operation of this type is the marking of a single location for future reference. This operation can be used, for instance, in establishing the closure of a contour. As said above, closure can be tested using a combination of tracing and marking. The starting point of the tracing operation is marked, and if the marked location is reached again the tracing is completed, the contour is classified as closed. In figure 4.4. there is a task that consists of determining visually whether there are two X on the same curve. The correct answer is perceived immediately. To establish that only a single X lies on the closed curve, one can use the above strategy of marking the X and tracking the curve. When the tracing is completed, we have reached the same X, as opposed to a second one. The problem cannot be solved *by pre-existing detectors specialized for the task.* Ullman suggests that the perception of the X on the curve involved the application of visual routines that *employ operations* such as marking and tracing.[20]

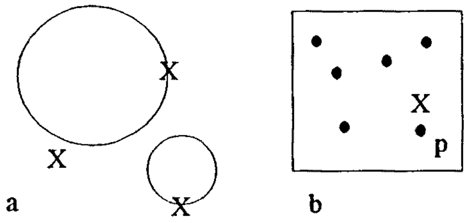


Figure 4.4. Marking. The task in *a* is to determine visually whether there are two X's on a common curve. The task could be accomplished by employing marking and tracing operations. *b*.The use of an external reference: the position of a point *p* can be defined and retained relative to the predominant X nearby (From Ullman 1996, p.306).

4.1.2.5.Integration of elementary operations. For Ullman, then, such "spatial-relation tasks", which seem to be achieved without effort, are in fact the product of complex processes, that do not take into account the specific problem at hand, such as inside/outside,

[20] Ullman suggests that the task of *counting* can be subserved by a mechanism of this type. Counting, for him, could be performed not by a special mechanism but by simple visual routines that employ elementary operations such as shifting and marking. It could be achieved by shifting the processing focus among the items of interest without scanning the entire image systematically. This counting is interesting for the rest of empirical studies we are examining.

but that are obtained through an indirect strategy. That is, the processes that account for the tasks don't *draw upon the information* about inside/outside relations, even though they are used to account for such tasks. The operations examined, such as shift, bounded activation, boundary tracing and marking are some of the different elementary operations that can be used in visual routines. For example, suppose that a scene contains several objects, such as a man in a car at one location, and a dog at another. We first *locate* the objects that we are interested in, such as the man-figure, then we undertake the visual *analysis* of it, and following that we *shift* our gaze and processing focus to the dog. The visual analysis of the man-figure is maintained in the incremental representation, and this information is still available at least in part as the gaze is shifted to the dog. In addition to this information we keep a *spatial map,* a set of spatial pointers, which tell us that the dog is at one direction, and the man at another, that is, we *mark* the map. Although we no longer see the man clearly, we have a clear notion of what exists where. The "what", is supplied by the incremental representations, and the "where" by the marking map.

In such a scheme, we do not maintain a full panoramic representation of the scene. After looking at various parts of the scene, our representation of it will have the following structure. There would be a retinotopic representation of the scene in the current viewing direction. To this representation we can apply visual routines to analyze the properties of, and relations among, the items in view, such as analysing whether the man is inside the car. In addition, we would have markers to the spatial locations of items in the scene already analyzed. These markers can point to peripheral objects, and perhaps even to locations outside the field of view. If we are currently looking at the dog, we should see it in fine detail, and therefore we will be able to apply visual routines and extract information regarding the dog's shape. At the same time we know the locations of the other objects in the scene (from the marking map) and what they are (from the incremental representation). To obtain new information, we would have to shift our gaze back to the man-figure and apply additional visual routines. This implies that the analysis and the extraction of information has to be made on the fly.

But then, this brings up the problem of how do we assembly visual routines. How are routines constructed in response to specific goals? How is such a process controlled? In all the examples, the goal was set up externally, and an appropriate routine was applied.

Sometimes, routines are invoked in response to internally generated goals. These stored visual routines can constitute what Ullman calls "perceptual programs", analogous to stored motor programs for executing movements. This pre-stored set of routines can improve the performance and the speed-up of familiar perceptual tasks.

The visual processor described by Ullman has thus three properties: i) the capacity to establish abstract properties and relations , ii) the capacity to establish a large variety of relations and properties, including newly defined ones, and iii) the requirement to cope efficiently with the complexity involved in the computation of spatial relations, namely:

Abstractness. The satisfaction of inside/outside problems provides an example of the visual system's capacity to analyze abstract spatial relations. As a matter of fact, the concept of being "inside" is abstract, because it does not refer to any particular shape, but can appear in many different forms. More formally, a shape property *P* defines a set *S* of shapes that share this property. The property of closure, for instance, divides the set of all curves into the set of closed curves that share this property, and the complementary set of open curves. Similarly, a relation such as "inside" defines a set of configurations that satisfy this relation. Clearly, in many cases the set of shapes *S* that satisfy a property *P* can be large and unwieldy. It therefore becomes impossible to test a shape for property *P* by simply contrasting it against all the members of *S* stored in memory.

Abstract shape properties and spatial relations are properties and relations with a large set of elements that can nevertheless be established efficiently by a computation that captures the regularities in the set. Closure, for instance, is an abstract property because it must be established by some process that makes use of general characteristics of closed curves. Our visual system can clearly establish abstract properties and relations of this type. The implication is that it should employ sets of processes for establishing shape properties and spatial relations. The perception of abstract properties such as insidness or closure would then be explained in terms of the computations employed by the visual system to capture the regularities underlying different properties and relations. These computations would be described in terms of their constituent operations and how they are combined to establish different properties and relations.

Open-Endedness. Our visual processor is able to solve not only one, but a large number of different properties and relations' tasks. It is a basic tenet of Ullman's that this implies that the computations that establish different properties and relations share their underlying fundamental operations. In this manner a large variety of abstract shape properties and spatial relations can be established by different processes assembled from a fixed set of operations. Given such possibilities, Ullman uses the notion of "visual routines", which refer in fact to the processes of elementary operations that establish shape properties and spatial relations.

Moreover, a mechanism is required by which new combinations of basic operations can be assembled to meet new computational goals. One can impose goals for visual analysis, such as "determine whether the green and red elements lie on the same side of the vertical line". That the visual system can cope effectively with such goals suggests that it has the capacity to create new processes out of the basic set of fundamental operations.

Complexity. The complexity of basic operations amounts to the use of the same mechanism by different routines that establish different properties and relations. A special case of the complexity consideration concerns, for instance, the need to apply the same computation to different spatial locations. The ability to perform a given computation at different spatial positions can be obtained by having an independent processing module in each location. In this sense, the orientation of a line segment at a given location may be determined in the primary visual cortex and be largely independent of other locations. In contrast, the computations of more complex relations such as inside/outside independent of location cannot be explained, according to Ullman, by assuming a large number of independent "inside/outside modules", one at each location. Routines that establish a given property or relation at different positions are likely to share some of their machinery in the same way as elementary operations are shared by different routines.

Ullman argues that the sharing of basic operations imposes certain constraints upon the computation of spatial relations. For example, the sharing of operations by different routines will restrict the simultaneous perception of different spatial relations. The application of a given routine to different spatial locations will be similarly restricted. In applying visual routines the need will consequently arise for the sequencing of elementary operations as well as the need to selecting the location at which a given operations is applied. In summary, the requirements discussed above suggest, for Ullman:

> (1) Spatial properties and relations are established by the application of visual routines to a set of early visual representations.
>
> (2) Visual routines are assembled form a fixed set of elementary operations.
>
> (3) New routines can be assembled to meet newly specified processing goals.
>
> (4) Different routines share elementary operations.
>
> (5) A routine can be applied to different spatial locations. The processes that perform the same routine at different locations are not independent.
>
> (6) In applying visual routines mechanisms are required for sequencing elementary operations and for selecting the locations at which they are applied.

4.2.Arithmetical cognition

The area of cognitive or mental arithmetic asks the question "how do people do arithmetic in their head?", that is , how does a person who has studied and used arithmetic (and mathematics) in school accomplish routine acts such as adding or multiplying single-digit numbers? Specifically, cognitive psychology is interested in how a person's knowledge of numbers and mathematics is organized in memory, as well as how this knowledge is accessed and applied in various settings? I review here a specific proposal about cognitive arithmetic, that has been presented by Stanilas Dehaene (1992, 1997).

4.2.1.Task

The word arithmetic has several senses, but one of the most common of them is the theory of the natural numbers and of similar systems of numbers, such as the integers. Arithmetic can be considered a formal axiomatic theory. Therefore, we could say that the competence in arithmetic corresponds to an ability that accords with such axioms. The simplest of all formalizations of arithmetic was provided by Peano's axioms. Succinctly, these are understood as follows:

i) 1 is a number.

ii) Every number has a successor, denoted as *Sn* or simply as $n + 1$.

iii) Every number but 1 has a predecessor (assuming that we consider only the positive integers).

iv) Two different numbers cannot have the same successor.

v) If a property is verified for number 1, and if the fact that it is verified for n implies that it is also verified for its successor $n + 1$, then the property is true of any number n.

Additionally, the ability in arithmetic would have to accord with some of the fundamental laws of the domain that can be derived from this formalization, such as the laws of commutativity and associativity for addition and multiplication as well as the distributive law, namely:

1) The commutative law of addition: $a + b = b + a$

2) The associative law of addition: $a + (b + c) = (a + b) + c$

3) The commutative law of multiplication: $ab = ba$

4) The associative law of multiplication: $a(bc) = (ab)c$

5) The distributive law: $(a + b)c = ac + bc$

6) $a^m a^n = a^{m+n}$

7) $(a^m)^n = a^{mn}$

8) $a^m b^m = (ab)^m$

9) $a^m / a^n = a^{m-n}$ $m > n$

Furthermore, it would have to accord with the precise notions of arithmetical theory such as counting, with all the properties attached to that notion, such as, for example, that it should be the unique method to quantify the objects of a set.

4.2.2.Cognition

It would seem justified to believe that the cognitive mechanisms used to satisfy our ability in arithmetic corresponded with a unified set of rules and facts that could be applied uniformly in different contexts. However, things seem to be a more complex. Dehaene suggests a number of basic mechanisms and processes used to solve arithmetic problems.

4.2.2.1.Mechanism: The "Accumulator" or Analogue mode. Dehaene affirms that there are two largely distinct number-processing pathways. One processes numbers as symbols and the other transduces them into approximate quantities. The latter, which Dehaene calls "approximate mode" (and which is also known as "accumulator"[21]), is a domain where tasks such as measurement, comparison of prices or approximate calculations are calculated in a fashion similar to a mental "number line". This mode is an *analogue* encoding. According to Dehaene, many animals, such as rats and pigeons, use this analogue mode to make calculations. Dehaene even proposes it to be a sort of "number sense" that provides a direct intuition of what numbers mean. Dehaene suggests that digits are not compared at a symbolic level but are initially recorded and compared as quantities. The animal brain works its minor arithmetical operations with this mechanism. This is a mechanism that can only handle continuous estimates, rather than discrete quantities. It is fundamentally imprecise and seems unable to precisely keep track of the items that it counts. Pigeons, for instance, cannot distinguish 49 from 50, because they cannot represent these quantities other than in an approximated and variable fashion. This means that for an animal 5 plus 5 does not make 10, but only *about 10*. The mechanism is a "fuzzy calculator". Therefore, animals seem to be able to perceive "numbers", that is, numerical quantities, though provided its fuzziness,

[21] The origin of this idea is usually attributed to Meck and Church (1983). They suggest that a single mechanism underlies both animals' ability to determine some sort of number sense, and their ability to measure duration. Basically, their proposed mechanism works as follows: a pacemaker puts out pulses at a constant rate, which can be passed into an accumulator by the closing of a mode switch. In its counting mode, every time an entity is experienced that is to be counted, the mode switch closes for a fixed interval, passing energy into the accumulator. Thus the accumulator fills up in equal increments one fore each entity counted. In its timing mode, the switch remains closed for the duration of the temporal interval, passing energy into the accumulator continuously at a constant rate. The mechanism contains several accumulators and switches, so that the animal can count different sets of events and measure several durations simultaneously. The final value in the accumulator can be passed into working memory, and there compared with previously stored accumulator values.

the notion of "numerosity"[22] perception is usually preferred to that of "number".

Dehaene extends the availability of this mechanism to humans. His hypothesis is that humans are endowed -since they are born- with a mental representation of quantities very similar to the one that can be found in rats, pigeons, or monkeys. Like them, humans are able to rapidly enumerate collections of visual or auditory objects, to add them, a and to compare their numerosities. It has shown in different experiments that babies of even six months of age are sensitive to the numerosity of auditory and visual sets (see Wynn 1995). Dehaene speculates that these abilities not only enable humans to work out the numerosity of sets, but also underlie their comprehension of symbolic numerals such as Arabic digits. In essence, the "number sense" that humans inherit from evolutionary history plays the role of a tool that favours the emergence of more advanced mathematical abilities.

Moreover, Dehaene contends that even if adults can use the symbols of, for instance, Arabic code, they cannot dispense with the analogue code. According to Dehaene each time humans are confronted with an Arabic numeral, the brain cannot but treat it is as an analogical quantity and represent it mentally with decreasing precision. To enter this putative approximation mode, Arabic and verbal numerals are first translated from their digital or verbal code into a quantity code. This encoding stage is fast, unconscious and independent of which particular number is coded. The input modality is then neglected and numerical quantities are represented and processed in the same way as other physical magnitudes like size or weight. The translation allows to retrieve immediately the meaning of a symbol such as 8 which, according to Dehaene, specifies that 8 is a quantity between 7 and 9, closer to 10 than to 2, and so on. This translation from symbols to quantities imposes an important and measurable cost to the speed of our mental operations. Dehaene nevertheless holds that the same numeral may also be processed via the traditional symbolic transcoding and calculation routines.

There is a widely known effect that supports the hypothesis that mental representations of numbers are, at least, entertained as quantities: the number-comparison task. This tasks shows that the time to decide which of two numbers is the larger (or the

[22] Numerosity is a notion used to escape the ambiguity of the word "number", and it is used to refer specifically to a measurable numerical quantity.

smaller) smoothly decreases with the numerical distance between them. This *distance effect*[23] is identical whether the comparison bears on Arabic numerals or whether it bears on physical parameters such as line length, pitch or numerosity. In all cases, the response time is a logarithmic function of the distance (numerical or physical) between the items, and similar accord or congruity effects are found. Even in same-different judgements, a distance effect emerges. The time needed to judge that two digits are different varies with the numerical distance between them. Nevertheless, the speed with which humans compare two Arabic numerals does not depend solely on the distance between them, but also on their size. The number line, the mental representation of numbers in the analogue medium, depicts quantities in a fashion not unlike the logarithmic scale on a slide rule, where equal space is allocated to the interval between 1 and 2, between 2 and 4, or between 4 and 8. As a result, the accuracy and speed with which calculations can be performed necessarily decreases as the number gets larger.

This analogue medium has other curious properties. In a task where subjects were asked to judge the parity (odd vs. even) of numbers from 0 to 9, Dehaene found that the assignment of "odd" and "even" responses showed an interaction of number magnitude with response key. Regardless of their parity, larger numbers yielded faster responses with the right hand than with the left, and the reverse was true for small numbers. This was termed by Dehaene as the SNARC effect (spatial-numerical association of response codes). Other experiments showed that the SNARC effect is governed by the *relative* magnitude of the numbers within the range of numbers tested. When numbers in the 0-5 range were used, numbers 4 and 5, the largest, elicited a faster response with the right hand than with the left. This pattern was reversed when numbers in the 4-9 range were used, where the same number 4 and 5 were now the smallest. This was interpreted as meaning that the SNARC effect seems to reveal a natural mapping of the numerical continuum onto the extracorporeal physical space. The number line would extend horizontally from left to right. This would be related to certain convention, such as that of writing from left to right. As a matter of fact, the reverse effect was found in Arabs who write from right to left. However, the existence of an analogue representation for numerical magnitude does not

[23] The origin of this idea is in Moyer and Landauer (1967).

necessarily imply the existence of a first-order isomorphism between number and any particular physical continuum.

The existence of this "analogue mode" has had consequences not only in how we perceive numbers, but obviously on how arithmetic, and mathematics in general, have appeared in our culture. Dehaene illustrates this idea with a sketch of the mathematical cultural evolution:

Evolution of oral numeration. The *starting point* was the mental representation of numerical quantities in the way we share with animals. The first *problem* was how to communicate these quantities through spoken language. The *solution* was to allow the words "one", "two" and "three" refer directly to the subitized numerosities 1, 2, and 3. The second *problem* was how to refer to numbers beyond 3. The *solution* was to impose a one-to-one correspondence with body parts (for example, there are cultures that the number 12 corresponds to the left breast). The third *problem* was how to count when the hands are busy. The *solution* was to turn the names of body parts into number names (12= "left breast"). The fourth *problem* was that there is only a limited set of body parts, compared with an infinity of numbers. The *solution* was to invent number syntax (12="two hands and two fingers"). The final *problem* was how to refer to approximate quantities. The *solution* was to select a set of "round numbers" and invent the two-word construction (e.g. ten or twelve people).

Evolution of written numeration. The first *problem* was how to keep a permanent trace of numerosities. The *solution* was one-to-one correspondence. Engrave notches on bone, wood, and so on (7=IIIIIII). The second *problem* was that such a representation was hard to read. The *solution* was to regroup the notches (7=~~IIII~~ II), and replace some of these groups with a single symbol (7=VII). The third *problem* was that large numbers still required many symbols (e.g., 37=XXXVII). The *solution* was to denote numbers using a combination of multiplication and addition (345= 3 hundreds, 4 tens, and 5). The fourth *problem* was that this notation suffered from the repetition of the words "hundreds" and "tens". The *solution* was to drop these words, resulting in a shorter notation ancestral to modern place-value notation (437= 4 3 7). The final *problem* was that this notation is ambiguous when units of a certain rank are lacking (407, denoted as 4 7, is easily confused with 47). The final *solution* was to invent a placeholder, the symbol zero.

4.2.2.2.Mechanism: The Symbolic Code. Dehaene understands symbol systems as a supplementary tool that humans avail themselves to be competent in various tasks. Linguistic symbols parse the world into discrete categories, and allow humans to move beyond the limits of approximation. Hence, humans are able to refer to precise numbers and to separate them categorically from their closest neighbours. Language symbols such as the Arabic numerals can label and discretize any continuous quantity. With symbols we can discriminate 8 from 9, which is not possible using the Analogue Mode. Furthermore, thanks so such a tool we can develop formal rules for comparing, adding or dividing two numbers. In Dehaene words, the scaffolding of mathematics can then rise.

4.2.2.3.Mechanism: Creative intelligence. It is a truism that children and adults often err in the most elementary of calculations. According to Dehaene mental calculation is very difficult because poses unnatural problem for the human brain. An innate sense of approximate numerical quantities may well be embedded in genes, but when faced with exact symbolic calculation, humans lack proper resources. Yet, humans find ways to face arithmetic problems. It seems that in the first six years of life, a profusion of calculation algorithms see the light (Siegler and Jenkins 1989). In Dehaene's terms, children reinvent arithmetic. Spontaneously or by imitating their peers, they imagine new strategies for calculation. They also learn to select the best strategy for each problem. The majority of their strategies are based on counting (see below) with or without words, with or without fingers. Children often discover them by themselves, even before they are taught to calculate. By the fourth year of age, children have mastered the basics of how to count. The first calculation algorithm that all children figure out for themselves consists in adding two sets by counting them both on the fingers. Most children realize by themselves that they need not recount both numbers, and that they can compute 2+4 by starting right with the word "two". This is called the "minimum strategy". In addition children of five spontaneously think of counting from the larger of the two numbers to be added. This indicates the they have a very precocious understanding of the commutativity of addition.

They quickly master many addition and substraction strategies, which they select for each particular problem. For the problem 8 + 4 some remember that 8+2 is 10, they manage to decompose 4 in 2+2, then they are able to simply count "ten, eleven, *twelve*." Calculation abilities, according to Dehaene, do not emerge in an immutable order. Each child behaves like a cook's apprentice who tries a random recipe, evaluates the quality of the result and decides whether or not to proceed in this direction. Siegler and Jenkins (1989) contend that children compile detailed statistics on their success rate with each algorithm. Little by little, they acquire a refined database of the strategies that are most appropriate fore each numerical problem.

4.2.2.4.Mechanism: Memory. The hypothesis that memory plays a central role in adult mental arithmetic is now generally accepted. The first articulated idea in this sense was proposed in 1978 by Ashcraft and Battaglia, who showed that young adults hardly ever solve addition and multiplication problems by counting. Instead, they generally retrieve the result from a memorized table. These tables are memorized during school years. However these tables are difficult to memorize, and therefore the brain uses all the aid it can obtain to retain them. Verbal memory is the most evident. In many countries recitation remains, for example, the prime method for teaching arithmetic.

4.2.2.5.Process: Quantification. For Dehaene, quantification consists in grasping the *numerositiy* of a perceived set and accessing the corresponding mental token or *numeron*. According to Dehaene, humans use three different quantification processes: *counting*, *subitizing* and *estimation*.

Counting. Counting is for Dehaene the Swiss Army knife of arithmetic, the tool that children spontaneously put to all sorts of uses. Dehaene bases his proposal about counting on that of Gelman and Gallistel (1978), which seems to be the most widely held hypothesis. In their model, counting accords with the following principles:

1) *One-to-one correspondence*: Each element of the counted set must map onto one and only one numeron.

2) *Stable order*: The numerons must be ordered and mapped in a reproducible sequence onto the items to be counted.

3) *Cardinality*: The last numeron used during a count represents a property of the entire set.

4) *Abstraction*: Counting applies to any collection of entities (from physical objects to mental constructs).

5) *Order irrelevance*: The order in which different elements of the counted set are mapped onto numerons is irrelevant to the counting process.

One important consideration of Gelman and Gallistel is that competence for counting and competence for language are largely distinct and independent. Counting is therefore accessible in principle to non-linguistic animals as well as to prelinguistic human infants.

Subitizing. In experiments with timed numerosity judgements, that is, where subjects have to guess the number of items in a set in a short interval of presentation time, counting should predict that response time should increase linearly with the numerosity of the display. However, such a pattern is found, by Dehaene, only over a limited range of numerosities, basically over the range of 4-6. For numerosities 1-3 experiments show that subjects must use another strategy other than counting. For numerosities larger than 7 the latency seems not to grow, though accuracy drops severely. This pattern suggests that counting is only used in the range of 4-6.

To account for such results the term "subitizing" was coined as a process responsible for fast responses to small numerosities, and the term "estimation" for the less accurate process used preferentially with large numerosities (Mandler and Shebo 1982). Both processes are still not widely accepted as distinct from counting. There are some proposals concerning what mechanism underlies the subitizing strategy. Mandler and Shebo argue that subitizing is a form of recognition of canonical configurations of visual items. In visual displays with a constant but small number of items, the disposition of objects necessarily forms invariant or canonical spatial configurations which may be recognized in parallel. For example [one] may be corresponded with a "dot", [two] with a "line" and [three] with a "triangle". Our visual system may recognize "threeness" in a triangular configuration, whatever the exact nature and arrangement of the constituent objects, just as it can recognize a cow regardless of viewpoint, size, colour, etc.

Estimation. The accuracy of subitizing decreases with numerosity. Number 4 seems to be the first point where subitizing starts to make a significant number of discrimination errors, confusing it with 3 or 5. In numerosities larger than 5 we use a sort of estimation process. When confronted with a crowd, humans may not know whether they are 81, 82 or 83 people, but we can estimate 80 or 100 without counting. Such approximations seem to be generally valid. Errors are made in certain conditions. For instance, humans tend to overestimate numerosity when the objects are regularly spread out, and conversely we tend to underestimate sets of irregularly distributed objects. Yet, we are normally accurate. In experiments (Ginsburg 1976) it has been shown that one single exposure to veridical numerical information, such as a set of two hundred dots dutifully labelled as such, suffices to improve our estimations of sets of between ten and four hundred dots. Our perception of large numbers follows laws that are strictly identical to those than govern subitizing: we are for example subject to a distance effect.

4.2.2.6.Process: Processing of quantities. The language faculty has given humans the ability to develop number notations, that are especially well-suited to meet calculation and communication needs. Most of adult number processing relies heavily on these notational devices, thereby explaining the predictive power of models assuming a mental "algebra" of symbol manipulations. Every linguistically literate subject can produce and understand numerals in at least two numerical notations: Arabic and verbal. With the mastery of a system such as Arabic notation comes the ability to calculate, to predict by symbolic manipulation the result of a physical regrouping or partitioning act without having to execute it.

Single digit operations. Although several models remain in competition to account for our ability to calculate single-digit additions or multiplications, all of them share the notion that in the adult, arithmetical facts, such as 2 x 2 = 4, are memorized and retrieved from a stored mental network or lexicon. It seems that evidence in cognitive arithmetic experiments fit the analogy of stored addition and multiplication tables with a lexicon (see Ashcraft 1992). For example, activation among related facts can account for the difficulty of rejecting problems like 5 x 7 = 30, where the proposed result falls in the same row or column as the correct result. The evidence suggests, moreover, that arithmetical storage is accessed automatically. For instance, it seems that multiplication and addition are initiated irrepresively from the presentation of an arithmetical problem (LeFevre et al 1988).

There are some properties of interest in such operations. The first is what is known as the *problem-size effect*. The time to solve a single-digit operation problem increases with the size of the operands. For example, computing 9 x 8 might take 200 milliseconds more than computing 2 x 2. The increase is generally non-linear: calculation time correlates well with the product of the operands, with the exception of ties (e.g., 3 + 3, 4 x 4) for which response time is constant or increases only moderately with operand size.

Multi-digit calculation procedures. Evidence seems to prove that calculation with multi-digit numerals involves the sequential combination of elementary arithmetical operations using a specific algorithm learned at school (Ashcraft and Stazyk 1981).

The evidence seems to back a clear developmental trend from early counting-based strategies in children to adult memory retrieval. One frequent strategy is the counting on procedure, in which children start with the larger of the operands and count upward as many times as is required by the smaller of the operands. However, several other strategies are available to the child: guessing, decomposing the problem (e.g., 4 + 8 = (4+ 6) + 2), retrieving the answer from memory, etc. As I have said, the use of such strategies does not follow a developmental sequence; rather, individual children typically switch between strategies from trial to trial, and which strategy is selected depends on the reliability and speed of the available strategies as measured over previous calculation trials (Siegler 1987). Progressively, memory retrieval wins over other calculation processes.

4.2.2.7.Dehaene's cognitive model. Dehaene proposes a model to account for the above findings. He calls his proposal the *triple-code model*. Dehaene's model clusters mathematical abilities into three groups according to the format in which numbers are manipulated. First, there are those abilities, like verbal counting or arithmetical fact retrieval, that are parasitic on the general spoken or written language-processing system, and which use verbal numerical notation. Addition and multiplication tables are just part of a learned lexicon of verbal associations, since the counting sequence is not different from other automatic series, e.g., the alphabet or month names.

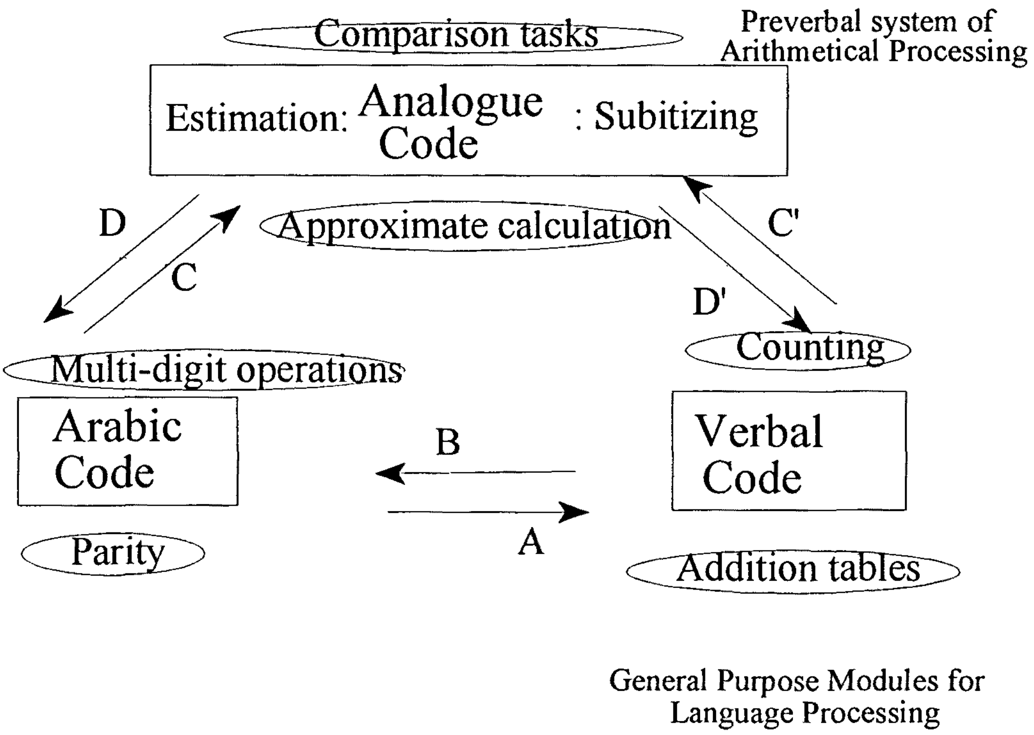

Comparison tasks
Preverbal system of Arithmetical Processing
Estimation: Analogue Code : Subitizing
D
C
Approximate calculation
C'
D'
Counting
Multi-digit operations
Arabic Code
B
A
Verbal Code
Parity
Addition tables
General Purpose Modules for Language Processing

Figure 4.5. Schematic representation of Dehaene's model. The three codes are depicted as rectangles. The analogue code accounts for estimation and subitizing abilities. Arrows indicate translation processes. Tasks which depend on each code are enclosed in circles. (From Dehaene 1992, p. 31).

On the other hand, there are abilities like multi-digit calculation or parity judgment which require the mastery of a dedicated notational system, such as Arabic notation. Such a system is clearly dependent on linguistic competence and literacy. Nevertheless, the Arabic subsystem is dedicated to numerical material and it may, according to Dehaene, be conceptualized as separate. Finally, the model considers the abilities to compare and to approximate numerical quantities as a third separate cluster. These abilities are present in animals, emerge in infants before the acquisition of language and are assumed to constitute a distinct preverbal system of arithmetic reasoning. In human adults the magnitude code would play a central role in approaching the quantity that a numeral represents and in checking the meaningfulness of calculations. The model has supporting evidence in neuropsychological patients where a dissociation between codes is observed in certain cases (e.g. Warrington 1982; Temple 1989, 1991; Temple and Vilarroya 1990; Temple, Jeeves and Vilarroya 1990; Temple, Jeeves and Vilarroya 1991).

Representation. Numbers are represented in three different codes. The auditory verbal code is created and manipulated using general-purpose language modules. In such a code an analogue of a word sequence (e.g., /six//hundred/) is what is mentally manipulated when we do arithmetic. In the visual Arabic code numbers are manipulated in Arabic format on a spatially extended representational medium. In the analogue code quantities are represented as inherently variable distributions of activation over an "analogical number line".

Each representation is directly interfaced by notation-specific input-output procedures. An Arabic numeral-reading procedure categorizes strings of digits for input into the visual Arabic representation. Conversely, an Arabic numeral-writing procedure converts the internal Arabic code into a motor program of writing gestures. Similar auditory input, spoken output, written input and written output procedures interface with the auditory verbal representation. *These procedures are not specific to numbers and also take part in the production and comprehension of oral and written language.* Finally, the analogue magnitude representation is also assumed to receive direct input from dedicated visual numerosity estimation and subitizing procedures.

There are communication between the three codes of representation, as shown in the picture with the paths A, B, C, C', D, D', for which there is empirical evidence (Cohen and Dehaene 1991).

Numerical procedures. Each number-related task can be decomposed into a sequence of component processes, each requiring a specific numerical format for input. The format in which numbers are manipulated must be assessed separately for each subcomponent of a task.

Several code-function assignments have been proposed by Dehaene. In "numerical comparison" the Arabic input is transformed into an analogue magnitude code before the comparison can be performed. "Multi-digit operations" seem to involve the mental manipulation of a spatial image of the operation in Arabic notation. Conversely, "addition and multiplication tables" are stored as verbal associations. Finally, access to "parity information" is postulated to depend exclusively on the Arabic code.

4.2.3.Conclusion

For Dehaene arithmetical knowledge is embedded in a variety of specialized abilities. Some recognize digits, for instance, and others translate them into an internal "analogue" quantity. Still others recover arithmetic facts from memory or prepare the articulatory plan that enables humans to calculate. The brain has the ability to connect these elementary processes into a useful sequence. The fact that we can take out from memory an arithmetical fact does not mean that we perform an arithmetical operation. Likewise, the algorithms that we learn in school and we use to perform arithmetical operations are a "shortcut" to accord with a given arithmetic operation. The mechanisms are a sort of *surrogate* operations of real arithmetical operations. Moreover, we could even say that the notion of *numerosity* is not possessed, but a vague notion of magnitude that we entertain in the putative analogue mode of Dehaene.

Arithmetic cognition has thus to be seen as a fractioned set of numerical abilities, among which faculties such as quantification, number transcoding, calculation or approximation may be isolated. The point is that these mechanisms are not specialized arithmetical mechanisms relying on mathematical principles that warrant functional compliance, nor are they sensitive to the principles of arithmetic. In sum, the brain has managed to use strategies to deal with calculation problems in a very accurate way despite its inability to function as an arithmetical mechanism.

4.3.Object cognition

My reference on this project will be the book *Spatial Representation* edited by Naomi Eilan, Rosaleen McCarthy and Bill Brewer (1993). Some of the contributors of this book share the view that our perceptual systems seem to be remarkably accurate in extracting and using information about the motion of objects in space. Examples of this competence include our abilities to locomote without colliding with stationary or moving objects, and to anticipate the trajectories of transforming objects in order to intercept, follow, or avoid them. These perceptual abilities are complemented by our capacity to predict the current and continuing motion of objects with a very scarce and incomplete information about the structure of objects and events. As a matter of fact, the movement of our bodies and of our eyes, as well as relationships of occlusion among objects limit the quality of information sampled. The general held idea is that our perceptual systems have evolved to take advantage of certain invariant sources of information in the environment to perform such activities. In particular, invariants in the form of descriptions of the motion of objects in the world may have been internalized and used to constrain alternatives among the multitude of possible interpretations of the partial information available. The suggestion is that there is what Cooper and Munger (1993) call a "mental physics" internalized by our perceptual systems over the course of evolution that reflects certain external physical regularities.

4.3.1. Task

In classical physics, two general types of information are used to describe the behaviour of moving objects. Kinematic information describes the pure motion of bodies without regard to mass: position, velocity and acceleration of an object. Dynamics describes, on the other hand, the forces causing movement or acting on objects with mass.

According to Peacocke (1993), we experience objects as material objects. This implies that a thinker's mental representation of objects must be suitably sensitive to all the substantial properties that objects necessarily have. Therefore, for Peacocke a constitutive account of what it is to have the notion of object requires the capacity to employ particular kinds of physical principles that spatial objects accord with. Peacocke asserts that normal mature subjects must employ, or have a grasp of, dynamical principles, rather than kinematic ones. Peacocke argues that part of what is involved in having such an experience is that perceptual representations must employ the notion of force and mass. Specifically, given the claim that the link with force is an essential property of space-occupying matter, and that physical objects are, at least, groupings of space-occupying matter, this means that a subject's mental representation of physical objects must be 'suitably sensitive' to the link with force and mass. This is because, among other reasons, for something to be a quantity of matter is for changes in its state of motion to be explicable by the mechanical forces actioning on it, as well as for its changes of motion to exert such forces. Therefore, in order to qualify as being able to reason about objects, we must *attribute the capacity to reason according to dynamic principles.*

4.3.2.Cognition

4.3.2.1.Object perception. Spelke and Van de Walle (1993) aim at identifying the way in which infants (and by extension that of adults) segregate a visual scene into units as well as predict the movements of those units, which should provide the core of the notion of physical objects.

Empirical evidence support the idea that infants perceive object unity as early as three months of age (e.g. Kestenbaum, Termine and Spelke 1987), and that they can also perceive the distinctness of objects under two conditions. First, young infants perceive two objects as distinct if the objects are spatially separated by a gap. Two objects are perceived as distinct not only when the objects are separated vertically, so that the gap is visible, but also when the objects are separated in depth, so that the gap cannot be seen directly. Second, young infants perceive two objects as distinct if the objects undergo separated motions even if the objects remain in contact throughout the time that they move. Two adjacent objects are perceived as distinct if one object moves while the other is stationary, and also if each object moves rigidly in a different direction. In contrast, infants do not appear to perceive the boundary between two objects that are both stationary and adjacent, even if the objects differ in colour, texture and form. The Gestalt relationships that specify the boundaries of stationary objects for adults -colour similarity, smoothness of edges and figural goodness- do not appear to be effective for infants.

Let us turn now to the case in which two objects undergo a common rigid motion (e.g. Spelke et al 1989). When two objects are adjacent and move together, infants appear to perceive one connected body. When two objects are separated in depth, however, a different finding is obtained. Although the objects are perceived as distinct when they are stationary, they are perceived as a single body when they move together. In the case of partly occluded objects, when a visible object moves partly out of view, young infants appear to perceive a persisting, unchanging body. Children infer that the visible surfaces that occlude an object that moves rigidly behind them are connected to it. This applies to young infants of at least three months. In newborns, on the other hand, it is not found. Whatever processes account for developmental changes in the first months of life, it appears that three- and four-month-old toddlers perceive visible objects in according to some characteristics that can be described as principles. For such infants, perception of visible objects evidently depends on processes that operate quite late, according to Spelke and Van de Walle, in perceptual analyses. These processes take as input representations of arrangements and motions of surfaces in three-dimensional visible layout, not lower-level representation of arrangements and displacements of images in the two-dimensional visual field. On the other hand, such a process seem not to depend on distinct visual mechanisms, haptic mechanisms and the like, but it depends on a single mechanism operating on representations of the layout obtained from any perceptual mode. Spelke and Van de Walle's experiments on haptic perception provide evidence that infants perceive haptically presented objects in the same way as in the visual domain. This may indicate that reasoning about objects is, according to Spelke and Van de Walle, an amodal system.

Other experiments show that infants, as adults, represent each object moving continuously over space and time as a single object even if objects enter and leave the field of view whenever the perceiver or the objects move (e.g. Baillargeon 1992). This is labelled by Spelke and Van de Walle the *continuity constraint*, according to which objects move only on connected paths from one place and time to another. Additionally, Spelke and Van de Walle have shown that infants, like adults, represent hidden objects according to another physical constraint, the *solidity constraint*, which dictates that objects move only on non-intersecting paths, such that no parts of distinct objects ever coincide in space and time.

Spelke and Van de Walle argue that their evidence point to a certain structured organization of infants perception. Specifically, infants treat a single object as a spatially connected body that retains its connectedness as it moves. When two surfaces can be seen *not* to be connected (because they are separated by a detectable gap or because their connectedness is broken as they move), the cohesion constraint dictates that the surfaces lie on different objects. Second, infants presume that distinct objects are not connected and that they do not become connected when they move. When no spatial gap or relative motion can be seen to separate two surfaces, the boundedness constraint dictates that the surfaces lie on a single object. Infants sensitivity to these two constraints can be encompassed by a single principle:

> **Principle of cohesion**: Surfaces lie on one object if and only if they are connected.

Third, when two objects are separated in depth they are perceived as distinct when they are stationary, but they are perceived as a single body when they move together. If two partly hidden surfaces move rigidly together, infants infer that surfaces are in contact somewhere out of view. This follows from the physical constraint of "no action at a distance": Distinct objects do not move together if they are separated by a gap. Conversely, if two partly hidden surfaces move independently adults infer that they are separated by a gap. This *inference follows from the physical constrain of "action on contact": surfaces move together* if and only if they are in contact. This can be captured by a principle:

> **Principle of contact**. Surfaces move together if and only if they are in contact.

Fourth, infants perceive objects travelling on a connected path over space and time without any gaps, in accord with the continuity constraint. And finally, infants reckon distinct objects travel on separate paths over space and time and therefore paths cannot intersect in such a way that the objects coincide in space at any moment in time, in accord with the solidity constraint. These two constraints can be encompassed by a single principle:

> **Principle of continuity.** An object traces exactly one connected path over space and time.

In sum, infants appear to organize the perceived layout into bodies that are cohesive, bounded, move independently of bodies from which they are spatially separated, move together with bodies with which they are in contact, and persist even if they are hidden.

Spelke and Van de Walle have also investigated whether infants consider that objects move in accord with two further physical constraints on object motion: gravity and inertia. A number of experiments were reviewed, such as the following (Spelke et al. 1992). Infants were habituated to a ball falling behind a screen that covered two surfaces and which was revealed once at rest on the first surface in its path. When the upper surface was removed, the ball was dropped behind the screen as before, and being revealed either in a new position on the lower surface or in its former position, this time in midair. The latter position was inconsistent with gravity and inertia: the ball appeared to stop falling in the absence of support and to change its motion in the absence of obstacles. Experiments provided evidence that, according to Spelke and Van de Walle, infants are sensitive neither to gravity nor to inertia, though they confirmed that they are sensitive to the constraints of no action at a distance and action on contact.

The final proposal of Spelke and Van de Walle is that the principles that encompass object perception and physical reasoning are closely connected and possibly identical. For these researchers, there is a single system of knowledge that guides perceiving and reasoning about objects not only in infants but also in adults. There appears to be a general process of object perception in infancy applicable to all material bodies. Infants perceive the unity and boundaries of objects by analysing the spatial arrangements and motions of surfaces in accord with the principles of cohesion and contact. There also appears to be a general process of physical reasoning in infancy. Infants reason about the behaviour of material bodies in accordance with the principles of contact and continuity. There is, according to Spelke and Van de Walle, a single conception of material bodies. Additionally, they suggest that the principles guiding perception and reasoning in infancy are central to the physical conceptions in adults. They describe spatial development as a primitive single system of knowledge that will be enriched, not overturned, by further knowledge. Children may learn, for example, that material bodies tend to have simple shapes, to move smoothly and to fall when unsupported. But adults will nevertheless keep the centrality of the principles by singling out objects, and even if they learn about gravity and inertia they might have such a sensitivity incompletely represented, as we will see below.

4.3.2.2.Perception of object motion. Cooper and Munger (1993) set up their objective to determine the second considerations made above, namely, which aspects of the regular behaviour of moving objects might constitute external invariants that would be most useful to internalize. They are interested in the question of what physical principles the perceptual system may have internalized by examining and whether observers are sensitive to and able to use these kinematic and dynamic sources of information in making perceptual judgments about ongoing visual events.

Experiments reported on before Cooper and Munger (e.g. Gilden and Proffitt 1989) are consistent with the idea that constraints of kinematic geometry are internalized and arise from the paths of apparent motion experienced between two successive views of an object differing in position and orientation. Likewise, in other experiments (McCloskey and Kohl 1983) it is shown that the majority of subjects are able to make trajectory judgements accurately. Finally, observers seem also to be able to determine accurately whether or not collisions are natural, and they can also use information about the relative velocities of objects after impact to assess their relative masses. It would appear that kinetic or dynamic information can be computed from full perceptual exposure to visual events.

Cooper and Munger devise certain situations to test these provisions. One case provides successive, discrete views of an object undergoing a transformation that must be inferred from the pattern of presentations. Observers must judge whether the queried position is the same as the final presented view. The second case provides the subject with a continuous, animated display of an object transforming in some well-defined fashion. At some point in the transformation, the object disappears momentarily and then reappears, continuing to transform in the manner depicted before the disappearance. Observers must judge whether the point of reappearance is at the correct location on the transformational trajectory, while the transformation continues "unseen" during the disappearance interval. For both types of stimulus situations, Cooper and Munger pose a number of questions:

(1) How accurate are subjects in predicting or in remembering the locations of moving objects?

(2) How are judgements affected by manipulation of kinematic and/or kinetic properties of the depicted events? (as evidenced by the extraction and uses of such

sources of information)

(3) How might we account for overall patterns of performance on these tasks? In particular, under what sets of conditions will the analogy with the physical motion of objects provide a useful conceptual framework?

The first case is the inference of trajectory of object motion from successively presented static views. The evidence so far is, as mentioned above, that performance is generally quite accurate. There is however a slight distortion for the position of an object in the direction of its implied motion. Such a distortion has been termed "representational momentum", in reference to the possibility that the perceptual system might embody a principle analogous to physical momentum.

Cooper and Munger examined the contribution of a source of dynamic information -depicted by the mass of an object- to the magnitude of distortion effects in the implied motion paradigm. The prediction was that if such distortions result from a cognitive analogue to physical momentum, then variations in apparent mass should influence the size of the experimental effect in a manner similar to variations in implied velocity. Cooper and Munger successfully replicated the finding of a systematic distortion in memory for the final position in the sequence. However, they found that objects differing in perceived mass do not result in memory distortions of differing magnitudes. The results of this experiment fail to support one particular consequence of the notion that the representation of a moving object has internalized momentum.

In general, the results question the usefulness of models or analogies based on dynamics for explaining the highly reliable distortions in memory for the position of an object implying directional motion. According to the analysis offered in classical physics, an object with greater mass should exhibit more momentum than an object with smaller mass, because more force is required to change the current pattern of motion. In addition, more streamlined objects should take more time to stop when moving through the same medium. In Cooper and Munger's experiments, neither rated object masses nor object shapes affected the magnitude of the momentum effect or directional memory distortion, though they have replicated the basic error in memory for position. Cooper and Munger add that the only factors that appear to change the *size* of this effect in a principled way are kinematic ones, that is, variations in the inferred velocity or change in velocity of the sequence of static views. This suggest that the system producing the error is sensitive to parameters of object motion, or kinematic information, but not to factors associated with the causes of that motion or dynamic information. A number of additional findings made by Cooper and Munger suggest that the momentum analogy is misleading in accounting for the phenomenon. In other experiments they report memory errors for an object's position which is determined by the global structure of the event implied in a sequence of static views, rather than by local characteristics of the implied motion in views immediately preceding the test display.

The second experimental paradigm that Cooper and Munger undertake are the studies in which observers are asked to predict, rather than to remember, the position of a transforming object. The procedure requires observers to extrapolate or to generate predictions concerning the continuing appearance of a transforming object that is momentarily obscured from view. Results of their experimental set up show that, first, overall performance when the point of reappearance is shifted away from the correct location becomes increasingly accurate as the angular size of the displacement increases, in either the forward or backward direction. Second, there is a marked asymmetry in the function relating percentage of 'correct reappearance' responses to test object position. Specifically, cases in which the object reappears *before* the correct position (termed *undershoots*) are uniformly less detectable than in cases in which the object reappears in a position *beyond* the correct location (termed *overshoots*). Third, the average percentage of 'correct reappearance' judgements is greater to small undershoots than to objects presented in the objectively correct position. In other words, the peak of the response curve is shifted in the direction of undershoot or, what Cooper and Munger call, 'backward' errors.

Cooper and Munger hold that the general picture that emerges from these experiments is similar to the findings for memory errors for the final position of a sequence of static views. That is, kinematic factors seem to affect the magnitude of the extrapolation error, whereas dynamic variables have little or no influence on the extent of undershoot. Stimulus variations corresponding to dynamic factors that have been assessed are object *size* and *mass*. Object size has no effect on the extent of undershoot, nor did size interact with other factors that influenced this error. Apparent mass of shaded triangular and rectangular prisms also fails to change the pattern of results, despite additional slight modifications on the experimental situation. In summary, variation in object characteristics that parallel changes in dynamic factors have no influence on the magnitude of extrapolation errors, but there is some evidence that variations in kinematic information do affect the accuracy of the position-prediction judgement.

In sum, it seems that people reliably make systematic, directional errors when asked to remember the final position of an object inferred to be moving or to extrapolate the trajectory of an object movement when a continuous transformation is momentarily interrupted. Analogies to dynamics or the forces causing the motion of objects may not provide a useful framework for characterizing the basic, unelaborated principles internalized by our cognitive systems. However, we have to accommodate these findings with the competence to be attributed, that is, a full-blooded notion of object which requires a sensitivity to the notions of mass and force.

4.4.Artificial Life

Artificial Life (henceforth AL), or Bottom-Up AI, Animat approach, Behaviour-based AI, Animal Robotics, as it is also known identifies a loose network of engineering that shares the common goal of understanding intelligent behaviour through the construction of artificial systems. The researchers also share a number of assumptions and hypotheses about the nature of intelligence. I will succinctly present them as conceived by Steels (1994). First, AL is interested in the behaviour of an individual or a group of individuals, focusing on what makes behaviour intelligent and adaptive and how it may emerge. The main emphasis is not on the physical basis of behaviour, but on the principles that can be formulated at the behavioural level itself. An example of a theory at the behavioural level is one that explains the formation of paths in an ant society in terms of a set of behavioural rules without reference to how they are neurophysiologically implemented. Given this emphasis on behaviour, the term *behaviour-oriented* is accepted as the way to appropriately distinguish the field, particularly from the more knowledge-oriented approach of classical Artificial Intelligence (AI).

Second, AL researchers contrast their methodological strategy with that of traditional AI. The latter consists basically in constructing computational models, namely, models based on process-oriented descriptions in terms of a set of data structures and algorithms. When the description is executed, that is, when the algorithm is carried out causing the contents of the data structure to be modified over time, phenomena can be observed in the form of regularities in the contents of the data structures. When this shows a strong correspondence with natural phenomena, it is called simulation. Conversely, AL's methodology consists in constructing a physical device whose physical behaviour gives rise to phenomena comparable to the natural phenomena in similar circumstances. The device will have components with a particular structure and functioning that have been put together in a particular way. The methodological steps are as follows: A phenomenon is identified (e.g. obstacle avoidance behaviour), and artificial system is constructed that has this competence, the artificial system is made to operate in the environment, the resulting phenomena are recorded, and these recordings are compared with the original phenomena. Potential misfits feed back into a redesign or reengineering with the original phenomena. The fundamental difference between both strategies is that when constructing a simulation, one selects certain aspects of the real world that are carried over in the virtual world. However, this selection may ignore or overlook essential characteristics that play a role "unknown" to the researcher.

Finally, AL is strongly oriented towards biology. Intelligence is seen as a biological characteristic in AL, in contrast with traditional AI where intelligence is defined in terms of knowledge: a system is intelligent if it maximally applies the knowledge that it has (Newell 1982). AL's notion of intelligence is: behaviour is intelligent if it maximizes the preservation of the system in its environment. This is complemented by the idea that a system is capable of adapting and learning if it changes its behaviour so as to continue maximizing its intelligence, even if the environment changes. The biological orientation also shows up in a focus on the problem of how complexity can emerge. Behaviour-oriented AI studies the origin of complexity at different levels: from components and complete agents to multiagent systems. Systems at each level maximize their self-preservation by adapting their behaviour so that it comes closer to optimal. Different dynamics, such as cooperation, competition, selection, hierarchy and reinforcement have been identified as crucial for the emergence of biological complexity.

4.4.1.Task

The level of the task is termed as "functionality" by Steels. A functionality for Steels is something that the agent needs to achieve, for example, locomotion, recharging, avoiding obstacles, finding the charging station, performing a measurement, or signalling another agent. Steels explicitly acknowledges the terms of task, goal or competence as synonyms of functionality.

4.4.2.Cognition

The basic units of investigation in AL are termed as "behaviour systems" (task-achieving module or schema being synonyms). In order to understand this notion we have to introduce the technical definition of *behaviour* and those of *mechanism* and *component* that Steels present. *Behaviour* is understood as a regularity in the interaction dynamics between an agent and its environment, for example, maintaining a bounded distance from the wall, or having a continuous location change in a particular direction. One or more behaviours contribute to the realization of a particular functionality. Behaviours belong, according to Steels, to the vocabulary of the observer. By looking at the same agent in the same environment, it is possible to categorize the behaviour in different ways. A *mechanism* is a principle or technique for establishing a particular behaviour, for example, a particular coupling between sensing and acting, the use of a map, or an associative learning mechanism. A *component* is a physical structure or process that is used to implement a mechanism. Examples of components are body parts, sensors and software.

A behaviour system is, for Steels, the set of all mechanisms that play a role in establishing a particular behaviour. In this sense, for instance, we could have a "homing in" functionality achieved by a "zigzag behaviour" toward a goal location that is the result of a "phototaxis mechanism". Steels argues that there is not a simple one-to-one relation between a certain functionality, a given behaviour, and a set of mechanisms achieving the behaviour. Additionally, behaviour systems form a real unit in the same way that a society forms a real unit. The interaction between the different mechanisms and the success in the behaviour to achieve tasks that contribute to the agent's self-preservation give a positive enforcement to all the elements forming part of a behaviour system.

4.4.2.1.Emergence. Steels contends that behavioural emergence can be defined from two points of view: that of the observer and that of the components of the system. From the point of view of an observer, Steels calls a sequence of events a behaviour if a certain regularity becomes apparent. This regularity is expressed in certain observational categories, for example, speed, distance to walls, and changes in energy level. In this sense, a behaviour is emergent if it can only be defined using descriptive categories that are not necessary to describe the behaviour of the constituent components. Moreover, an emergent behaviour leads to emergent functionality if the behaviour contributes to the system's self-preservation and if the system can build further upon it.

Emergence can also be defined from the viewpoint of the components implicated in the emergent behavior. In order to do this we have to distinguish between controlled and uncontrolled variables. A controlled variable is directly influenced by a system; for instance, a robot can directly control its forward speed, although maybe not with full accuracy. An uncontrolled variable changes due to actions of the system, but the system cannot directly influence it, only through a side effect of its *actions*. For example, a robot cannot directly control its distance to the wall; it can only change its direction of movement, which will then indirectly change the distance. We must also distinguish visible variables from invisible variables. A visible variable is a characteristic of the environment that, through a sensor, has a causal impact on the internal structures and processes, and thus on behaviour. For example, a robot may have a sensor that measures distance directly. Distance would then be a visible variable for this robot. An invisible variable is a characteristic of the environment, which we as observers can measure, but the system has no way to sense it, nor does it play a role in the component implicated in the emergent behaviour. For example, the robot could just as well not have a sensor to measure distance. Then for a behaviour to be emergent in this second sense, regularities must involve uncontrolled variables. A stricter requirement is that the behaviour involves only invisible variables. Therefore, when a behaviour is emergent, we should find that none of the components is directly sensitive to the regularities exhibited by the behaviour and that no component is able to control its appearance directly.[24]

The particularity of emergent behaviour can be seen in contrast with the more traditional AI strategy. According to Steels, agents can satisfy functionalities in two ways. First, a designer can identify a functionality that the agent needs to achieve, investigate possible behaviours that could realize the functionality and then introduce various mechanisms that sometimes give rise to the behaviour. Second, existing behaviour systems in interaction with each other and the environment can show side effects, which is what, according to Steels, gives rise to emergent behaviour. Side effects are in fact the most basic form of emergent behaviour. This behaviour may sometimes yield new useful capabilities for the agent, in which case Steels speaks of *emergent functionality*. The fact is that in engineering, increased complexity through side effects is usually regarded as negative and avoided. Yet, Steels argues that in nature, this form of complexity is preferred. For an agent operating independently in the world, it has advantages because less intervention from a designing agency is needed. Emergent functionality has, nevertheless, disadvantages because it is less predictable.

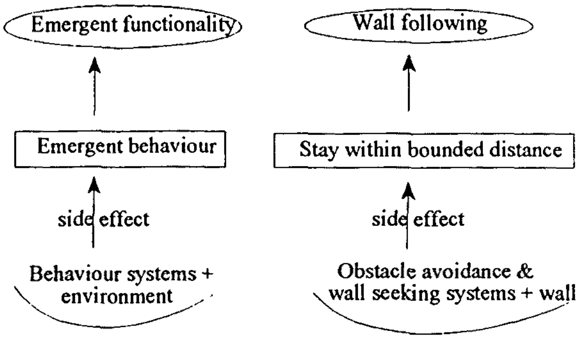

Emergent functionality
Emergent behaviour
side effect
Behaviour systems + environment
Wall following
Stay within bounded distance
side effect
Obstacle avoidance & wall seeking systems + wall

Figure 4.6. Emergent behaviour occurs as a side effect of the interaction between behaviours and the environment. New descriptive categories are needed to describe it.

Examples of emergent behaviour. Steels presents work done in his lab which can be enlightening for our discussion. The task, or functionality, that the experimenters want their robot to achieve is that of "wall

[24] This is related to idea of *self-organizing systems*. Such systems are described as systems in which a high-level pattern emerges from the interactions of multiple (simple) components without the benefit of a controller or a leader, and sensitive only to the local causal stimuli. The observed patterns of the system are explained by the collective behaviour (under specified conditions) of a large assembly of simple components, none of which plays a leading role. It is true that in these systems the actions of the parts cause the overall behaviour and that the overall behaviour guides the action of the parts. This is called sometimes "circular causation". Such a phenomena relies on what is known as *collective variables*, variables that are fixed on higher-level features, but which are not trackable on properties of simple components. In other words, the causal properties of the mechanism accounting for the emergent property need not comply with the description.

following". The behavioural regularity needed for this task is to have a bounded distance between the agent and the wall. This regularity can be attained in a directly controlled, nonemergent, way by measuring the distance and using feedback control to steer away or toward the wall. Provided that the distance is required to describe the behaviour causing wall following, then distance is a visible variable. Conversely, maintaining a distance from the wall can be achieved in an emergent way by the simultaneous operation of two behaviour systems. The first one achieves regular obstacle avoidance, for example, in terms of a dynamic coupling between infrared reflection and deflection of the path as described earlier. The second behaviour system exhibits wall seeking. This behaviour system maintains an internal variable *c*, which reflects what could be described as "motivation of making contact with the left wall." The variable *c* decreases to 0 when contact is made with the left wall (sensed by infrared reflection) and moves up otherwise; it influences the deflection of the forward motion path toward the wall. The higher is *c*, the stronger the deflection. The two behaviour systems together implement an attraction and repulsion behaviour that added up and in the presence of a (left) wall gives the desired (left) wall-following. Such a robot will, on encountering a wall on the right, first move away (thanks to the sensor) and then quickly veer back to re-encounter the wall (thanks to the bias). The cycle will repeat and the robot will follow the wall by in effect repeatedly 'bouncing off it'. An analogous behaviour system is needed for making contact with a right wall.

The point to notice is that the behaviour of wall following here emerges out of the interaction between the robot and its environment. Wall following is emergent because the category "equidistance to the (left/right) wall" is not explicitly sensed by the robot or causally used in one of the controlling behaviour systems. In other words, the competence that as theorists we might describe is that of "following a wall", whereas the "cognition", is not programmed in any sense to "follow walls".

Steels presents still another example. He invites us to imagine a robotic agent that needs to position itself between two poles so as to recharge itself. The charging station is indicated by a light source. One solution would be to endow the robot with sensors that measure its position relative to the poles and with a subroutine that computes a trajectory between the poles. An alternative solution relies on two simple behaviour systems whose environmental interaction yield portioning between the poles as a kind of side effect. The behaviour systems are (1) a phototaxis system that yields a zigzag approach to any light source and (2) an obstacle-avoidance system that causes the robot to turn away when hits something. With these two simple systems in place, the target behaviour emerges smoothly and robustly. So long as the agent may approach the charging station from any direction, it might seem that an additional positioning behaviour is required, which makes sure that the agent enters the charging station between the two poles. However, a positioning behaviour system is not necessary. The obstacle-avoidance behaviour causes retraction and turning away when the poles are hit. Because the robot is still attracted by the light source, it will again approach the charging station but now from a new angle. After a few trials, the robot enters the charging station as desired. The positioning behavior is emergent because the position relative to the poles of the charging station is *irrelevant* to describe the behaviour of the implicated behaviour of the implicated behaviour systems (obstacle avoidance and phototaxis). There is no separate structure in the agent that is measuring position with respect to the poles and causally influences motion based on this measurement. Nevertheless, the positioning behaviour occurs reliably without any additional structure in the agent. And in fact we could describe the system to be computing two rules, rule (1) being "approach the pole", while rule (2) being "don't touch the pole", and we would have a correct description of the behaviour systems supporting counterfactuals.

4.5.Conclusion

What is the picture that emerges from this review of empirical findings? My aim has been to present some real examples from cognitive science that seem to show the following idea: The notion of psychological reality might have to accommodate possibilities that adapt poorly to the notions presented in Chapters 2 and 3, and that seem to depict explanatory scenarios that might require certain adaptations of Grandpa's explanatory framework. The basic problem we have encountered is that the capacity manifested by some cognitive systems is not "internalized" in the cognitive mechanisms that account for it. Specifically, the capacity does not surface as a knowledge structure that inherits the concepts or information with which the capacity accords. We need now to account for such a "paradox".

Chapter 5

Conceptual room for theory accordance without internalization

The intrinsic features of a mechanism constrain but
do not determine its function.
Kim Sterelny

We have seen in the last chapter that some knowledge-attribution possibilities point to the fact that there are ways in which an adequate knowledge attribution clash with Grandpa's explanatory framework. A system may accord with a certain functional description at the task level which is different from the function computed by the internal structure of the cognizer. Specifically, a system might posses a number of processes dedicated to satisfying a certain function that might have nothing to do with the notions entertained at the task level. However, the task description remains the correct competence to be attributed. Sometimes it is the fact that many processes, without a precise and unique purpose, are engaged in different ways to fulfill certain functional requirements. In these cases, it seems that the cognitive system does not accord with the "functional analysis" of the behaviour. We have then a "paradox": a system that seems to comply with a given functional analysis, framed in Grandpa's terms, that does not implement such an analysis. In other words, the system does not obtain its goals *in virtue of* executing the functional specification of the task level, but in virtue of performing a different functional specification, which in turn manages to satisfy the task requirements. In short, full-blooded functional compliance does not necessarily require functional implementation.

In this chapter I will try to make room for such a possibility. In order to succeed in this aim, I will review how cognitive capacities are attributed to a system in Grandpa's framework. This will concern the question of levels of description and analysis. I will try to argue that the "paradox" cannot be framed within levels of explanation, that is, the difference between the task description and the internal functional description goes further than a difference in levels of explanation. Moreover, both functional properties to be attributed to a cognitive system belong to *the same level* of description, that of psychological or cognitive function. Accordingly we will need to make conceptual room and give naturalistic support to account for the proposal. In short, we have to answer the following question: How is it possible that there be two different but nevertheless correct *functional descriptions* of the same cognitive phenomenon? In this chapter I will try to show how it is *conceptually* possible for a system to satisfy two characterizations of the same functional characterization. In order to do that, I will first review the way in which the levels of explanation of the classical cascade are individuated and constrained. This will allow me to introduce the notion of "cascade blocking" presented in Franks (1995), where it is argued that some explanations block the cascade, in the sense that distinct levels don't compute the same function, even though explanations at each level are roughly correct. This observation will make space for a proposal that distinguishes perspectives rather than levels. Such a space will be filled by an account adapted from a proposal of Rowlands (1997), who draws a distinction in proper functions attributions at two different levels, what he calls the "organismic and algorithmic".

5.1.Levels of explanation

The first thing I will do now is to show why the examples we have reviewed present a problem for Grandpa's explanatory framework. Specifically, we have to provide some reasonable argumentation to accept the counterintuitive idea that Grandpa's three level explanation does not smoothly exhaust the explanatory requirements of such experimental designs.

5.1.1.Individuation of the task

As has remained clear throughout the last chapter, the cognitive system can perform a capacity that can be individuated both teleologically and computationally, so that the task can be fixed under an interpretation within Grandpa's explanatory schema. Consider the task of calculating the insideness and outsideness of a given object. On the one hand, there seems to be sufficient reasons, from a computational point of view, to justify the characterization of the task as the computation of "inside/outside" relations. It has been shown how the cognitive system is sensitive to such relations, that such a task can be singled out as an activity of the system performed in a determined and efficient way, and that such computations are fundamental in many tasks of visual cognition, such as in performing figure-ground segmentation. On the other hand, from a teleological point of view, the task can be considered to be selectively important to the point of being the reason for which the apparatus subserving visual cognition is there, and therefore is justified as the function of such a mechanism.

The same is true of our arithmetic capacity, though here it is easier to see that the legitimate functional attribution is the computation of arithmetic operations. After a period of learning, either by schooling or by a self-taught instruction, we are capable of performing arithmetic operations in accordance with the principles of arithmetic. Such operations are undertaken both in our everyday activities when we buy, sell or exchange money or objects, and as well as when we have to solve arithmetical problems posed in formal terms. We can even be justified in supposing that such arithmetical operations were performed by our ancestors when they had to decide, for example, how much fruit or eggs had to be brought home to feed the members of their families. For this very same reason it might even be assumed that such operations had selective value so that their satisfaction could be considered to have a teleological value.

As for the way in which we reason about objects, the justification for the consideration of it as a basic task of the cognitive system is perhaps stronger. And this is because the concept of object is considered to be fundamental to the conceptual structure of our cognitive system. It has been argued that a constitutive account of what it is to have the notion of object requires, first, a sensitivity to the fundamental properties of objecthood.

Additionally, the use of particular kinds of physical principles is also constitutive of the capacity of normal mature subjects to reason about and predict object motions. Such a constitutive basis underlies the remarkable precision of our perceptual systems in extracting and using the motion of objects in space. Examples of this competence include, as I have mentioned in the previous chapter, our abilities to locomote without colliding with stationary or moving objects, and to anticipate the trajectories of transforming objects in order to intercept, follow or avoid them. These perceptual abilities are complemented by our capacity to predict the current and continuing motion of objects on the basis of very scarce and incomplete information about the structure of objects and events. Obviously, the more fundamental in our conceptual structure something is, the more justified it is to characterize the task in teleological terms.

Finally, our artificial life-robot examples show an emerging capacity that can be individuated as a capacity of, for instance, "charging the battery". Such a function is the unique task the artifact accomplishes and, if the robot were a biological system, the function would probably have selective value, provided that it is what allows its continuing functioning.

5.1.2.Accomplishment of the task

When we examine how the previous tasks are accomplished we see that, either computation is subserved by a variety of different submechanisms none of which computes a part or the totality of the functions, (e.g., "inside" or "outside"). Nor do any of the submechanisms draw upon such information as "insideness" or "outsideness". In fact, the mechanisms are insensitive to such notions. The mechanisms are sensitive, rather, to a variety of aspects that are engaged, according to certain contextual constraints, in a non-unique manner with other mechanisms to fulfill the task.

A clearer example is arithmetic. "Arithmetical cognition" has to be seen as a fractioned set of numerical abilities, among which faculties such as quantification, number transcoding, calculation or approximation may be isolated. These abilities are subserved by a variety of mechanisms that are engaged in different ways depending on the requirements of the task, the domain in which they have to be applied, the particular learning strategies of the individual as well as other contextual factors. The fact that the principles of arithmetic are followed has more to do with the properties of the surrogate strategies (such as those of certain operational algorithms) rather than with the nature of the arithmetical knowledge of the cognizer. As for the way we reason about objects, our cognitive system seems to be employing kinematic principles to be efficient in perceiving and reasoning about objects and their motions rather than dynamic principles. In this respect, young infants appear to single out objects, and reason about object motion, according to the principles of cohesion, contact and continuity which are principles *not essential for the conception of objecthood.* On the other hand, the way in which we reason about object motions seems to be in accord with kinematic principles. Our perceptual system seems to take advantage of physical principles that reflect particular invariances in the environment, although they correspond to purely kinematic variables, velocity and change in velocity, rather than to dynamic variables such as mass and friction.

Finally, we have seen how it is possible to construct an artifact that not only accords with some functional requirement, such as "wall following", and we have seen that such task can be subserved by a group of mechanisms, none of which is by itself sensitive to the function being performed. In sum, it is possible that the cognitive system accords with a theory at the task level that describes a correct sensitivity of the system towards the theory, though the mechanisms that account for the efficiency do not compute the function specified by the theory.

However, we are not home yet, since it could still be argued that no argument has been provided by which the mechanisms described could not be taken as algorithms, possibly of a very basic sort, but algorithms nevertheless, which can be included in an explanatory cascade without much problem. The question is to know whether a disparity in two distinct but nevertheless correct functional attributions could be accounted for in such a framework. To show where the problem lies we should take into account exactly which sort of relationship the different levels of the cascade are supposed to entertain. To do so it will be also necessary to review how levels are individuated. For this reason it will be essential to acknowledge how these versions establish the relation between levels. Indeed, even if the appeal to levels of explanation is a fairly common one, it is a much trickier question to apply it to real-life examples. McClamrock (1991) for example, acknowledges that the difficultly of determining exactly when the idealization that underlies each level specification concerning the behaviour of components is appropriate. Just exactly how close the real behaviour and the ideal must match may be a question with no perfectly general and systematic answer. Even though the practical application is tricky, what is contentious is the nature of the relation between levels. None of the authors really approaches the question of how one level accounts for another level. At most there is an appeal to gross generalities.

5.2.Level individuation and interlevel relationship

In general, theorists that have discussed the issue of levels use the term "levels of organization" to refer to the explanatory cascade within cognitive science. They usually refer to the fact that complex systems are to be seen as typically having "multiple levels of organization". As we saw in Chapter 2, this model of multiple levels can be compared to a hierarchy, with the components at each ascending level being some kind of composite made up of the entities present at the next level down. In this change of levels there is a further decomposition and lessening of the degree of abstraction of the activities, where components in the higher-level explanation are further decomposed.

5.2.1.Nomic Specification

On the one hand, we can take levels on an ontologically strong view, with the idea that when we are describing levels we are describing intrinsic properties of the hierarchy which would cross-classify that of their explanatory role. In this sense, we should be able to find some way to individuate such levels, and this could be provided by the laws proper to each level of the hierarchy:

> **Principle of Nomic Autonomy of Levels**: Levels in science are individuated on the basis of laws proper to the level, that is, objects and events fall under a level, or share in a property, insofar as they fall under the scope of the same laws.

Newell (1982), for example, identifies three different properties intrinsíc to such individuation:

> (a) The specification of a system at each level always determines completely a definite behaviour for the system at that level.
>
> (b) The behaviour of the total system results from the local effects of each component of the system processing the medium (level) as its inputs to produce its outputs.
>
> (c) The immense variety of behaviour is obtained by system structure, i.e., by the variety of ways of assembling a small number of component types.

Succinctly, we could say that each level establishes or describes different generalizations that are only expressible at one level. The type of generalization concerned in cognitive science is presented by Pylyshyn (e.g. Pylyshyn 1984, p.33), who points to the fact that certain sets of properties have something in common that can be said to pertain to a given level. These generalities are captured in terms of a precise taxonomy or vocabulary, one for the physical level, another for what he calls the 'functional' level, and still another for the psychological or semantical level. Each vocabulary or taxonomy may lead to the postulation of a new level. In the case of the functional level it is a vocabulary in which theoretical constructs are identified by their role in explaining or generating the behavioural regularities in question (for example, the symbolic terms that appear in any computational or information-processing model of a psychological process). At such a level, the level of the algorithm, it is unnecessary to appeal to representations in the level of algorithm. For Pylyshyn, we do not need the notion of representation to explain how a device works, though we need it to explain how it performs the function intended by its designer. In other words:

> (...) the content of the representations- or, what the properties in question actually represent- is not part of the explanation itself. (Pylyshyn 1984, p.26)

At the task level Pylyshyn appeals to the intentional vocabulary and the laws or principles of rationality. This level explains the way in which content enters causal explanations. For example, when we try to account for our logical reasoning we could first give a picture of the symbol and rules we mentally employ. Modern logic has to some extent -at least, within deductive logic- made it possible to specify principles such as validity in purely formal or syntactic terms, that is, without reference to the content of nonlogical terms. But this, and not even other rational rules, could be specified in the following way:

> Every rational rule might be specified in a formal notation, and the set of such rules that characterize an ideally rational individual might be listed; but one could still not express syntactically what it was that all such rules have in common which distinguishes them from other syntactically well-formed rules - because *that* is a semantic property (a property that cannot be stated without reference to what is represented by the expressions in the rules). (Pylyshyn 1984, p.38, n.7)

This focus on capturing generalizations has many aspects. One is of course that they typically allow for reasonable explanations and predictions on the basis of ignoring detailed information about the physics of a determined system. So, for example, we can predict the distribution of inherited traits of organisms via classical genetics without knowing anything about DNA, or predict the answer a given computer will give for an arithmetical problem while remaining ignorant of the electrical properties of semiconductors. What's critical here is not so much the fact of multiple realizability; rather, it's the indifference to the particularities of lower-level realization that's critical. The fact of the matter is that a certain level is indifferent to the explanation of the system implying that if the task level is satisfied, regardless of implementation. This would account for the behaviours under consideration.

As McClamrock (1991) has noted, considering a level in this sense depends on idealizing about the behaviour of particular lower-level structures; we might even see them simply in terms of their normal input/output functions and their local contribution to the behaviour of the larger system rather than in terms of the details of their internal structures. Generality is achieved by singling out subsystems via input/output specifications and ignoring the internal structure by which they might produce that function. For example, the analysis of the behaviour of a given computer running, say, a programme written in LISP can be appropriate, but it leaves totally unspecified how a given primitive LISP function (such as car(students) - i.e. "give me the first item on the list students") is calculated. The

LISP program is completely compatible with any way of representing the lists in memory, or even with different underlying machine architectures (e.g. it could be implemented on a Von Neumann machine or by a conectionist machine). The question of exactly when such idealization about the behaviour of components is appropriate has nevertheless no definite answer.

5.2.2.Explanatory role

On the other hand, we can individuate levels of a hierarchy by their explanatory role. In this sense, some theorists propose explanations of a system's behaviour at every level from the higher to the lower. When applied to a cognitive system these explanations have to provide its functional characterization, where the processes are characterized less in terms of their intrinsic state and more in terms of the overall functional role they play in the workings of the device:

> **Principle of The Explanatory Role of Levels:** Levels in scientific explanation are individuated on the basis of the functional role of the level in accounting for the behaviour of the hierarchy to which it belongs.

Hence, the individuation of a level can be established by the theorist's explanatory requirements. As a matter of fact, the core model behind Grandpa's explanatory framework, that of Marr (1982), specifies each level according to explanatory considerations. The task level is that which responds to the question of *what* the system computes and *why*, whereas the algorithmic level acknowledges *how* such a function is computed, and the implementation level specifies *how* the algorithm is implemented.

What determines the recognition of a level? A new level can often be recognized by the presence of constraints on the behaviour of a system over and above the constraints that can be expressed in terms of the available principles. As Pylyshyn (1984, pp.37-38) has noted, concerning for example the constraints of the algorithmic level, there are generalizations that would be viewed as nonsensical or bizarre or incoherent by someone who interpreted the algorithmic states as having representational content. For example, there could be a regularity among algorithms that would be interpreted as the rule that if one believes that all ravens are black, and that there is a raven in the tree, then that person goes into a state corresponding to the belief that the sky is falling. "What are the principles that prevent such rules from occurring? Why don't certain rules permitted by the constraints of the functional architecture, in fact, occur?" wonders Pylyshyn (ibid., p.38) For him, the reasonable answer is that there may be additional principles, perhaps universal in human cognitive systems, not stateable in terms of constraints on the functional mechanisms, principles which are imported from outside the specific level to explain these particular constraints.

These constraints are sometimes context-dependent properties, that is, properties that depend on occurring in the right context, and not just on the local and intrinsic properties of the particular event or object itself. For example, the position of a given DNA sequence with respect to the rest of the genetic material is critical to its status as a gene; type-identical DNA sequences at different loci play different hereditary roles. So for a particular DNA sequence to be, say, a green-eye gene, it must be in an appropriate position on a particular chromosome. A similar case can be made for mechanical systems like a valve, whose functional properties depend on the context being considered. The very same physically characterized air flow valve can be a choke in one context (i.e. when it occurs above the fuel jets) and a throttle in another (when it occurs below the jets) (McClamrock 1991). In sum, whether a given valve is a choke or a throttle depends on its surrounding context. By "contextualizing" objects in this way we shift from a categorization of them in terms of local and intrinsic properties to their context-dependent functional ones. However, considerations about context-dependence can and should arise at more than one level of analysis of a complex system, and may have quite different answers at the different levels. For example, the properties of DNA sequences as objects of chemistry depends only on their local physical structure. But their properties as genes depend on their overall contribution to the phenotype; and what contribution they actually make to the phenotype is highly dependent on context, i.e., its dependent on where the sequence is in relation to the rest of the genetic materials, and on the precise nature of the coding mechanisms which act on the sequences. Accordingly, to get the story right about which constraints are the relevant ones, one needs to know what a complex system is doing at the higher level in order to find out at the lower level how it accomplishes that task - that is, we often need to know the function of the complex system being analyzed to know what aspects of structure to look at. As McClamrock (1991) has noted, the important fact is that regardless of salience at the lower levels, what picks out which lower-level properties are important is understanding the overall workings of the complex system is the higher level. Therefore, to take a well-known example, we could interpret that the heart is basically a noisemaker, and then the lower-level properties which will seem most significant might be things like the resonant frequency of the various chambers, or the transient noises created by the movement of the various valves. Therefore, understanding the behaviour of a complex system requires knowing which aspects of the complex mass of lower-level properties are significant in making a contribution to the overall behaviour of the system, and this is obtained by acquiring some sense of the higher-level functioning of the system. However, and this is important for our discussion, according to McClamrock (1991) contextualizing or de-contextualizing can be done without a concurrent shift in level - that is, we might reinterpret the functions of parts against a broader background of the system without at the same time shifting the level size of the parts.

This explanatory approach to levels has the advantage of not making a claim about the actual number of levels of organization in cognitive systems. If that were the case, the framework would be claiming that cognitive systems ought to be "ontologically" organized in only three levels of organization. However, there is no reason to believe that an explanatory framework implies such an assumption. This does not mean that the brain could be special in this way; someone might possibly try to offer some reason to think that cognitive systems are to be discharged in three levels. Yet, the number of autonomous levels of organization of a system seems rather an empirical fact about each particular type of system we consider. The number of actual algorithmic levels of organization in any given information-processing system (including the brain) is an entirely empirical question. It could be zero (as eliminativists imply) but it could also be far more than one (as in the nested virtual machines in a real computer).

Accordingly, the distinction between levels of organization and levels of explanation would be orthogonal to the whole matter. For each of those levels of organization or decomposition, there could be three general kinds of questions that we can pose, or three kinds of explanations we might try to give: questions about that structure itself; questions about the functional, context-dependent properties of the parts and relations in that structure as well as their contribution to the functioning of the system as a whole; and questions about the implementation of the primitive parts of that algorithmic structure. Or to put it in a way even closer to Marr's, we might see the three perspectives of the algorithm, content of computation, and implementation as having something like the questions of *what function, how is it accomplished and how is it implemented.* Let's see now the inter-level relationship under the explanatory approach, which has been labelled "classical cascade" by Franks (1995).

5.3.Franks' and the blocking of the classical cascade

Franks grants Dennett the picture of an explanation in cognitive science as a "triumphant cascade" (Dennett 1987, p.227), an explanation that cuts across the three levels of Grandpa's framework. The fact of its pervasiveness and acceptance compels Franks to label it as the "classical cascade". The cascade consists in the constraining of the higher levels to the lower levels, while the lower levels fill the details left unresolved by the higher levels. In this sense, the task level provides a set of predictive relations between inputs and outputs, but it adds no explanation to that mapping. The algorithmic level has to account for the mapping specified by the task level and provides, in turn, a sequence of states which can map intervening states on to each other. This allows for allowing finely grained predictions, but without specifying the physical properties of the states. The implementation level must account for the algorithm being computed and also give the physical details of the mechanism. The strategy is to go from a functional specification to a physical one via the interpolation of an algorithmic level. Each level of explanation is autonomous yet constrained by the other levels. This means that we can find explanations of different types of systems that share a task-level account and an algorithm account, though they differ in the implementation details. Cognitive-science practice takes it that the one-to-many relationships holding downwards through the levels provide for an autonomy of description at each level. A task-level account makes no direct commitment about the algorithm employed to compute the cognitive function; similarly, a formal specification of the algorithm makes no presuppositions about the physical implementation of that˙algorithm. In general, all implementational descriptions incorporate an algorithm and a task-level account; likewise, any algorithm account incorporates a task-level account. Therefore, proceeding downwards through the levels provides for a progressively more concrete description of the faculty in the agent, leading to a physical explanation. Proceeding 'upwards' provides for subsumption of a particular concrete case under the algorithmic description that covers the equivalent class of implementations, which is in turn subsumed by the general functional description that also provides a semantics for the equivalent class of algorithms.

For Franks, a basic requirement for a successful cascade can be termed the 'inheritance of the superordinate': given a particular task level starting point, any algorithm must compute the same function, and any implementation must implement the same algorithm and compute the same function. In other words, lower levels inherit their superordinates. Where this fails a mismatch occurs between the between the description of a given faculty at one level, and its counterpart description at a lower level. If a task-level functional mapping fails to be inherited, then the algorithm specified at the next level down is filling the details of a different function. In such a scenario, a functional mismatch between levels appears, with the consequence that the functional constraints of higher levels are not complied with by the lower ones, even if all of them contribute to explain the same cognitive phenomenon; the cascade then cannot be said to apply to the same task, since the semantics given for different levels is not the same. Similarly, an algorithm specified at the algorithm level might not be inherited by the implementational description. In both cases the cascade fails to hold, since moving through the levels either changes the function under consideration, or changes the algorithm.

At this point, Franks borrows Frege's analogy to explain the mismatch between levels. The analogy is in the distinction between a map and the terrain that the map represents. A map can represent a terrain more or less accurately. The cascade is a sort of map of the terrain of the cognitive faculty under investigation: a task-level description is a map of the actual function an agent computes, an algorithm level is a map of the actual algorithm employed, and an implementation-level description is a map of the actual brain operations. A successful cascade is a map in which the same function is computed through all three levels, and the same algorithm all through.

The conundrum here is, of course, whether we can have a task level that gives an account of the function performed by the agent and an algorithm which does not fill in the details of the same functional mapping; while belonging to the explanation of the same phenomenon.[25] Franks' point is that this is an intrinsic consequence of the way in which cognitive theories are conceived, and that therefore the mismatch does not reflect an exception to be condoned or an incorrect way of designing theories. Making an argument out of an exception would not be worth the effort. The problem, according to Franks, is that the idealization implicit in competence theories implies the blocking of the cascade. That is, in failing to fulfill this requirement, competence theories can block a cascade, since idealizations of function and algorithm prevent inheritance of the superordinate. Then, and this is the relevant point, he claims that

> (...) unless the competence account specifies *the same function* as the performance [or algorithm] account (that is, performance inherits the superordinate), then competence *fails to provide an account of the implicit knowledge that explains* (subject to appropriate contextual factors) *performance*. (ibid.,p.480, my italics)

In sum, the classical cascade fails whenever a subordinate fails to inherit the superordinate levels' specifications, which is more a necessary than a contingent fact of competence theories. The problem, according to Franks, is that those mismatches are inherent to the idealized character of competence theories.

My view is that even if we can grant the point of the mismatch, it is not the nature of idealizations what induces the relevant mismatches; rather mismatches occur when levels draw upon different information that reflects two different perspectives on functional attributions. However, I present first Franks' points to develop later my own proposal.

[25] The problem should be seen in a conceptually ulterior stage than that of deciding *what sort of function* a given algorithm is computing. This, as we saw above, depends on the functional interpretation of the algorithm. Therefore, Frank's account assumes that the attributions are justified.

5.3.1.Idealizations and the failure of the cascade

Franks focuses on the case of grammatical competence, specifically within Chomskyan competence accounts. He argues that the most fully specified competence accounts of language involve a commitment to idealizations of the language faculty. For Franks there is a general incompatibility between the requirements on a given cascade and idealizations in task-level descriptions. The sort of idealizations implied in such accounts are due to *idealizing* the nature of the faculty, and in virtue of *abstracting away* from the details of mechanisms, e.g., the Chomskyan idea of "abstracting away from particular mechanism of the speaker". For Franks, such idealized descriptions cause difficulties for the cascade when they are intended as a scientifically realist claim about the nature of cognitive functioning. Franks develops this idea specifying a taxonomy amongst idealizations:

- **Simplifications.** Certain parameters are 'bracketed off' from consideration.

- **Extrapolations.** The map is credited with logical or computational powers that 'outstrips' its actual ability. There are three types of extrapolation:

  Generalizations: Explanations that are treated as non-*ceteris paribus*, that is, they are generalized to all circumstances, e.g., red objects appear red (overlooking the fact that they appear white when the ambient light is red).

  Abstractions: An explanation that subsumes the performance of individual agents in order to construct a composite function for the set of such agents.

  Overextension: Some input-output mappings not only are not performed by the agent in question, but also are not performed by any agent, regardless of their cognitive capacities.

According to Franks, under a realist stance the theorist has two options in the face of such idealizations. One is to preserve the cascade, which demands maintaining the idealization through the rest of the levels, so that at the next level down we specify an algorithm that idealizes the algorithm of the terrain. However, such a cascade does not explain the cognitive faculty originally under investigation, since according to Franks, it is idealized at the task level. The other option is to allow the idealization, though then the mismatch arises, so long as we cannot 'graft', in terms of Franks, the idealized task level onto the algorithm level, thus blocking the cascade. In other words, if we take the idealization of the faculty, which abstracts away from certain cognitive mechanisms, then the mismatch appears since the idealizations cannot be discharged.

How do such mismatches block the cascade? If we take the 'grafting' option, then when the idealization is a simplification in terms of Franks, the problem is that the specification of the task-level function will leave out aspects of algorithmic and implementational processes that are not semantically interpreted in the theory. On the other hand, when the idealization is an extrapolation at the task level, a description at such a level will be supersets of both the domain and range of the functions actually computed by the algorithm and the implementation. As a matter of fact, an extrapolated function encompasses, as we saw above, the function that is performed by the algorithm. Whether or not this still supports the cascade depends upon the view of the task level. With the view that function is specified as a function-in-extension, then the cascade might be maintained, since the extrapolated function is subsumed by the individual function, the latter having only a restricted domain and range. On the function-composition version (that is, the version in which the function specifies the information-processing constraints), the cascade is blocked, because the task level is intended as an analysis of the information-processing nature of the task, and a different set of functions and subfunctions defines a different task. If the individual function removes certain subfunctions, but preserves others, it is not clear that the same task is being computed, or if the overall input-output mapping is being altered. When generalizing, the cascade fails for those individuals for whom the *ceteris paribus* clauses are not satisfied, though it still holds for those for whom the clauses are satisfied. Similarly, when abstracting the function, a cascade for the community as a whole may hold: individual explanation gives way to social explanation. However, in overextension, the cascade of course fails. Where only some of the mappings are overextensions, the range of performable mappings (for which the cascade may still hold) may be empirically discoverable; where all of the mappings are overextensions, there will be greater difficulty. The problem would then be how to separate aspects of the function that are performable from those that are superfluous.

It is the contention of Franks that competence theories involve an overextension of function, thereby failing to support a cascade for cognition. Franks applies this to the Chomskyan enterprise of explaining linguistic competence. He interprets Chomsky as proposing that the task level expresses an idealized (overextension) function-composition view, and that the algorithm level reinstates the factors idealized away at the task level, thus defining a resource-limited function and a psychological algorithm. This has two implications. First, there is a *prima facie* inconsistency between the task level as an idealized function-composition description, and the algorithm level as a non-idealized psychological algorithm. Since the task level is suggestive of a certain style of processing, it should not diverge radically from the algorithm level but, since it is an idealization, it must do so if the algorithm level is resource-limited. The second implication is a functional mismatch between the task level and the algorithm level: human linguistic processing is a finite task, restricted by memory and time limitations. In short, the psychological algorithms treat grammar as specifying decidable functions -even if the task level rules of grammar employ undecidable functions. This finiteness is how competence/task-level function overextends in the case of Chomsky to performance/algorithm function.

Such a mismatch can be found, for instance, between a competence theory of grammar and a parser. Matthews (1991) notes the example of the parser proposed in Fodor, Bever, and Garrett (1974). This parser would, according to Matthews, incorporate heuristic procedures that would not inherit the function specified by the grammar at the task level, though it would comply with it. The grammar would be a version of Chomsky's Extended Standard Theory (EST), whereas the parser would be a set of pattern-action rules similar to the rules of Newell and Simon's production systems. In this situation, the theory could well account for the ability even if it did not necessarily describe the actual causal structure of the cognizer. Nevertheless, it could account for the ability because it would be the *unique explanation* of why such an ability is performed. The system would not even have to draw upon the same information as that of the theory. The parser would nevertheless bear what Matthews calls an *explanatorily transparent relation*, since the syntactic generalizations that are captured by means of the theoretical constructs of EST (e.g. rules, principles, and structures) are *explained* in terms of the organization and operation of the mechanism postulated by the parsing theory. Then, if we take this parser to describe the salient causal structure of the system, then it follows that, for Chomsky, its lack of correspondence is compatible with a correct attribution of the "grammar" by the speaker. However, Franks would surely argue that, if the parser is taken to represent human language processing, and there is an idealized competence function, a mismatch appears. In short, the competence theory of Chomsky blocks the cascade.

Franks proposes therefore to *separate* the explanatory cascades for Chomskyan competence and performance accounts:

> Competence and performance are properly construed as two separate aspects of the faculty for understanding and producing language. Competence is the declarative causal engine of performance, and performance is the procedural or behavioural outcome of the interaction of competence with limiting psychological factors. The claim then would be that it is inappropriate to view competence and performance as both taking part in a single cascade explanation of language use, but rather we should view them as playing a role in separate explanations: one cascade for competence and another for performance. Since competence and performance are different faculties, it is not appropriate to expect them to each have the same functional or algorithmic profile. After all, there is no obvious reason why a cause and an effect should be at all similar. (...) An alternative view (and one perhaps more in the spirit of the rejoinder to my argument) would be to postulate a more complex relationship between the competence cascade and the performance cascade, such that the competence cascade provides input to performance, which is also constrained by processing limitations. In essence, this requires taking a competence account of function and algorithm and adding in facts about other cognitive faculties - including resource and memory limitations. (Franks ibid., p.497-498).

As I said above, I am in a sense inclined to grant the main point of Franks' argumentation, namely, that there can be mismatches in Grandpa's explanatory cascade, and that these mismatches are in fact intrinsic to cognitive-science theorising. However, it is my view that the problem is not the idealizations; or rather, the relevant mismatches for me are not caused by the idealizations. On the contrary, they are the consequence of the relevant motivation. As a matter of fact, it may be true that the mismatches are caused in some cases by certain types of idealizations, such as simplifications, generalizations or abstractions. However, these idealizations are harmless for cognitive-science theorising. It is sufficient tô take them into account and deem that they make the cascade unsteady, rather than block it. My view is that the worrying mismatches are the ones that come out as overextensions, but not because they are, in the case of language, "an account of consistent and humanly possible languages" (Franks 1995, p.488), but because what we consider as two levels of the same explanatory cascade are really two different explanatory projects. This will be developed below, but we need, for the moment, to extend the notion of mismatch presented by Franks so that it can help us in that direction.

5.4.Accordance without inheritance

5.4.1.Basic notions

To begin with, I take it that the right choice of functional cognitive theory should be the function-composition view, that is, the version that specifies the informational processing constraints of the task to be accomplished. Then, given this assumption, the right way to specify a mismatch should be the following. First, we should establish the possible mismatch within informational framework. Second, we stipulate inter-level inheritance as follows:

> **Informational Criterion of Inheritance:** The information drawn upon by a system remains constant through the levels of an explanatory cascade.

Accordingly, we will have a mismatch if the following condition holds:

> **Level Mismatch:** A mismatch occurs between levels of an explanatory cascade when two levels do not draw upon the same information.

However, we need a complementary condition that can guarantee that both accounts should be considered to belong to the same explanatory cascade:

> **Condition of correlation**: Two levels belong to the same explanatory cascade if, other things being equal, they contribute to the explanation of the same cognitive phenomenon.

It is my view then that a case can be made to the effect that the empirical evidence reviewed in the last chapter lends support to the idea that mismatches frequently occur in cognitive science. To start with, we have seen that the *task* level can be specified by the computation of the set of spatial relations. However, when we have examined how such tasks are accomplished, that is, when we try to specify the *algorithm* level, it turns out that, for example, an inside/outside computation is subserved by a variety of submechanisms *which do not draw upon the information* of "insideness" or "outsideness". According to the above condition of inheritance, we can say that an explanatory mismatch occurs. Additionally we can also say that, provided that the principle of correlation holds, both accounts can be said to belong to the same explanatory cascade, so long as they both contribute to the explanation of the same cognitive phenomenon, e.g., the ability to engage in an efficient perception of spatial relations.

Secondly, we have also seen, nevertheless, that the way the cognitive system is capable of satisfying a fractioned set of numerical abilities that *do not draw upon the information* that a theory of arithmetic specifies, if only for the fact that certain arithmetical operations (those which for example correspond to subitizing) are satisfied without a notion of quantification as understood in accord with arithmetical principles, but rather with a vague notion of *numerosity*. We can nevertheless talk of both accounts as belonging to the same explanatory cascade because they explain the very same phenomenon, i.e., the satisfaction of arithmetical operations.

Thirdly, we have also seen that there is an account at the *task* level of our ability to perceive and reason about objects and their motions. However, our cognitive system employs kinematic principles to be efficient at perceiving and reasoning about objects and their motions, rather than dynamic principles. Yet, these principles draw upon information other than that of the task account, even if they contribute to the same explanation.

Finally, we have seen how it is possible to construct an artifact that accords with some functional requirement, say, the *task* of "charging the battery". Such a task can be subserved by a group of mechanisms, none of which draws upon the notion or information of "battery" or "charging". In sum, it is not only possible, but also empirically evident, that mismatches occur in cognitive theorising.

5.4.2.Perspectives rather than levels

Clearly, it is inappropriate to view the incompatible accounts of a given cognitive function as both taking part in a single cascade explanation of a function, but rather we should view them as playing a role in separate explanations. The mismatch between the examples presented can be properly construed as two separate aspects of the faculty for understanding and producing the function in question. This implies a difference both in the *epistemological stance* of the theorist and a difference in *functional properties.* One is the account of a certain cognitive structure, and the other is the behavioural[26] account of the agent specified in terms of the relationship of the agent and the environment. The point will be developed below. For the moment note that each account will serve different explanatory aims, and will give different perspectives of a functional attribution. An account of how an agent satisfies a determined functional requirement will be provided by an analysis of the satisfied *task* which, according to Marr's consideration about the critical features of the computation of the system -the characterization of his computational top-level-, should be determined by the constraints of the problem. This analysis will give an idea of how a system seems to comply with a functional description, framed within a given functional pattern. However, the relevant point is that this characterization should not imply its epistemic attribution, since it could be that the system is performing a process in which the task is internalized in some way in its cognitive mechanisms. Rather, the internal processes could very well contribute to the satisfaction of the *task* by computing another function. Finally, the need for considering the *task* as a correct attribution should be justified by the

[26] Behaviour should be taken here in its broadest conception, rather than in the behaviouristic way. A behaviour which includes (Kim 1996, p.28):

i) Physiological reactions and responses: increase in pulse rate;
ii) Bodily motions: an arm's raising
iii) Actions involving bodily motions: typing, greeting, shopping
iv) Actions not involving bodily motions: reasoning, guessing, calculating

teleological causal powers of the task. In other words, the fact that the mechanisms of the agent fulfils a certain function is reason enough for their existence.

As a matter of fact, I contend that the mismatch reflects a difference in explanatory projects both having their place in a theory of functional attribution. To support my argument, I will appeal to the distinction made by Rowlands (1997) between two different content attributions, within the account of teleological semantics. I will transfer the contrast between content attributions to the case of functional attributions, with the conviction that the metaphysical foundation of the distinction will hold. However, to do this I will first have to make conceptual room for such a difference within normal cognitive function attributions, and to look subsequently to the naturalistic support within theories of function.

5.5.Explanatory projects within cognitive functional attribution

A plausible answer to the question of how we attribute functions within cognitive science seems to imply that we individuate cognitive functions by performing a behavioural specification and then we infer that the system has the function of being competent in such behaviour. In Grandpa's framework we identify a cognitive function when we provide an abstract formulation of the information-processing task that defines a given psychological ability. The identification of such an efficiency in terms of an informational-processing stance transfers, in a manner of speaking, a capacity to the system that accounts for the efficiency. In short, we tend to conflate efficiency *in some task* with a capacity *for that task*. So if we show that we are capable of solving arithmetic problems, then we infer that we have some *arithmetic* capacity, or that if we show that we can distinguish between well-formed sentences from nonsense string of signs, then we infer that we have some *grammatical* capacities. Even when we confront some complex task efficiencies, such as the task of solving equations of Newtonian mechanics, we tend to attribute knowledge in the domain of the efficiency. In other words, cognitive function attributions are normally task-dependent. But the move to think of the cognitive system as task-oriented or structured is not a trivial one. A task or a behavioural description is, after all, a setting against which we describe a certain behaviour. So there must be an underlying assumption that does the work of attribution, namely, a certain support for transferring the behavioural identification to the cognitive structure, and this we could precisely call the:

> **Principle of transference**: The individuation of a cognitive capacity is transferred from that of its behavioural efficiency.

Cummins, who has thoroughly analyzed such issues, reaches the contention of that principle in the following way. First he presents what a cognitive function ascription is for him, namely,

> (...) to ascribe a function to something is to ascribe a capacity to it that is singled out by its role in an analysis of some capacity of a containing system. When a capacity of a containing system is appropriately explained via analysis, the analysing capacities emerge as functions (1983, p.28)

Cummins calls such an explanatory strategy "functional analysis".[27] Functions are assigned when analysing a complex capacity into a set of simpler capacities that are to be explained by subsumption under laws. The function of an item is its contribution to the overall capacity, and the overall capacity is explained in terms of the contributing capacities of parts of the system. We then explain a cognitive capacity by a sub-class or type of functional analysis. This analysis accounts for a capacity whose inputs and outputs are specified via their semantic interpretations. The capacity for addition, for example, is the capacity to produce as an output the correct sum of the inputs. The outputs must be interpretable as numerals representing the sum of the numbers represented by the numerals in the inputs. Two inputs (or outputs for that matter) count as the same thing -i.e. as tokens of the same type- in the case that they have the same interpretation. Capacities specified in this way are called by Cummins, information-processing capacities (ibid., p.34). On the other hand, he specifies the conditions that underlie why a certain behavioural efficiency should be taken as a capacity of the system, namely:

> Whether or not a device was designed to add, it may be an adder -i.e., be capable of adding- in virtue of

[27] As Boden (1988, p. 169) says, we must remember that the identification of a task of some psychological domain is logically prior to its analysis.

> the fact that it is possible to interpret outputs as numerals representing sums of numbers represented by numerals interpreting inputs. What makes a device an adder is simply the possibility of such an interpretation. When we discover that such an interpretation is possible for a device d, we have discovered that d has a certain information-processing capacity. To explain this capacity we need to find some way of interpreting the causal sequences connecting inputs and outputs as steps in an addition algorithm. (ibid., p.42)

However he gives a caveat to this assertion which, in a certain sense, puts the issue back into contention:

> I am ignoring here the possibility of information-processing instantiations. Decimal adders are typically instantiated as binary adders, which are in turn instantiated as truth-function computers. But they needn't be. Mechanical calculators and, for all I know, people physically instantiate decimal-addition algorithms. (ibid., p.197)

The problem is in the precision "for all I know". The question here is not that he simply ignores the possibility of other *representational* instantiations, such as the fact that one instantiation can be in the above case a "decimal adder" or a "binary adder". What is more, he also ignores other *computational* instantiations. He presents different situations in which the idea is made clearer:

> Since we do this sort of analysis [functional analysis] *without reference to an instantiating system*, the analysis is evidently not an analysis of an instantiating system. The analysing capacities are conceived as capacities of the whole system. Thus functional analysis puts very indirect constraints on componential analysis. (ibid., pp. 29-31, my italics)

As he also says, and this is easily forgotten, a functional analysis is often very difficult, so that when we reach some analysis of the capacity we tend to automatically infer its instantiation. Marr also has a say here:

> (...) some judgment has to be applied when deciding whether the computational theory for a problem has been formulated adequately. (Marr 1977, p.130 in Haugeland)

In the case of chess he says that "one presumably wants a computational theory that has a general application, together with a demonstration that it happens to be applicable to some class of games of chess, and evidence that *we play games in this class*" (ibid. p.130, my italics). The fact is that he considers that "the kind of judgment that is needed seems to be similar to that which decides whether a result in mathematics amounts to a substantial new theorem, and I do not feel uncomfortable about having to leave the basis of such judgments unspecified." Hence, even if we accord to a certain functional pattern of adding, it could turn out that we were computing some other function. The cases presented in the last chapter are a few examples of what I am getting at.

One particularly revealing example, presented by Patricia Kitcher (1988), exposes the distinction between task and process in Marr's theories, showing that compliance with a top-level theory does not guarantee implementation. Kitcher begins by characterizing Marr's theory as an optimizing theory. For her, Marr assumes throughout that if the function of stage *S* in visual processing is to compute *I*-2 from *I*-1, then that stage carries out a computation in an optimal way. It employs exactly the information needed for the derivation. Additionally, she credits Marr with the assumption that, in general, the visual system is well designed for the extraction and representation of information about the shape, spatial location, and orientation of object surfaces. In other words, Marr's computational approach tries to understand visual processing by looking at the information-processing task to be carried out by the whole system or by one of its subparts. The analysis starts with the task, tries to determine the resources available for carrying out the task, and then proposes an account of how the task must be performed. Therefore, for example, the "fundamental theorem of stereopsis" offers sufficient conditions for matching items in the left and right images of a scene so that the disparity between the images may be calculated (Marr 1982, 112-115). However, Kitcher notes that, regardless of the correctness of such a computational theory, that is, the sufficient conditions for deriving this information have been correctly specified, it could turn out -contrary to Marr's design for a theory vision- that this information is of virtually no use in figuring out how we actually determine the distances of viewed objects:

> I will illustrate the problem by means of an example. In the title paper of *The Panda's Thumb*, Stephen Jay Gould (1980) presents a case where beginning with the task to be accomplished is picking up the

> wrong end of the stick (...) Gould's point is that taking an engineering approach to the grasping capacity of the panda's thumb is not very fruitful because, from a design point of view, the thumb is pretty klutzy. In Michael Ghiselin's terminology, it is not a lovely contrivance, but a contraption. (...) The functional decomposition of the task carries very little of the explanatory burden of the whole account. Only a very loose task description is part of the story (...) Most of the details of the capacity are provided by the "implementation" level. (...) To sum up, the worry that Marr's theory could not turn up to be wrong is mistaken. If our visual system turns out to be not terribly orderly or not well designed, form an engineering point of view, then Marr's project of a unified theory of vision will fail. Marr and his co-workers are gambling that vision is an elegant contrivance with well-organized, unified subprocesses. It is a bold gamble that they could lose (1988, p.22).

In short, we have the non-trivial possibility that under some correct interpretation the machine or system might show a behaviour that can be effectively described as A though it has no *knowledge* of A. The fact is that one thing is what *efficiency* the system shows and quite another *what* the system does. Indeed, the risk in embracing unconditionally such a principle is that of conflating two very different things, namely, a capacity in a *dispositional* sense (and as we will see, even in a teleological sense) and a capacity as an *epistemic* notion. A capacity in a dispositional sense points to an effect, a behavioural effect in our case, whereas when we refer to the capacity in an epistemic sense, we point to the internal cognitive structure of the system that can account for such an effect. When we say that a cognitive system is capable of adding, what we are claiming is that the system can solve adding problems But when we say that a system has knowledge compatible with a theory of adding, we imply that there is a cognitive structure that implements such a theory. Both ascriptions can be connected, that is, they can correspond, but they needn't do so. And that's the point. We may be able to add without any knowledge of arithmetical principles. Proving that we can add reveals nothing about our knowledge of adding principles.

This does not mean that we cannot infer knowledge from behavioural descriptions, but simply notes that individuating the function of a system and individuating the knowledge that makes it possible are not two birds of a feather. The relevant point here is not that some functional attributions are *incorrect* because the actual mechanism that accounts for it does not correspond with the functional description. This may happen, but it is of little interest here. Rather, the important situation is when we have a *correct* attribution, like the "following walls" robot, whereas the mechanism that accounts for it does nŏt *map* the "following walls" function. It is a *correct* attribution insofar as it may be evolutively significant, since the "following walls" function could be selectively significant. Or, alternatively, the fact that humans look like logical compliers, that is, in accord with some theory of logic and its principles, may be *evolutively* necessary for us to survive, even if we have not internalized a theory of the principles of logic. In other words, it might be that the cognitive effect for which the system is there is produced by some causal principles that do not match the structure of the effect. As Hardcastle puts it:

> One good and easy way to model the deductive reasoning processes in humans is to use what we write on paper or explicitly acknowledge as our steps in reasoning. However, this fact does not show by any means that how we actually reason in our heads conforms to recognizable step-wise functions. Anyone who has taught introductory logic knows first-hand that people do not in fact reason deductively, even when they think they are. (...) We need to keep in mind that how we express ourselves on paper or publicly may not accurately reflect what is going on the inside. (Hardcastle 1997, pp.378-9)

The cognitive psychologist could nevertheless respond to such a challenge by saying that if that is the case, then the problem is that the description is *incorrect*. Suppose that it is true of humans that their ability in logic is due to some probabilistic module that is applied -generally correctly- to logical problems. Then the psychologist could say that, when we register logical inferences, we ought to say that we perform statistical ones. The bottom line is that the informational process of a task must be undertaken by the system in the form established by the task description. If this doesn't occur, then either the task cannot be accomplished or the attribution of the task to the system is faulty.

Any explanation in psychology should give a information-processing description of how subjects solve problems, perceive linguistic items or forms, which then should be sufficient to sustain the transfer principle. Accordingly, it seems either trivial or uninteresting to debate whether the functions of our cognitive systems are **a** or **a'**, since the important question is giving an account of how we do the task in question. However, if this is the contention of the psychologist, then the point should be made explicit once again. The idea is that within a certain descriptive account, a system might be complying with the whole functional-descriptive pattern, by being the effect of some process, while the process might be responding to some other causal principles. Again, Hardcastle is of help:

> Maybe Horgan and Tienson [1996] would say that if it did turn out that we performed rigorous proofs by pattern recognition instead of appealing to explicit rules, then we really do not reason deductively after all, for they claim that "deductive reasoning is possible only for a creature whose cognitive states have formal structure... [It] proceeds in steps that ... attend to the exceptionless relations that are determined by the structured content of those cognitive states. This is not simply a matter of being pushed along by cognitive forces" (p.104). But this just begs the interesting question for those interested in modelling actual human cognition. We do something that resembles deductive reasoning in problem solving. The question is how we are supposed to understand those processes, regardless of whether it turns out to be the actual derivation. (Hardcastle 1997, p.378)

Clark makes a similar point:

> These programs characteristically attempt to model *fragments* of what we might term recent human achievements. By this I mean they focus on tasks that we intelligent, language-using human beings perform (or at least think we perform) largely by conscious and deliberate efforts. Such tasks tend to be well structured in the sense of having definite and recognizable goals to be achieved by deploying a limited set of tools (e.g., games and puzzles with prescribed legal moves, theorem proving, medical diagnosis, cryparithmetic and son on). They also tend to be the tasks we do slowly and badly in comparison with perceptual and sensorimotor tasks, which we generally do quickly and fluently. (1989, p.17)

Marr has frequently dwelled on this issue. He has insisted that behind many cognitive abilities attributions there could be a *wrong* characterization of the task that the system is apparently and optimally up to: "I have no doubt that when we do mental arithmetic we are doing *something* well, but it is not arithmetic" (Marr 1977, p.140 in Haugeland). Marr's idea is obviously that it is at least contentious, though counterintuitive nevertheless, that, in this case, *solving* arithmetic problems *implies* having an arithmetic *capacity*, or what is the same, that a *manifested* capacity implies an *internal* capacity.[28] Here there are other

[28] Boden (1988, p.170) interprets the assertion in the following way: "If Marr is right, it would not be only Newell and Simon who are wasting their time: similar strictures would apply to recent computational work offering a systematic account of children's successes and errors in subtraction. In short, according to Marr, *all* psychological research on arithmetical (and other types of) problem-solving is a theoretically useless activity. It will remain so until we can identify the basic task(s) and the relevant information-processing constraints. Until then, we had better

examples about the same issue:

> (...) the figure-ground "problem" may not be a single problem, being instead a mixture of several subproblems which combine to achieve figural separation (...). There is in fact no reason why a solution to the figure-ground problem should be derivable from a single underlying theory. (...) There may exist no Type 1 theory of English syntax of the type that transformational grammar attempts to define (...) An abstract theory of syntax may be an illusion, approximating what really happens in the sense that (...) the behaviour of [a] set of processes that implement [a certain task] and which, in the final analysis, are all the theory there is. In other words, the grammar of natural language may have a theory of Type 2 rather that of Type 1. (Marr 1977, pp.133-135 in Haugeland)

One of Marr's recurrent ideas is to insist on the importance of correctly characterizing a function in order to reach a robust explanation of the capacities of a system. Marr indicates many times how important it is to isolate an "information processing problem", although he never provided a clear criteria of how to address such a problem. He is conceptually close to Haugeland (1981), though, in equating the explanation in psychology as giving an explanation to an input/output ability, where the inputs function as "posing problems" to the system. As a matter of fact for Marr the most important task of cognitive science is finding *good problems*. The goal of a good AI strategy is, in Marr's terms, the *identification* of "interesting and solvable" information-processing problems. For Marr this is precisely the fundamental task of a cognitive explanation applied to Artificial Intelligence, the identification of his "theory of computation", which is the abstract formulation of "*what* is being computed and *why*". He considered it fundamental because among other reasons, for him, once a computational theory has been established for a particular problem, it never has to be done again:

> Although algorithms and mechanisms are empirically more accessible, it is the top level, the level of the computational theory, which is critically important from an information-processing point of view. The reason for this is that the nature of the computations that underlie perception depends more upon the computational problems that have to be solved than upon the particular hardware in which their solutions are implemented. To phrase the matter another way, an algorithm is likely to be understood more readily

concentrate on what Marr would term simpler (though by no means simple) problems -such as low-level vision, parsing, the perception of music, and (possibly) semantics".

> by understanding the nature of the problem being solved than by examining the mechanism (and the hardware) in which it is embodied. (Marr 1982, p.27)

For Marr if one begins by hypothesizing a particular algorithm used by an organism without first understanding exactly what the algorithm is supposed to be computing, one runs the danger of simply mimicking fragments of behaviour without understanding the principles or goals of the capacity to be explained.

This will lead us to the following point: Marr believes, together with, for example, Newell and Simon, that it is the *nature of the problem* that dictates how we solve it, and therefore opens the necessary space to differentiate the two perspectives that we are aiming at: The *task* performed and the *cognition* implicated. The *task* that the system might comply with "depends only on the nature of the problem to which it is a solution" (Marr 1982 p.129):

> If we believe that the aim of the information-processing studies is to formulate and understand particular information-processing problems, the structure of those problems is central, not the mechanisms through which their solutions are implemented. Therefore, in exploiting this fact, the first thing to do is to find problems that we can solve well, find out how to solve them, and examine our performance in the light of that understanding. The most fruitful source of such problems is operations that we perform well, fluently, and hence unconsciously since it is difficult to see how reliability could be achieved if there was no sound underlying method. (1982, p.347)

Therefore, we could say that for Marr efficiency is unconstrained by the internal mechanism, so that he leaves us the space of describing a functional competence for which the internal mechanisms block the "explanatory cascade". The idea of putting the weight of functional individuation on the nature of the problem was introduced by Newell and Simon:

> Now if there is such a thing as behaviour demanded by a situation, and if a subject exhibits it, then his behaviour tells more about the task environment than about him. We learn about the subject only that he is in fact motivated toward the goal, and that he is in fact capable of discovering and executing the behaviour called for by the situation. If we put him in a different situation, he would behave differently (Newell and Simon 1972, p.53)

However, it would seem that we are dealing here with the issue of whether (sub)system A *has the function F*, instead of (sub)system A *functions as an F* (Griffiths 1993). Under this distinction, the examples presented up to now could be interpreted in the following way: our cognitive system could be functioning as an 'adder' while having the function of 'multiplying'. Here we could have a conflict with the etiological notion of function, for the 'proper function', the function of A could be that of multiplying, which in essence is the selectionist advantage, whereas in our terms the function would be the 'adding'. How can we resolve such tensions?

5.6.Theories of Function

There are two distinct concepts of function which are acceptable to Grandpa. One comes from the works of Larry Wright (1973) and the other from, e.g., Bigelow and Pargetter (1987) and Robert Cummins (1975, 1983). These two traditions allow theories of functions to be divided into two categories: selectionist or etiological theories and dispositional or design theories. Both face the challenge of answering questions of the following sort: How is the function of the heart individuated from its variety of effects? Why is the heart's function pumping blood, instead of making certain sounds, even that of filling a certain space in the body?

5.6.1.Etiological/Selectionist Theories

Larry Wright gave a version of the theory of functions from a teleological point of view that has been taken as the role model for a teleological explanation of function. Wright proposed that functions are distinguished from mere effects by their explanatory significance. The function of something is the effect it has, which in turn explains why it is there. Hearts are found where they are because they pump blood, not because they make certain sounds or simply fill a certain space. According to some authors, there are certain details in Wright's proposal that required modification which were developed in very similar though different views by Ruth Millikan (1984, 1986, 1989a, 1989b, 1989c) and Karen Neander (1991a, 1991b). The idea is that an explanation of function in Wright's manner requires some process of selection.

The fundamental idea of selectionist theories is that the function of *x* is *f* just in case things of the type *x* got replicated because they (at least sometimes) did *f*. In terms of Millikan, for an item A to have a function F as a 'proper function', it is necessary (and close to sufficient) that A originated as a 'reproduction' (to give one example, as a copy, or a copy of a copy) of some prior item or items that, due in part to possession of the properties reproduced, have actually performed F in the past, and A exists because (causally, historically because) of this or these performances. In terms of Neander, it is a/the proper function of an item (X) of an organism (0) to do that which items of X's type did to contribute to the inclusive fitness of 0's ancestors and which caused the genotype, of which X is the phenotypic expression (or which may be X itself where X is the genotype) to increase proportionally in the gene pool. The proper function of an item is its *N*ormal function, where, following Millikan, the capitalized "N" indicates that this is a normative sense of normal as opposed to a causal or dispositional one. Even if in these proposals there is an appeal to natural selection, Godfrey-Smith (1996) argues that we can extend the notion to other sorts of selections, rather than simply selection on genetic variants. Conscious selection by a planning agent, selection through reinforced learning and other types of individual-level adaptation, even cultural processes follow this principle. In this sense he deems this notion of function to be "teleonomic"[29]: a function is to do whatever explains why it is there.

Proper functions are the sorts of functions that biologists assign to the biological systems. The notion is sometimes introduced by pointing out that proper functions are what things are *for* whilst other functions are not. This is normally expressed by the opposition between *having the function* F from mere *functioning as an* F. In this sense, proper functions differ from other functions in that they can be cited to explain the presence of a functional item. The presence of the liver can be partially explained by its capacity to store glycogen and secret bile. In other words, these functions enter into an evolutionary explanation of the presence of the liver. Additionally, a distinctive part of the biological

[29] The term is used to refer to those of traditional teleological theories that can be given a foundation in the operation of natural selection. Godfrey-Smith (1996) uses it to include the other sort of selection.

concept of function is that where there are functions there can be a *malfunction*. In other words, the concept is *normative* . If we have a theory of the function of cognition, we have a theory not just of what cognition actually does, but also of what is supposed to do, what its "proper" function is, in Millikan's terms. To be in the relevant circumstances but do something different is to malfunction.

Explanations that cite an item's proper functions are *teleological*. The existence and form of an item seem to be explained by its goal, or purpose, rather that its antecedent causes. The etiological approach to proper functions is an attempt to demystify these teleological explanations. To ascribe a proper function to an item is to claim that earlier items of the same type had the effect with which we now identify as a proper function and that their having had that effect explains the presence of later items of the type.

On the other hand, explanations that cite proper functions are *relational* explanations. Millikan-Normal functions are generally defined relative to some environmental object or feature. For example, the function of the chameleon's skin is to make the chameleon the same colour as its immediate environment, etc. The existence of the feature is due to the fact that it has evolved to meet certain environmental demands. Therefore, the representational characteristics of a given cognitive mechanism derive from the environmental objects, properties or relations that are incorporated into that mechanism's relational proper function. In other words, if a cognitive mechanism *M* has evolved in order to detect an environmental characteristic *E*, then this is what makes an appropriate state *S* of *M* to be a state about *E*; this is what gives the state *S* the content that *E*. Accordingly, the representational content of cognitive state *S* derives from the relational proper function of mechanism *M* that produces *S*.

5.6.2. Design/Dispositional account

Cummins' dispositional notion is also central to our discussion. As we saw earlier, Cummins argues that the practice of assigning functions derives from an explanatory strategy that he calls "functional analysis". Functions are assigned when analysing a complex capacity into a set of simpler capacities that are to be explained by subsumption. The function of an item is its contribution to the overall capacity. The overall capacity is explained in terms of the contributing capacities of parts of the system. Accordingly, design theories of functions define the function of a mechanism or process in terms of its functional role, that is, in terms of its contribution to some capacity of the system to which the process or mechanism belongs (Cummins 1983). Design theories thus relativize functions to capacities of containing systems:

> The (or a) function of *x* in a system $\sum$ is *f* relative to capacity *C* of $\sum$ just in case $\sum$'s capacity *C* analyzes (in part) into *x*'s capacity to *f*.

Cummins uses ascriptions of function without reference to its biological relevance, be it ecological, evolutive or whatever. His explanation does not address the issue of why a thing functionally characterized exists, but how some system shows some sort of capacity or disposition with such a functionally characterized thing being part of it. For Cummins, we individuate the heart's function as that of blood-pumping, as a component of an explanation of how a biological system manages the problem of providing oxygen to every cell in the body. Additionally, for Cummins, to malfunction is to do something other than the thing which explains how this component contributes to some particular capacity of the overall system. As a matter of fact, the position Cummins is not concerned with is to distinguish teleological from other non-teleological notions of function. This has led Millikan to argue that Cummins analysis does not contribute to the understanding of "proper functions" of biological items and human artifacts, since many "Cummins-functions" are not proper functions, and conversely there are some proper functions that are not Cummins-functions.

5.7.Rowlands organismic-algorithmic distinction

Etiological/teleological versions of function have nevertheless received some criticisms. Rowlands (1997) summarizes them in two fundamental problems. The first is what he calls the *problem of indeterminacy.* A biological function has been considered for some (e.g. Fodor 1990) to be indeterminate in the sense that there is no fact of the matter that could determine which interpretation of the function of an adapted biological mechanism is the correct one. The second problem, closely related according to Rowlands, is what he calls the *problem of transparency*. Ascriptions of biological function "are transparent in the sense that a statement of the form 'the function of biological mechanism M is to represent Fs' can be substituted *salva veritate* by a statement of the form 'the function of mechanism M is to represent Gs' provided that the statement 'F iff G' is counterfactual supporting" (Rowlands 1997, p.280). Accordingly, statements of functional ascriptions are transparent or extensional, which is incompatible with capturing the intensionality of psychological ascriptions. For Rowlands such problems stem from two related misunderstandings:

> (i) a confusion of the different levels of description at which biological functions can be specified, and (ii) a confusion over the objects to which the contents underwritten by these functions can be attributed. (Rowlands ibid., p.280)

In order to dissolve such problems Rowlands elaborates a proposal about proper-function attributions. Rowlands first draws a contrast between a cognitive state and a cognitive mechanism, which has the effect of freeing cognitive states of having to account for proper functions:

> An organism's cognitive state tokens are (often) caused by events occurring in that organism's environment. And there are mechanisms, typically neuronal, that mediate those causal transactions. Each of these mechanisms will, presumably, have an evolutionary history and, therefore, will possess a proper function. Moreover, it is plausible to suppose, this proper function will be precisely to mediate the tokenings of cognitive states.(Rowlands ibid., p.282)

Then, he exemplifies the indeterminacy problem with the noteworthy example of the sight-strike-feed mechanism of the frog that we already saw in Chapter 3. Let us remind the story. Frogs catch flies by way of a strike with their tongue. It is natural to suppose that mediating between the environmental presence of a fly and the motor response of the tongue strike is some sort of mechanism that registers the fly's presence in the vicinity and causes the strike of the frog's tongue. In other words, the presence of the fly might cause the relevant mechanism to go into state S, and its being in state S causes the tongue to strike. The teleological story goes on to consider that the content of state S is that of "fly" or "fly, there", deriving this content from the fact that the proper or Normal function of its underlying mechanism is to detect the presence of flies. However, there is a problem with this sort of story. The present account assumes that the proper function of the mechanism is to register the presence of flies in the vicinity. Yet, there is an alternative construal of the function of the internal mechanism: what the mechanism in question has been selected to respond to are little ambient black things.[30] In this case, the proper function of the mechanism is to mediate between little ambient black things and tokenings of a state that causes the frog's tongue to strike. This state will, then, be about little ambient black things and, therefore, mean that there are little ambient black things in the vicinity. The proper function of the mechanism is different in each case and, hence, the content is distinct in each case: the frog's mechanism is functioning Normally even when the frog strikes at a little ambient black thing that is not a fly but a lead pellet (it is usually referred as "BB") that happens to be in the vicinity. The problem of *indeterminacy* arises here because there is no fact of the matter that could determine which of these interpretations is the correct one. Evolutionary theory is neutral on both claims, as long as a sufficient number of little ambient black things in the frog's environment at selection time are flies (or edible bugs). Therefore, the teleological approach is thought to entail the indeterminacy of mental content.[31]

On the other hand, the problem that Rowlands has labelled of *transparency* states that the proper function of a selected mechanism will not decide between any pair of equivalent content ascriptions where the equivalence is counterfactual supporting:

> Or in a more formal way, the context *"was selected for representing things as F"* is transparent to the substitution of predicates reliably co-extensive with *F*, and, a fortiori, it is transparent to the substitution of predicates necessarily co-extensive with *F*.

Consequently, teleological ascriptions offer no hope of constructing contexts that are intensional as "believes that....". Teleology cannot, therefore, explain the intensionality of mental ascriptions. Obviously, both problems are related. The problem of indeterminacy arises because there seems to be no fact of the matter that could favour the interpretation

[30] Rowlands notes at this point that he refers here to environmental entities, not dots on a retinal image, that is, he is constructing the problem avoiding the proximal-distal opposition.

[31] This problem is related, or is a version of, what is known, after Fodor (1990), *disjunction problem*.

of the frog's mechanism as detecting flies, from another that interprets it as detecting little ambient black things. It is an indeterminacy that comes from the environment, since the account fails to distinguish different between interpretations limited to different environmental objects. Accordingly, these environmental correlates of cognitive mechanism proper function can be substituted interchangeably by which we have the problem of indeterminacy entails that of transparency.

Rowlands goes on to introduce the concept of *affordance*, which will help him constrain the content individuation of a cognitive state. He traces the notion back to J.J.Gibson (1979), for whom the affordances of the environment are, for a given creature, what it *furnishes* or *provides*, whether this benefits or harms the creature. A non-rigid surface, for example, like the surface of a lake does not afford support or easy locomotion for medium sized mammals, although it does for a water bug. Affordances are relational properties of things, and have to be specified relative to the creature in question. Different substances of the environment have different affordances for nutrition and manufacturing. Different objects of the environment have different affordances for manipulation. The relevant point is that what is important for the survival of an organism is not so much the objects *in* its environment but the affordances *of* its environment: it is not what an object is but what it *affords*.

Any organism that can recognize objects but not detect the affordances of those objects would not survive. On the other hand, any organism that can detect the affordances of objects even though it is incapable of recognizing those objects could survive. From the point of view of survival, it is the affordances of the environment and not the objects in the environment that are of importance. For Rowlands the important point is:

> At some level of functional specification, the function of a mechanism will be to enable the organism to detect a given affordance of its environment. (ibid., p.288)

This leads him to distinguish two levels of content specification: *organismic* and *algorithmic*. Roughly, he differentiates them at two levels that could be equated with the conceptual and sub-conceptual levels. However, even if his distinction could be placed *within* such levels, he is looking for some other properties. The point of the departure is the difference between attributing the perceptual content of an organism that could proceed with the locution "...sees that *P*". If a subject, say Jones, sees that *P* and makes the content attribution true, what underwrites this content attribution will be Jones' visual apparatus. However, it does not follow that we can attribute the same content to the visual apparatus. Rowlands points out that the visual apparatus *cannot* see. The visual apparatus is rather what allows the *organism* to see, and we thus can attribute visual contents to a person because of the proper functioning of the eye. Then, Rowlands makes an important distinction:

> Suppose we have an organism *O*, sensitive (i.e. able to detect) to some feature of the environment. On the basis of this sensitivity, let us suppose, we can attribute the perceptual content C_O to the organism. However, *O*'s sensitivity to this feature of the environment is underwritten or realized by mechanism *M*. And *M*, let us suppose, has this role because of its evolutionary history and the corresponding proper or Normal function with which this evolutionary history has endowed it. Because of this proper function we can, according to the teleological approach, attribute the content C_M to *M*. It does not follow, however, and, indeed, it is usually false, that C_O = C_M. This is true even though it is *M* that allows *O* to be sensitive to its environment in a way that warrants the attribution of content C_O to it. That is, even though it is the proper function of *M* that warrants the attribution of content C_M to it, and even though it is the fulfilling by *M* of its proper function that allows the content C_O to be attributed to *O*, it does not follow, and, indeed, is almost always false, that C_O = C_M The content attributable to a mechanism *M* and the content attributable to an organism *O* do not generally coincide, *even where it is the mechanism M that underwrites the attribution of content O*. (ibid., pp.288-289, the italics are mine)

In short, the content can be attributable at both the organismic and the sub-organismic levels, and both contents do not coincide.

For Rowlands, if we have to "adopt a teleological theory, we shall need a teleological account of both forms of content attribution" (p.289). Moreover:

> The above case, seems to require us to distinguish two proper functions of the mechanism *M*. These proper functions underwrite two importantly distinct levels of content-attribution. At what I shall call the *algorithmic* level of description, we might get the following sort of account: the proper function of mechanism *M* is to detect *G*s. For example, a proper function of one component of the visual system might be to detect texture density gradients in the structure of light surrounding the organism (the *optic array*). For any system that could do this, the content-attribution of the form "detects that *d*", where *d*

> represent the density gradient would be warranted. This is content that we attribute not to the organism as a whole, but to a mechanism possessed by an organism. It is content attributable at the sub-organismic level. On the other hand, at what I shall call the *organismic* level of description, we might get the following sort of account: the proper function of *M* is to enable *O* to detect *F*s. For example, a proper function of the component of the visual system responsible for detecting texture density gradients might be to enable the organism to perceive a roughly horizontal ground receding away from it into the distance. (Rowlands ibid., p.289)

Now we are in a position to look at the properties of this distinction. Organismic descriptions of the proper function of a mechanism derive content attributions for the organism as a whole. Algorithmic descriptions of the proper function of that mechanism derive content attributions for the mechanism itself. Organismic descriptions and algorithmic descriptions are, then for Rowlands, non-equivalent. The organismic description designates a function of the mechanism, what he calls the *organismic proper function*, which is different from, and more important, *irreducible* to, the function, that is, the algorithmic proper function designated by the algorithmic description. They are therefore different accounts.

To exemplify his position he offers the account of the rattlesnake's representation of its prey. The rattlesnake has a certain prey detection mechanism that is activated only if two conditions are satisfied. First, the snake's infrared detectors must be stimulated. Second, the visual system must get positive input. The former condition is satisfied when there is a localized source of warmth in the environment, the latter when there is a localized source of movement. When these two mechanisms are stimulated, the rattlesnake attacks. Rowlands remarks that the usual prey of the rattlesnake, the field mouse, complies with such conditions quite appropriately. However, the rattlesnake can be easily fooled with, for example, an artificially warmed imitation of a mouse. What is then the proper function of the snake's prey detection system? Rowlands applies his argument to the answer of this question and considers that we have to distinguish between the organismic and algorithmic proper functions of the snake's prey detection system. On the one hand, there is the *organismic* proper function of the mechanism:

> This is best specified in terms of the affordances of the environment to which that mechanism gives the

> snake sensitivity. Thus, the organismic proper function of the mechanism is to enable the rattlesnake to detect a certain affordance of the environment, namely *eatability*. This allows the attribution of content such as "eatability!", or "eatability, there!" to the rattlesnake. (Rowlands ibid., p.290-1)[32]

On the other hand, there is the *algorithmic* proper function of the system:

> The mechanism enables the rattlesnake to detect when the environment affords eating. It achieves this, however, by way of a certain algorithm, namely, the detection of warmth and movement.(...) the algorithmic proper function of that mechanism is to detect warmth and movement.(...) [Such a proper function] warrants content-attribution such as "warmth, there" or "movement, there" to the *mechanism*. (Rowlands ibid., p.291)

Rowlands makes a point that is crucial to his task of accounting for a teleological semantics, namely, that although the mechanism seems to represent *mice*, the detection of mice should be seen not as a proper function of the mechanism but as a Normal "consequence" of the mechanism fulfilling its organismic and algorithmic proper functions. At the organismic level of description what is important, for Rowlands, is that the snake eats, not that it eats mice; in other words, any appropriate source of food is satisfactory, since it allows survival. The mechanism is there because it allows the snake to survive by obtaining any sort of food. Likewise, the algorithm level gets the snake sensitive to warmth/movement, and therefore it cannot distinguish between mice, small rates, voles, squirrels, etc, nor need it distinguish these animals, since they are all eatable and equally important for survival: "To attribute to the mechanism the content 'mouse' or ' mouse, there', would be to attribute more semantic detail that is really there." (ibid., p.292). However, this precision is of secondary interest for me, since I want to use the distinction in accounting for different functional attributions rather than for content individuation:

> The proper functions of the mechanism are (i) to enable the rattlesnake to detect when the environment affords eating, and (ii) to detect when the environment exhibits warmth and movement. (Rowlands ibid., p.292)

[32] Here Rowlands makes a difference between eatability and edibility, the latter being what can be eaten by an organism, and the former adding the property of nourishment.

Summing up, Rowlands distinguishes two proper function attributions, one directed at the organism or agent, which is a proper function individuated by its selectionist advantage, and another attribution directed at the mechanism that allows such a function to be satisfied.

5.8.Conclusion

So far so good. Now I want to take advantage of both the notion of explanatory mismatch and that of Rowlands' distinction between two content attributions in order to develop my proposal.

First, we have seen that a mismatch can appear between the description of the faculty at one level of Grandpa's explanatory cascade, and the description employed at a lower level. We have argued that this mismatch does not compromise the robustness of the explanation; one system may accord to a task-level description even if the analysis of the task is not inherited by the mechanisms of the system.

On the other hand, we have presented Rowlands' identification of a difference in proper function attributions. Rowlands uses his distinction to sustain a difference in content attribution *for a specific proper function account*. Yet, even if Rowlands is actually interested in the issue of content attribution, specifically in solving the problems of transparency and indeterminacy -intrinşic in the way content is specified for a determined system-, Rowlands points to a genuine distinction.

However, Rowlands' identification is only the first step. The second is to see that both content attributions belong to different explanatory projects. Indeed, since the question is not just that there are two different content attributions; rather the crucial point is to see *why*. My aim will be to investigate the origin of the distinction between 'organismic' and 'algorithmic' properties, which will allow us to solve the problem of the explanatory mismatches within Grandpa's framework.

Chapter 6

Capacities, operations and functions

Paradoxes are useful to attract attention to ideas.

The paradox we found in Chapter 4, namely that cognitive systems may accord with a competence theory without internalizing it, has been partly resolved in the last chapter. I have argued that we can appeal to a distinction made in the discussion of naturalistic notion of function to the effect that we should recognise two sorts of *explanatory projects* underlying cognitive science theorising. One corresponds to the explanation of the way in which cognitive agents seem to comply with a class of functionalities that might, in turn, be selectively relevant. It is an explanation that couples an agent with its environment, and is characterized as a functional pattern to be satisfied according to a determined functional pattern that is what (partially) accounts for why the system is there. On the other hand, there is the project of explaining the processes within the system that account for the satisfaction of the task. My claim is that it is inappropriate to view both accounts as taking part in a single cascade explanation of a certain function at issue; rather we should view them as playing a role in separate explanations. This forces us to view them as conceptually independent; they are not *necessarily* but *contingently* connected. Rowlands makes the point of the mismatch between the functional demand and the contributions of the mechanisms responsible for satisfying it in the following way:

> Typically, our algorithms will only approximately carry out the functions specified at the ecological level. We have, for example, in our visual system, mechanisms whose function is the preservation of colour constancy in vision. Their function is to ensure that we can continue to discriminate an object's colour even when illumination changes. The mechanisms don't, however, work perfectly; your visual system can be fooled, for example, under sodium lightings. (Rowlands 1997, p.296)

The mismatch should thus be properly construed as two separate aspects of the faculty for understanding and producing the function in question. My view is that this distinction can have its place in a theory of functional attribution. The agent-in-an-environment account can be viewed as a full-blooded causal notion because it is a teleological notion, whereas the *other* can have a place as the object of a functional analysis of the system.

I shall call *cognitive capacity* the notion that gives content to the ecological, the agent-in-an-environment, account, whereas *cognitive operation* will be the label for the notion that corresponds to the explanation of the intrinsic processes of a system. In fact, one could see what follows as an attempt to provide a different technical notion of cognitive function than the one presented by Cummins (1983), as well as to account for the notion of *process* that McClamrock (1995) proposes. For clarity of exposition I will use the expression capacity* and operation* to refer to the notions that I have just introduced. My aim now is to develop such a distinction between explanatory projects that we have identified in the explanation of a system's cognitive capacities.

Two provisions before we proceed. First, I want to clarify that my proposal points only to a distinction in explanatory projects in cognitive science that may have been overlooked and which have obscured or misled some discussions in the field. In this sense, the distinction is orthogonal to the question of Intentional Realism,[33] since it holds no ontological commitment about the reality of folk psychological notions. In other words, it should be compatible with the two more popular positions on the issue: eliminativism and realism. It should be compatible with realism because the distinction capacity*/operation* need not deny that a certain folk notion, like happiness, belief or understanding English,

[33] I understand Intentional Realism as a realist stance about intentional notions such as beliefs, desires and the like. These notions characterize psychological states as having a specific content, as being about something of the world.

corresponds with an entity in the mind/brain. The only thing the proposal requires is the conceptual distinction between capacity* and operation* attributions. Similarly, the proposal should be compatible with eliminativism because even if it comes out that *no* folk notion can be attributed to the mind/brain, we shall still need to differentiate between the explanatory project of the agent-in-an-environment, and that of the cognitive architecture. Indeed, since *no* neurophysiological notion will ever explain why a certain population survives in contrast to another, when the sole difference between the two is an environmental variable.

Finally, the proposal is also intended to remain neutral about two other very popular issues within the philosophy of mind, namely, the distinction between narrow content and broad content, and the opposition between methodological solipsism and externalism. On the standard understanding of the notion of content, there are two ways in which content can be individuated. Contents individuated by reference to conditions external to the believer are said to be "wide" or "broad" contents. Beliefs whose content is individuated solely on the basis of what goes on inside the persons holding them are said to be "narrow" contents. On the other hand, the doctrine that defends the restriction of psychological explanations to what happens within the mind is known as methodological solipsism (Fodor 1981b) or autonomy principle (Stich 1983). It amounts more or less to asserting that cognitive psychology ought to restrict itself to postulating formal operations on those mental states that are directly related and within the mind: The idea is to explain behaviour in terms of the (narrow) causal relations among stimulus, hypothesized mental states and behaviour. It could seem as if my proposal had to commit itself to defend one of these positions. This would be a misunderstanding of my position, however, since the proposal laid out here concerns precisely the belief that some of the functional-attribution mismatches don't stem from opposite metaphysical positions in reference to content, or from opposite epistemological stances in psychology for that matter. Rather, the mismatches derive from different *explanatory aims*. That is, I submit the idea that there might be, at least, two different ways of looking at cognitive functions that are not mutually exclusive, since they focus on different properties of the system. Accordingly, each position is not to be considered incompatible with the other.

6.1.Capacity* as the agent-in-an-environment account

As I have just said, the class of capacities* corresponds to the specification of the environmental demands -functionalities- that cognitive agents can satisfy. These requirements are characterizations in terms of the properties of the organism as an agent-in-an-environment. Specifically, they must be, on the one hand, characterized by the theorist as being cognitive demands and, on the other, they must be *behaviours* of the agent that can be selectively relevant. A capacity* should in fact be characterized in terms of environmentally specified properties that needn't supervene on any particular localized proper subset of the properties of the overall cognitive system.

In the domain of cognitive processes, capacities* are classically understood as the class of *cognitive functions*. Specifically, when the theorist aims at explaining a cognitive function in this sense he aims at inferentially characterizable capacities (Cummins 1983), where there is a input-output condition and a rule of inference that describes the ability. Thus, to explain an inferentially characterized capacity is then to explain the capacity to conform to the characteristic inferential pattern. As we have seen all along, according to Cummins, to ascribe a function to some cognitive system is to ascribe a capacity to it that is singled out by its role in an analysis of some system's capacity. We explain a cognitive ability whose inputs and outputs are specified via their semantic interpretations. The capacity to add, we saw, is the capacity to produce as its outputs the correct sum of the inputs. The outputs must be interpretable as numerals representing the sum of the numbers represented by the numerals interpreting the inputs. Two inputs (outputs) count the same -i.e. as tokens of the same type- only in the case that they have the same interpretation. So long as the model of cognitive system is an information-processor, capacities specified by functional analysis are labelled information-processing capacities. This is the explanatory value of interpretation: we understand a computational capacity when we see state transitions as computations. In sum, a capacity is specified by giving input-output conditions, and what makes a capacity cognitive is that the outputs are cognitions, and what makes outputs cognitions is that they are cogent or, in terms of Cummins, epistemologically appropriate relative to inputs. Hence, outputs must be inferrable from inputs in an inferentially manner (the law specifying the capacity is a rule of inference).

I sustain such a characterization with one modification. Instead of understanding a cognitive capacity (or function, term that Cummins uses as an equivalent of capacity) as an ascription that is singled out by its role in an analysis of some capacity of *a containing system*, I take the notion as the identification of a demand that is satisfied by a system in an environment. The characterization of such demand would be nevertheless the same as in Cummins' framework. A specific capacity* of a system might be, for example, to perform arithmetical operations, such as addition, and therefore the requirement is for the system to do whatever is needed to master adding. Moreover, the demand is not restricted to the satisfaction of the function in extension (adding), rather, to what may be involved is the satisfaction of the demand under a determinate inferential pattern, which can be described as an informational-processing pattern, that is, a computation, information-preserving, systematic transformation of input to output. As a matter of fact we could describe a capacity* as an element of a system that maps an environment into a computation.

My proposal is then that a capacity* is individuated by a determined functional pattern (input-output mapping) for a *cognitive system* and an environmental context that constrains the function. This context can be subsumed under an adaptation of the notion proposed by Newell and Simon (1963, 1972) *task environment.* In *Human Problem Solving* Newell and Simon provided an analysis of how humans solve "intelligent" tasks. Their account was based on the characterization of the structure involved in problem solving. The characterization distinguished between the demands of the environment and the psychology of the subject. They argued that any instance of problem solving involved some constraint in the environment. Accordingly, they proposed that an account of problem solving abilities should have an environment specification related specifically to the task at hand. The term 'task environment' is thus used to refer to an environment coupled with a goal, problem or task, for which the motivation of the subject is assumed. The demand of a task environment is therefore a constraint on the behaviour of the problem solver that *must be satisfied in order that the requirement be attained.* As Newell and Simon put it, however, the environment per se does not make demands; rather, the problem makes them via the problem solver's commitment (or necessity I would add) to attain the solution. The features of the environment that give rise to these demands constitute the relevant structure or texture of the environment. The structure is given by the task invariants of the problem at hand. In this regard, for Newell and Simon, there are task invariants that must be specified independently of the subject that must satisfy the task. Marr (1977, 1982) went along with this idea when he considered that a critical feature of the task described as "addition" should be determined by the constraints of the problem. These invariants are those properties shared by all paths in an environment leading to a goal. For example, the task invariant for the tic-tac-toe game is making three in a row.[34]

What we should keep in mind is that the bottom line for Newell and Simon is to construe a "deterministic" framework where a cognitive system produces a stream of behaviours when placed in the environment and given the goal (implicitly -the behaviour of processes will carry implicitly the goal- or explicitly represented). Additionally, such an account *must* be independent of the description cognitive mechanisms that make it possible.

We should then see the issue as that of an agent that needs to satisfy a capacity* in order to attain its goals, be they selectively efficient, or simply contingently convenient. It is for this reason that it could be useful to view the capacity* as a sort of *demand* of the environment that the system faces to achieve its goals. This would be the first condition to subserve teleology accounts of functions. It is the condition of fitness understood as the capacity to satisfy the goals of the system, which can have selective value, although the notion of capacity* does not automatically imply the selective value of the capacity* identification. For that we need the *historical* condition. Then, we could characterize cognitive capacity* in the following way:

> **(Cognitive) Capacity*:** The (cognitive) task that system *S* undertakes in environment *E* to satisfy demand *D*.

And this is developed in:

[34] The details of Newell and Simon analysis of task environment are of little importance here because I will not endorse them, but I will give a sketch of them to understand the rationale behind the notion of task environment. Newell and Simon analyse the task environment in different constituents. The *problem-space* is defined in terms of the solver's representation of the problem, and it comprises the set of all problem-states that could possibly be reached by the available operators. An *operator* is a way of transforming one problem-state into another. To narrow down the problem-space, they introduced a *means-end analysis*, which converts the overall problem into a series of goals, sub-goals, on distinct hierarchical levels.

Demand specification: The goal to be satisfied.

Task specification: A task is a behavioural regularity which is characterized in a mapping that relates inputs into outputs in a given pattern *P* such as *P: I→O*.

Environment specification: The structure of environment *E* relevant for the capacity* fulfilment.

Capacity fulfilment:* The structure of the task satisfaction.

These notions will be developed below; there are some points that should be clarified, though. First, I understand a *task* in the way that Cummins understands functions, as inferentially specified behaviours, that is, behaviours that can be characterized as an input-output mapping and a rule of inference that describes the behaviour.[35] Such a characterization explains cognitive capacities via the semantic interpretation of inputs-outputs and intermediate states.

Secondly, a capacity* characterization must be made considering that *optimality* is the default assumption (see Chapter 1). This strategy normally over-idealizes the "functional" problem by presupposing that this function is optimally executed by the cognitive system. This condition is needed because a capacity* understood as a satisfaction of a demand posed by the environment is a *normative* requirement, that is, it is the way how the system *must* behave to achieve the goal. Only those ways that actually satisfy the demand will count as capacities*. Indeed we, as theorists, are not interested here in how systems do actually behave to satisfy the demand, or in what proportion for that matter, nor

[35] To be more precise the individuation of a capacity* is behavioural but not behaviouristic. It is behavioural because it concerns effects which environmentally constrained and attributed not only to the brain, but to the whole organism and whatever resources it resorts to satisfy the function. It is not behaviouristic because it takes behaviour in a broad stance, as we considered in the last chapter in reference to Kim's (1996) definition of behavior, that is, including physiological reactions and responses, bodily motions (typing, greeting, shopping), and actions not involving bodily motions (reasoning, guessing, calculating). In the specification of a capacity* we can, for example, include a description of informational inferences that the system must perform to satisfy the capacity*.

are we interested in revealing whether its satisfaction is actually performed in other than optimal ways; rather, we are, in fact, interested in designating the optimal way to satisfy the demand.

Finally, a capacity* is also a contingent notion. It depends on how the world is, or was, since it has been environmentally constrained. As a matter of fact, it is precisely for that reason that it is useful to distinguish between two explanatory projects in cognitive-function attributions. The causal relevant properties are normally concurrent with the environmental situation in which a given population dwells. What makes an individual, or a group of individuals, survive or fit in depends precisely on the characteristics of the environment that surround them and with which they interact. We need actually a notion that comprises an unbound class, and the class of capacities* belongs, in fact, to an open-ended class. We can always find instances of capacities* that are absolutely new or that have been lost in the course of evolution. The point is made by Newell and Simon:

> Now if there is such a thing as behavior demanded by a situation, and if a subject exhibits it, then his behavior tells more about the task environment than about him. We learn about the subject only that he is in fact motivated towards the goal, and that he is in fact capable of discovering and executing the behavior called for by the situation. If we put him in a different situation, he would behave differently (Newell and Simon 1972, p.53)

Similarly, a cognitive scientist might reveal new capacities* of the cognitive system when he submits subjects to perform certain experiments in a lab. There the subject has to attain certain goals, and hence the function studied is transformed into a functional requirement. The satisfaction of such a requirement can become a new capacity* by simply providing the goal the subject is looking for, such as economic compensation. Alternatively, if the world turned out to be organized in a way that men had to solve anagrams to obtain women's favours, then such requirement would make a capacity* of its satisfaction. In addition, this property will let us use the notion as a connection between etiological and dispositional accounts of function.

6.2.Operation* as the cognitive architecture

I take as a fact that a cognitive system is made up of a number of fundamental processes. We could say that the entire behaviour of the system is, in fact, compounded out of sequences of basic processes. By basic I mean that they correspond to a complete and characterized (cognitive) activity defined in the context of a theory The idea is that there must be a sufficiently general and powerful collection of operations to compose out of them all the macroscopic performances of a system. Then, an operation* of a cognitive system is that process[36] that accounts for the different functional properties of a cognitive system. This idea is obviously not new. It could be related to Pylyshyn's (1984) notion of *functional architecture*. For Pylyshyn a cognitive system is endowed with a number of built-in processes that define its functional architecture, by opposition to the "cognitive" one:

> Generalizations expressible in terms of the semantic content of representations are referred to in this book as "semantic-level generalizations," whereas generalizations expressible in terms of functional properties of the functional architecture are referred to as "symbol-level generalizations". (1984, p.32)

Each architecture aims at different explanatory models:

> The second criterion attempts to draw the architecture-process boundary by distinguishing between systematic patterns of behavior that can be explained directly in terms of properties of the functional architecture and patterns that can be explained only if we appeal to the *content* of the information encoded. (ibid., p.xvii)

Pylyshyn compares the notion of providing the functional architecture of a system with that of providing a manual that defines some particular programming language. Pylyshyn distinguishes between the functional architecture and the program that can run on it, and the difference is tied to the existence of a "language of programming" (ibid., p.94), regardless of the many levels of symbol processing at which computation can be viewed. Then, Pylyshyn's functional architecture is made up of a set of primitive functions, which

[36] I understand a process in the usual way, as a chain of events causally related.

correspond to the steps of an algorithm:

> If we know the architecture -that is, if we are given the set of primitive functions- we can determine whether a particular algorithm can be made to run on it *directly*. For an algorithm to run directly on a certain architecture, the architecture must contain primitive operations whose behavior is formally isomorphic to each elementary step required by the algorithm. In other words, for each elementary operation in the algorithm there must already exist some operation in the functional architecture whose input-output behavior is isomorphic to it. If to get the algorithm to execute, we must first mimic the input-output behavior of each elementary step in the algorithm, using a combination of different, available operations, we would not say the algorithm is executed *directly* by the available operations, that is by that virtual machine. (ibid., p.115)

The architecture must form, for Pylyshyn, a sort of cognitive "fixed point", so that differences in cognitive phenomena can be explained by appeal to operations among the fixed set of operations and to the basic resources provided by the architecture.

However, the notion of operation* I wish to propose has a fundamental difference with Pylyshyn's notion. Pylyshyn applies "functional architecture" to those basic information-processing mechanisms of the system for which a nonrepresentational or nonsemantic account is sufficient (1984, p.xvi). But as I have attempted to point out, operations* must be characterized by the locally computational properties *which must be relationally specified*, such as the kinematic principles that underlie our perceptual system abilities for reasoning about objects, which we saw in Chapter 4. According to Pylyshyn, processes or mechanisms needn't apply to an inferential account, a condition that I believe we cannot avoid if we want to give a complete account of cognitive resources at hand. Pylyshyn gives an example of his idea about functional architecture:

> For example, the successor operation we might call AFTER, which is irreflexive, antisymmetric, transitive, acyclic and connected over a specified set o names. Such a syntactic operation in the *functional architecture* can be interpreted freely as any relation in the semantic domain (...) In using such a formal operation in this way to realize an interpreted rule, we automatically inherit the formal properties of the built-in primitive relations. Thus we need not represent explicitly such properties of the relation as its reflexiveness, symmetry, transitivity, noncyclicality, and so on. (1984, p.100, my italics)

However, in my account of operation*, the primitive process represented by thẽ operation AFTER has the contents that the mathematical relation [after] establishes, and such a relation can be counted as an *inferential* step. In this regard, I specify the notion of operation* as a process in the mind/brain which can be given a fully representational account and at the same time counts as wired-in, preestablished in the structure of the brain. As a matter of fact, I do not see the need to drain content out of the functional level.

As far as I consider operations* to be the elementary processes of the system's cognition, I will use the expression *cognitive architecture* to refer to the complete and articulated structure of cognitive operations*. This will help me to distinguish my notion from that of Pylyshyn, which is specifically used to segregate "cognitive", or representational, accounts from functional ones. Then, we can specify the criteria to constrain the identification of operations*:

> **(Cognitive) Operation***: An inferentially characterizable process of a cognitive system's mechanism.

In this sense an elementary process is individuated by two elements:

> *Process:* A process is ¢haracterized by a mapping that relates inputs into outputs in a given pattern *P* such as *P: I→O.*
>
> *Mechanism:* The cognitive structure that implements the operation*.

An operation* can yield more complex operations* out of the composition of basic operations*. These complex operations* will be called second-order operations*. Provided that all what I am going to say about operations* concern first-order and second-order operations* in the same way, I will omit the distinction. As a matter of fact, it will be only necessary to refer to second-order operations* in reference to the strategy of functional analysis, since it is the way to decompose the system into operations* independently of the capacity* fulfilment. This will imply, for instance, discharging second-order operations* into first-order ones.

The notion of operation* just presented has the following properties. Provided that there is a one-to-many relation between a mechanism and the computations it can perform, a mechanism can instantiate one or more operations*. For example, any system implementing a complex computation will simultaneously be implementing many simpler computations. In general, there is no canonical mapping from a physical object to "the" computation it is performing. We might say that within every physical system, there are many computational systems. The question of whether a given system implements a given computation is still entirely objective. It has been argued, for example, that any given instance of digestion will implement some computation, as any physical system does, but that the implementation of this computation is in general irrelevant to its being an instance of digestion. With cognition the claim is that it is in virtue of implementing some specific computation that a system is cognitive. In other words, there is a certain class of computations such that any system implementing them is cognitive. We might go further and argue that every cognitive system implements some computation such that any implementation of the computation would also be cognitive, and would share many specific mental properties with the original system. So the fact that a mechanism can perform more than one computation is no flaw in the current account. The only thing we have to do is to narrow down the operation* that interests us in the explanation at stake. To this very limited extent, the notion of implementation is "interest-relative".

Secondly, an operation* contributes to the satisfaction of capacities*, or better yet, the demands that are derived from a capacity* specification. In other words, a cognitive system satisfies a demand by engaging operations* that contribute, in appropriate ways, to satisfy a capacity* fulfilment. It is in this way that we can conceive an operation* as a constituent of demand satisfaction, even though it is not a *proper part* of a capacity*. The way in which this is realized can take on different forms. The engagement of operations* may correspond, first, to a preestablished program, such as the way in which the brain engages elementary processes in the early stages of visual perception, and/or it may also be consistent with new ways of elementary-process composition that can be triggered by the environment, such as the competence needed in reasoning and predicting objects' motions; there is then the possibility of creativity, such as the one that "playing chess" may require.

Thirdly, there is also a one-to-many relation between an operation* and the capacities* to which it can contribute. That is, each operation* may contribute to a capacity*, to more than one capacity* or to no capacity* at all. Ullman sketched the idea in the following way:

> (...) in contrast with a counting network that is specially constructed for the task of detecting a prescribed number of items, the same elementary operations employed in the counting routine also participate in other visual routines. (Ullman 1996, p.308)

Operations* are intrinsic properties of systems and some of them might exist precisely because they have been *selected for*. In this regard, we can say that a mechanism was selected for the performance of an operation* if past instances of that kind of mechanism performed the operation*, where such performances increased the relative adaptedness of the organism, and where this increase in relative adaptedness explains why this kind of organism and this kind of mechanism exists in the population today. Yet, other mechanisms might be *selected of*, if past instances of the mechanism were maintained, or caused to proliferate, as a consequence of selection for.[37] As it is standardly conceived, selection of is a consequence of selection for. Moreover, some operations* might have been selected for the performance of *more* than one capacity*: capacities* are organism-environmental dispositions which can be fulfilled by different processes.

Finally, the number of elementary processes, of operations*, is limited for each system. The fact that the cognitive architecture of a system is fixed by the physical properties of the system implies that, in turn, the catalogue of operations* depends directly on the genetic programme of the species. In other words, every cognitive system has its potential *operations** preestablished, in the same way that it has its muscular and hormonal structures. Obviously, the unfolding of the genetic program has to be triggered by environmental factors, but the catalogue of possible operations* is "determined" by the genetic program. The sort of operations* that a cognitive system can perform does not thus depend on the nature of the tasks that a system can perform, but on the "cognitive

[37] "Selection of" and "selection for" is a distinction introduced by Sober (1984). An example can clarify the distinction. Consider a filter that will allow only balls of 1cm or less in diameter to pass. Suppose that all such balls are green. In the group that passes through the filter there is selection *for* size, but selection *of* both size and colour.

architecture" embedded in its brain.

Exactly what kinds of processes qualify as primitive (hence explanatory) operations, that is, as operations* is a difficult question. I agree with Pylyshyn that the question of which processes should be comprehended in what he labels functional architecture structure is an open question. He clearly accepts that a basic process is "an instance of problem-solving" (1984, p.109) but also accepts that such consideration is insufficiently precise.[38] As a matter of fact, I lack any robust criteria to support any operation* identification. But that was not the task I set out to do here.

The identification of true basic cognitive properties is perhaps one of the most difficult tasks in cognitive psychology. Therefore, the only thing that we can do is to provide a tentative list of possible candidates for operations*, that is, candidates for the primitive functions of the cognitive system. Pylyshyn constructs his list as abilities for symbol manipulation, like storing and retrieving symbols, comparing them, and treating them differently as a function of how they are stored. My view is that symbol manipulation is *one* of the possible operations*.[39] Symbol manipulation is perhaps one of the most relevant, powerful and useful operations*, since it allows a framework and a tool where

[38] A great deal of Pylyshyn's work in *Computation and Cognition* is actually devoted to the construction of criteria to distinguish the functional architecture from what he calls the cognitive architecture. He first considers that the criterion might be that no cognitive operator is considered primitive unless it can be realized on a computer, yet he concedes that the condition is not *sufficient*, since it relies on the realizability of the process on a functional architecture we have no reason to believe is correct. Another criterion dismissed is that of capturing generalizations, since merely setting our sight on what seems the most convenient level of abstraction is clearly also insufficient. Generalizations that no one has any idea how they can possibly be realized by some mechanism are interpreted as *descriptions* of phenomena, not *explanations*. As he then recognises there is no "simple and sovereign" method available that will ensure the correct, basic architectural functions we have hypothesized. He even asserts that "primitiveness is a theory-relative notion", and for that reason he advances two empirically based criteria. The first criterion he calls *complexity equivalence*, which is similar though weaker to the intuitive notion of the "same algorithm". The complexity of a process is related to the notion of the amount of resources used in a computation or how much computation a process does. Two different programmes are viewed as instantiating complexity-equivalent algorithms if there exists a certain kind of topological relation between them. If every linear series of nonbranching operations in a program can be mapped into a single operation with the same input-output function in the second program, then the programs are complexity equivalent. The second criterion is based on the notion of *cognitive penetration*, which is based on the assumption that the primitive operations of the functional architecture must not depend on goals and beliefs, hence, on conditions which, there is reason to think, change the organism's goals and beliefs. Therefore the operations that change in relation to change of beliefs should be considered *not* part of the functional architecture.

[39] We must distinguish here the metaphysical thesis that cognition is just symbol manipulation, from the thesis that I develop here, namely, the thesis which holds that, regardless of what the true nature of the medium in which cognition takes place is, one of the possible cognitive, or functional, processes of a cognitive system is "symbol manipulation".

many tasks can be accomplished (symbol manipulation can subserve, for example, the task of deriving logical arguments). For this reason, I consider that a characterization of the basic processes of a cognitive system should go much further than simply specifying them as symbol-manipulation processes. A robust description of basic processes would have to include characterizations such as the ones we saw in Chapter 4 in the case of, for instance, visual perception. My view is that the way to identify and characterize the elementary operations that correspond to the notion of operation* is an empirical question that cannot be established beforehand.

6.3.Surrogate operations*

In Chapter 2 I discussed Davies' notion of psychological reality. I presented objections to the notion of two sorts; one objection was directed at the Mirror constraint asserting that it was too weak a criterion to distinguish the reality of one theory against another; another objection was levelled at the scope of the notion considering that it was too strong since it left out certain conceptually possible and empirically attested possibilities. One such possibility was what I labelled the *extended mind*, following Clark and Chalmers' (1995) characterization. The idea was that when satisfying some tasks, a part of the world/environment functions as a process which complements those performed in the brain. In other words, we have to describe the efficiency in some cognitive function we may be forced to describe a competence that has to be attributed not only to the internal mechanism of the system, but to some resources provided by the environment. The examples go from the aid of pen and paper to execute arithmetical operations, up to the very use of language to derive arguments, including much more simple cases as the use of the physical structure of a cooking environment (grouping spices, or "tools") as an external memory aid, or the use of aids in some games.

I have argued that the attribution of the theory to a cognizer required the inclusion of such "extended" processes, provided that without this inclusion we would not be able to attribute an *appropriate* theory about some competence. The fact is that we need to *include* such processes in the explanation of the competence, of how a cognizer is capable of "complying" with the operation without the internalization of some of the processes established in the theory. Indeed, for suppose that the actual way we, for example, perform arithmetical operations require the postulation of certain processes that are not part of the mind/brain but of some external aid. As it were, the *correct* theory of arithmetical competence, the psychological appropriate theory, needs the external aid as part of its deployment. Then if we characterize the arithmetical competence of such a theory, we also have to include the "external steps". An account of such competence without the postulation of, for instance, the processes instantiated by the abacus would not get off the ground.

These possibilities were related to the notion of *epistemic action* introduced by Kirsh and Maglio (1994). Succinctly, epistemic actions are physical *external* actions that an agent performs to change her own computational state in order to make such mental computations easier, faster, or more reliable. The notion has its origins in some examples within cognitive science, especially concerning the actions that manipulate external symbols. In arithmetic (Hitch 1978) or in navigational skills (Hutchins 1995) various intermediate results of certain computations are recorded externally to reduce cognitive loads. Kirsh and Maglio have shown that these sort of actions occur without the need of symbol manipulation, especially in the case of the video game Tetris. More precisely, Kirsh and Maglio use the term epistemic action to designate a physical action whose primary function is to improve cognition by:

(1) Reducing the memory involved in mental computation, that is, space complexity.

(2) Reducing the number of steps involved in mental computation, that is, time complexity.

(3) Reducing the probability of error of mental computation, that is, unreliability.

The way Kirsh and Maglio accommodate this notion in an information-processing framework is by regarding epistemic actions as actions designated to change the *input* to an agent's information-processing system.

In order to account for such cases, I propose to consider these processes as a sort of processing "delegation" of the cognitive architecture into the world. In other words, I shall treat such processes in the same way that we treat internal capacities, just as if they were properly executed by the cognitive system, but adding the qualification of not being realized *by* the cognitive system. Therefore, I shall call such operations* *surrogate* operations*, a notion which comprises all those informational-processing processes that are not to be attributed to the intrinsic mechanisms of the cognitive system. An analogy could help us to come to terms with the notion. Suppose that we individuate one function of our ancestor *Homo habilis* as the capacity to hunt. We give an analysis of the function in different steps, such as: i) Tracing the prey, ii) cornering it and iii) killing it. Suppose that in step three we need to postulate the presence of "arms", without which the function cannot be manifested, that is, the prey cannot be killed by giving it presents or kisses for that matter. Then in this situation the *action* of "killing" could be labelled a surrogate operation* of the different actions undertaken by the *Homo habilis* to exert the capacity to hunt. This sense of surrogation is the same that I want to use in characterizing cognitive operations* in need of an external aid. Summing up:

> **Surrogate operation***: An elementary (cognitive) process which is instantiated partially in a cognitive system and partially in the environment.

As operations*, a surrogate operation* is individuated by two elements:

> *Process*: A surrogate operation* is characterized by a mapping that relates inputs the outputs in a given pattern *P* such as *P: I*→*O.*
>
> *Instantiation*: A surrogate operation* is instantiated in a physical mechanism that couples a cognitive system with some relevant structure of the environment.

Surrogate operations* have a number of properties that differentiate it from the notion of operation*. First, the characterization of a surrogate operation* is dependent on the role they play in the capacity* fulfilment. They don't have the computational properties intrinsically, but insofar as they take part in a capacity* fulfilment; their characterization is a matter of interpretation of the role they play in a capacity* fulfilment.

Second, the fact that a mechanism that instantiates a surrogate operation* can correspond to any type mechanism -that honours the computational properties required- makes the notion subject to a multiple realizability condition *within a given cognitive system and environment specifications*. In other words, the realization of surrogate operation* is contextually specified. There is then a one-to-many relation between the computational properties and the mechanisms it can be described to realize. Therefore, its specification depends on the context of the capacity* fulfilment.

Third, the selective value of a surrogate operation* must be attributed to an operation* of the cognitive system that instantiates the ability of the cognitive system of "taking profit" of such type of surrogate operations*. In the same way as we cannot attribute any "proper function" (in Millikan sense) to the axe used by *Homo habilis*, we cannot attribute proper functions to surrogate operations*.

Finally, the number of surrogate operations* is not limited. As we have defined the notion, the functional uses we can make of the world are open-ended, and only dependant on the use that a cognitive system can make of it.

6.4.Demands and environments

The notion of demand and that of environment presented in the capacity* specification can be filled out by two recent, yet remote, projects undertaken in cognitive science: Evolutionary psychology (see for example Barkow, Cosmides and Tooby 1992; Cosmides and Tooby 1987; Cosmides and Tooby 1994) and situated action (see for example Greeno 1989; Lave 1988; Suchman 1987; Winograd and Flores 1986). Evolutionary psychology is a research field that attempts to take advantage of some insights of the theory of selection. Specifically, adherents to this approach contend that there is a functional mesh between the design features of organisms and the adaptive problems that they had to solve in the environment in which they evolved. By understanding the selection pressures that humans faced in evolution, they try to figure out the design of the information-processing mechanisms that evolved to solve these problems. More concretely, evolutionary psychologists believe that the best way to discover the mechanisms that underwrite our psychological capacities is to first discover the functions that our psychological capacities fulfill. They further assert that the best way to discover the functions of our present psychological capacities is to first discover the *adaptive* functions that our distant ancestors possessed, and this implies appealing to what is known as the *functional mesh* (Davies 1996). Functional mesh refers to the adaptive fitting between organismic traits and environmental demands that resulted from evolution via natural selection. In sum, for evolutionary psychologists, the best way to discover psychological functions is via adaptive functions:

> To discover the structure of the brain, you need to know *what* problems it was designed to solve and *why* it was designed to solve those problems rather than some other ones. (Cosmides and Tooby 1994, p.47)

The fundamental notion for evolutionary psychologists is therefore that of "adaptive problems". These refer to demands of the environment whose solution promotes reproduction, such as detecting predators, foraging (hunting and gathering), resource competition, parental care, dominance and status, inbreeding avoidance, etc. These are the environmental demands that I would include in my framework of capacity* with a slight difference. The problems I intend to consider in a capacity* specification *need not be adaptive*. As we will see below, for that to happen we need to add the historical condition which will transform the capacity* into a proper function in terms of Millikan. Accordingly, *my* demands are the class of *possible* adaptive functions.

Situated action, on the other hand, approaches cognition by studying how certain structures of the world constrain human behaviour. Adherents to situated action believe that human knowledge cannot be divorced from the world; the mutual accommodation of people and the environment matters to understand how cognition emerges. Situated action emphasizes the importance of historical influences, social interaction, culture and other aspects of the environment.

According to both fields of research, environments will have to be constrained by a variety of factors. These correspond both to the particular evolutionary history of the system to be studied, and to the "computational affordances" that the environment presents. The latter refers, for example, to which parts of the environment can function as a surrogate operation*. Situated action has studied, for example, environments such as that of human-computer interaction (Winograd and Flores 1986; Kirsh and Maglio 1994). Others have studied more natural settings such as that of navigational abilities (Hutchins 1995). The evolutionary history refers, on the other hand, to the particular contexts in which systems have evolved. In this regard we might be interested in variables such as the following (Cosmides and Tooby 1994): Ancestral hominids were ground-living primates; omnivores, exposed to a wide variety of plant toxins and having a sexual division of labour between hunting and gathering; mammals with altricial young, long periods of biparental investment in offspring, enduring male-female mateships, and an extended period of physiologically obligatory female investment in pregnancy and lactation. They lived in small nomadic kin-based bands of perhaps 20-100; they would rarely have seen more than 1000 people at one time, and so on.

Provided that a full development of these issues would require a thorough and complete assessment, and provided also that my aim was specifically to make the theoretical point of the need to treat such issues, I will leave the question here. I essence my position is to assume the empirical research strategy undertaken by both evolutionary psychology and situated action, setting aside the additional considerations made by their adherents concerning the inability of more traditional approaches within cognitive science to account for human cognition. As a matter of fact, the hypothesis that I am presenting, that of legitimizing two different explanatory projects in cognitive theorising, allows us to accommodate both the traditional approach, Grandpa, and that of evolutionary psychology and situated action.

6.5.Functional composition

How is a capacity* satisfied? My thesis here is that a capacity* is achieved by the engagement of a number of (one or more) operations* and (one of more) surrogate operations* in a specified way within an environment. This is what I termed "capacity* fulfilment". As Ullman has put it:

> [I]n perceiving a given spatial relation different strategies may be employed, depending on various parameters of the stimuli such as the complexity of the boundary, or the distance of the X form the

> bounding contour. The immediate perception of spatial relations often requires, therefore, selection among possible routines, followed by the coordinated application of the elemental operations comprising the visual routines. (Ullman 1996, p. 311-312)

The way to clarify what I mean exactly by a capacity* fulfilment is to resort to the notion of virtual machine. The conjunction of operations* and surrogate operations* in an environment can be seen to form a **virtual machine** that executes, performs or computes the *capacity** in question. Virtual machine is a notion drawn from computer science, but used by some cognitive science theorists (cf. Pylyshyn 1984, Boden 1988, Clark 1989). Basically, a virtual machine is -in computer science- a machine (either abstract or physically embodied) that can be programmed in such a way that its behaviour mimics the behaviour of something else. In this sense a virtual machine refers to the functional architecture (understood in Pylyshyn terms) plus the interpreter of the program that must run on it. For example, a computer running a word-processing program is a "virtual word processor", although different programs can turn it into a virtual calculator or a virtual arcade game. The same virtual machine may have a variety of underlying physical implementations, as is the case when differently construed computers nevertheless offer the same capabilities and screen displays. A virtual machine for Pylyshyn (1984) is in fact its functional architecture, and therefore it would apply here to the notion of *operation**. However, I shall use the notion of virtual machine in a slightly different manner than Pylyshyn. I think that if we interpret the notion of virtual machine as "the machine the programmer can *think* is being used" (Boden 1988, p.162), we can apply the notion to the specification of how the cognitive system, plus whatever aids it resorts to use in, satisfies the *capacity** that is required. And indeed Clark uses it more or less in that sense:

> [A]t the very least, it is surely a mistake to think (...) in terms of one task, one cognitive model. For (...) our performance of any top-level task (e.g. mathematical proof) may require computational explanation in terms of a number of possibly interacting *virtual machines*" (1989, p.151, my italics)

In other words, we can view the virtual machine as the emulation of whatever information-processing model that can account for the capacity*:

> The thought I develop in this chapter concerns a possible multiplicity of virtual cognitive *architectures*. The idea is that for some aspects of some reasoning tasks, we might be forced to emulate a quite different kind of computing machine. For example, to perform conscious deductive reasoning, we might emulate the architecture of a serial Von Neumann machine. (...) I endorse a model of mind that consists of a multitude of possibly virtual computational architectures adapted to various task demands. Each task requires psychological models involving distinctive sets of computationally basic operations (Clark, ibid., p.128-129)

Even Pylyshyn entertains a version of this idea:

> The distinction between directly executing an algorithm and executing it by first emulating some other functional architecture is crucial to cognitive science. It bears on the central question of which aspects of the computation can be taken literally as part of the model(...) From the point of view of cognitive science, it is important to be explicit about *why* a model works the way it does and to independently justify the crucial assumptions about the cognitive architecture. That is it is important for the use of computational models as part of an explanation, rather than merely to mimic some performance (1984, p.74)

The bottom line is therefore that some capacity* fulfilment can be described in terms of a "virtual machine" which may comprise a number of different elements: the cognitive system as well as certain external aids (pen and paper, an abacus, etc.). It is also important to see that both operations* and capacities* can be seen to work at the same level of analysis within Grandpa's explanatory framework. In some idealized form of analysis of top-level descriptions, at the task level, the same *type* of generalizations, vocabularies and laws may apply to both.

How is a given capacity* fulfilment going to be satisfied? My hypothesis is that a capacity* is achieved by a process of what I will call **functional composition**. Specifically, I propose that there be a process that can be described as a capacity* fulfilment by the theorist and which is composed of elementary processes, i.e. there is determined decomposition of the capacity* in functional constituents; hence the causal effects of the capacity* must be explainable in terms of their composition. These elementary processes are characterized according to a functional analysis of the capacity*, that is, the decomposition of the capacity* into simpler capacities. As I have stipulated the notion of capacity*, its fundamental processes need not be attributed to the system. Hoˇwever, the composition of such elementary processes is necessary to construe the capacity, and must therefore be satisfied by the operations* and surrogate operations* that the cognitive system provides. Actually, the relation between the operations* and the elementary processes of the capacity* fulfilment can be parallelled with that of *role* and *agent*. One possibility can be that the role that each operation* and surrogate operation* plays in that satisfaction neatly corresponds to one elementary process; yet, as we have seen, it is possible, and in fact very usual, that the capacity* cannot be discharged in the set of elementary processes of a cognitive system. This is the bottom line of the distinction between capacity* and operation*. Therefore, for each elementary process of the capacity* we could need one operation*, more than one operation* or the realization of a surrogate operation*. In short, an explanation of a capacity* would have the following form. For a system S, a capacity* F, an operation* A, and an environment E:

i) S is there because it does F in E

ii) F is satisfied in E due to A

iii) F has analysis $f_1, f_2, ..f_n$ and·A has analysis a_1, a_2, $...a_n$

iv) The role of a_1 in E is to perform f_1, f_3

My belief is that the role that each operation* or surrogate operation* plays in a capacity* fulfilment corresponds to the notion of *function*, and in my view this actually corresponds to the way in which the term is normally conceived in discussions of biological functions. I will therefore use the expression function* to refer to this use of the term:

Function*: The role that a given operation* or surrogate operation* plays in a capacity* fulfilment.

Before we move on, it might be useful to make some clarifications. First, a capacity* fulfilment account describes a causal structure. The virtual machine has a mode of composition, a syntax (that is, there are a set of relations among its constituents that constitute its structure). A capacity* fulfilment is the causal product of operations* articulated in a functionally determinate way, i.e. how the states are causally determined by their constituents is functionally well-specified as are their outputs which are causally determined as a result of their causal composition. Second, the independence of the notion of capacity* from that of operation* implies the independence of both accounts. One consequence is that operations* might be functional constituents of other capacity* fulfilments. Another upshot is that a capacity* may not be a product of a linear decomposition in operations*. Operations* can be engaged, for example, in redundant ways, or may contribute only partially to the satisfaction of capacities*.

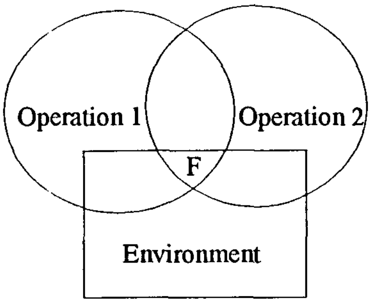


Figure 6.1. Capacity* (F) fulfilment by the conjunction of operations* in an environment.

Figure 6.1 shows an abstract picture of how a given capacity* fulfilment can be satisfied without having to discharge the analysis of the capacity* into simpler operations* attributed to the cognitive system, that is, how the capacity* can be satisfied without accordance with the cognitive system to a functional analysis. Each of the operations* can be seen as contributing to the description of the "capacity", and the theorist might be viewed as "taking a picture" of the processes going on in the system. The fact that the satisfaction of the capacity* can correspond to the deployment of one operation*, and therefore obtain an equivalence between an operation* and a functional requirement, so much the better. But the description of the capacity as a capacity* points to other explanatory goals such as the description of the operations*. Figure 6.1. also depicts the way in which a specific environment can constrain a capacity*. Furthermore, it shows that they should not be seen as elements of different levels, but simply as parts of different explanatory projects. Obviously, there are many other ways in which a capacity* can be fulfilled.

How is it possible that a certain combination of independent operations* may satisfy a capacity* that can be described in an informational-processing way, that is, as a pattern of transformations from state to state? The contribution of each operation* has been abstractly defined as a partial contribution to the fulfilment of the capacity*. The thesis is that each operation* may contribute in the satisfaction of one or more informational-processing steps of the functional description. A combination of operations* might redundantly contribute to one capacity*. The idea is that when we fulfill a capacity* there might be different cognitive processes subserving one functional requirement in the brain, so that, in general, operations* will not cut capacities* at their joints. Yet the capacity* must be satisfied *as if* there were some system that executed the function in the way described. In a certain sense we could say that it is only *in virtue of* achieving such a pattern that the capacity* is accomplished. As figure 6.2 shows, particular operations* may contribute in redundant and partial ways to the accomplishment of the functional pattern. In this case, the capacity* can be described as an input, two intermediate states -S1 and S2- and an output. This capacity* is fulfilled in this case by the contribution of different operations* that assume a role in the capacity* fulfilment. Operation 1 contributes to the completion of step 1, Operation 2 contributes to the transition from the input to step 1 and step 2, and Operation 3 to the attainment of step 2 and transition to output.

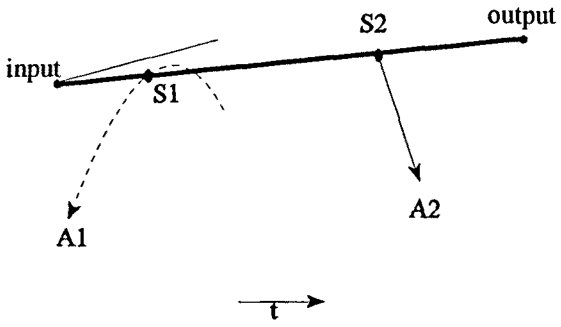


Figure 6.2. Operations* (A1 is represented by the broken line, A2 by the straight thin line, and A3 by the dotted line) contribution to the satisfaction of a capacity* (represented by the thick straight line between input and output). Each operation* contributes to the execution of some of the steps (represented by the points S1, S2,...) that intermediate between input and output.

Are there examples in cognitive science theorising that take this approach? Godfrey-Smith (1996, p.22) makes the most general case when he says that "cognition may well be a collection of disparate abilities and traits, each with different evolutionary history". Allport (1989) is one of the theorists that has provided experiments and analysis in this respect, focusing on the domain of attention. Allport points out that any complex system (humans, and other species) is a multi-functional organism:

> [W]hose subcomponents (sense organs, effectors, cognitive subsystems) are not in general uniquely dedicated to particular goals or to particular categories of action. Subcomponents must therefore be selectively engaged and coordinated to implement particular activities and particular goals (1989, p.649-650).

His idea that "there are many different functions of attention" (p.653) is, in my view, consistent with the idea that I have presented. Dehaene presents this very same idea from a more neurobiological perspective:

> (...) when our brain is confronted with a task for which it was not prepared by evolution, such as multiplying two digits, it recruits a vast network of cerebral areas whose initial functions are quite different, but which may, together, reach the desired goal. Aside from the approximate accumulator that we share with rats and pigeons, our brain probably does not contain any "arithmetical unit" predestined for numbers and math (...) To comply with the requirements of mental arithmetic, our brain has to tinker with whatever circuits it has, even if that implies memorizing a sequence of operations that one does not understand (...) [These circuits] function automatically, in a restricted domain, and with no particular goal in sight (...) The computational power of the human brain resides mostly in its ability to connect these elementary circuits into a useful sequence under the sway of executive brain areas (...) These executive areas are responsible, under conditions that remain to be discovered, for calling the elementary circuits in the appropriate order, managing the flow of intermediate results in working memory, and controlling the accomplishment of calculations by correcting potential errors. (Dehaene 1997, p.6, p.134, p.203)

Marr suggests a similar idea when he considers the possibility of analysing a certain "capacity" in a multiplicity of non-related abilities:

> [T]he figure-ground "problem" may not be a single problem, being instead a mixture of several subproblems which combine to achieve figural separation (...). There is in fact no reason why a solution to the figure-ground problem should be derivable from a single underlying theory (1977, p.133 in Haugeland).

Here figural separation might be the goal that the system must fulfill through the engagement of different abilities. Marr also considers the possibility in other capacities:

> [T]here may exist no Type 1 theory of English syntax of the type that transformational grammar attempts to define (...). An abstract theory of syntax may be an illusion, approximating what really happens in the sense that (...) the behavior [corresponds to a] set of processes that implement [a certain task] and which, in the final analysis, are all the theory there is. In other words, the grammar of natural language may have a theory of Type 2 rather that of Type 1. (ibid., p.135).

What this amounts to is that the theory of English grammar can be correct for a subject even if he does not actually implement it. Marr speaks of an illusion here because he is committed to the idea that *all* theories at the task level must be implemented. Yet, as we have seen, it might be the system complies with the theory even if it is not implemented

Boden (1988) introduces the notion of "quick and dirty" mechanisms as the constituents of functional "neat" satisfactions:

> [T]here are reasons for thinking that evolution and/or individual learning have provided us with a number of "quick and dirty" methods of making inferences or interpretations of various kinds (1988, p.228)

This idea is central in my position. Operations* can be engaged to undertake whatever capacity* the organism needs to fulfill. As Allport recognizes, these operations* can implement a combination of partial solutions that solve the problem the organism faces. This can be done by what we have called "virtual machine emulation". Allport advances (in reference to a task of *environmental monitoring)* that the solutions for a given problem could be:

> (...) expected to operate at many different levels. These should include *fast*, relatively crude or approximate systems, operating on rule-of-thumb (that is, associative) criteria, both learned and unlearned, as well as possibly slower and more sophisticated systems. (1989, p.653)

Then the engagement of such operations* allows the satisfaction of the capacity, granting the organism with behavioural coherence. In the case of visual attention the capacity* value is located in the satisfaction of *attention* tasks, though it is subserved by operations* not directly related to attention constrained and maintained precisely by the capacity* that ensures the coherence and preservation of the behaviour. According to Allport's theory, one such operation* can be, in the case of visual perception, "selecting objects of poss̆ible priorization". This process will be on-line, and will feed other operations* within the whole attentional processing, being an autonomous constituent of the capacity*. Being cognitively independent, some of the operations* do not have to be engaged at the same time as the actual capacity*. In the case of visual attention "this constraint implies that (possibly very complex) processes of perceptual grouping and segmentation may have logically to *precede* the effective focusing of visual selection-for action" (Allport ibid., p.650).

As we saw in Chapter 4, the research in the domain of Artificial Life can produce the same results. The picture was that the behavioural coherence of Artificial Life artifacts *emerges* as a result of the dynamics of *self-organizing systems*. Such systems are described as systems in which the observed patterns of the system are explained by the collective behaviour (under specified conditions) of various "behaviour systems". It is true of these artifacts that the actions of the "behaviour system" cause the overall behaviour and that the overall behaviour guides the action of the parts. This accounts for the emergent property which need not comply with the description in terms of the behavioural regularity. This possibility was manifested by the "following walls" and the "self-charging" robots. The emergentist solution relied on two simple "behaviour systems" whose environmental interaction yield positioning between the poles as a kind of side effect. The "behaviour systems" were: (1) some phototaxis system that yields a zigzag approach to any light source, and (2) an obstacle-avoidance system that causes the robot to turn away when it hits something. With these two simple systems in place, the target behaviour emerged smoothly and robustly. In fact we could describe the system to be computing two rules: (1) "approach the pole", and (2) "don't touch the pole". This could sustain a correct description of the artifact's behaviour that supports counterfactuals.

6.6.Operation* and capacity* conceptual independence

One of the upshots of the thesis I have just presented is that the project of explaining a cognitive capacity* happens to be *conceptually* independent of the project of explaining how it comes to be satisfied. The fact is that we can draw a natural distinction between characterizations of a functional system in terms of intrinsic, computational properties of the system, and characterizations in terms of properties that it has in virtue of performing as an agent-in-an-environment. The capacity* is individuated by the pair environment-(sub)system and the operation* is specified by the intrinsic properties of the system. These are the mechanisms that make the rich and varied behaviour of a given system possible. The explanations for both characterizations are *independent*; they are contingently, rather than necessarily, connected. The capacity* characterization is *normative* and *contingent* of the system: It is the requirement that the system *must* comply. The operation* characterization is *descriptive* and *necessary* for the system. In this regard, suppose an individual, say Fred, and his *doppelgänger*[40]. Both share the same sort of *processes*, the same operations* though at the same time they do not share the same *capacity** fulfilments. Suppose that in the duplicate world arithmetic had been formalized in a different way. Then the *capacities** of both the original and the dopplegänger would be different, even if their *operations** were the same.

The independence of both capacity* and operation* projects has also certain pragmatic consequences. For instance, the natural strategy for a theorist could be first to determine the capacity*, and then the operations* that make it possible. But nothing precludes the opposite strategy. Moreover, sometimes it could happen that the theorist might be interested in revealing the basic cognitive properties of the system, regardless of which informational-processing problems a determined system is able to solve. Yet, the former strategy will always be more economical so long as it will pick out the relevant cognitive activities of the system. We might, for instance, discover that some operations* happen to have no use in the cognitive economy of the agent.

Generally, what we label as a capacity* specification is characterized in terms of its environmentally specified properties that needn't supervene on any particular localized proper subset of the properties of the overall system. In contrast, an operation* characterization of a process of a system is specified in terms of properties that supervene on computationally or physically local properties, regardless of the way in which we ground its contents. (Note that the theorist still have the choice between narrow or wide contents.)

[40] A dopplegänger is a figure proposed in philosophy-of-mind discussions about content, and corresponds to a duplicate of a subject such that it shares with the original all intrinsic physical properties, and lives in a duplicate world that also share all physical properties with the original world.

In the case of reasoning about and predicting an object's motion, the capacity* (the capacity of being sensitive to dynamic principles) is environmentally characterized: We describe how the system *must* behave to attempt an efficient engagement with objects in a specific environment. On the other hand, the operations* are characterized looking both at the intrinsic properties of the mechanisms as well as to the relational properties of their states: We describe which are the "reasoning about objects" properties of the system.

At a higher level, we can show other examples. One, mentioned by McClamrock (1995), concerns the way in which certain animals obtain nutrients. Primates don't synthesize vitamin C and therefore must take advantage of the correlation in the environment between the foods they eat (which contain vitamin C-fruits) and their colour (bright colours like reds and oranges) to detect those foods. By focusing the detection apparatus to colours rather than some more hidden but directly relevant property of the object, one gives up some reliability -one might mistake flowers for oranges- for a huge gain in the cost-effectiveness of the search. You use mechanisms that can fairly directly detect the colours of objects but not ones that distally detect their vitamin C content. The use of these contingencies illustrates the exploitation of a kind of mismatch between the task, the capacity*, of the system and the process, operations*, of the system. The capacity* is finding foods containing vitamin C in the local environment, and the operation* is detecting coloured edible things. So long as both execute a different function, it results in a blocking of the explanatory cascade.

Other examples are less clear, even if they are better to understand the distinction I wish to draw. The adaptive function of the human heart, for example, is to pump blood. This is so, at any rate, on the assumption that past hearts pumped blood. That's the etiological account, under which we are entitled to attribute the function of "pumping blood" to the "heart". However, in my account the capacity* should be attributed to the whole organism, or at least, to the whole cardiovascular system. One of the capacities* of the cardiovascular system is to "pump blood". To individuate that capacity* we need also to appeal to the environment and to other parts of the organism; for the satisfaction of a "pumping blood" function in an efficient manner it is required that certain conditions of the organism apply. For example, the cardiovascular system must be able to adjust the pumping of the blood to the oxygen requirements of the body, and the heart has to be sensitive to oxygen demands of the rest of the body, otherwise the heart wouldn't contribute to the survival of any individual. In transplanted hearts the graft is hardly sensitive to the oxygen needs of the rest of the body, so long as it is a denervated heart. It is a malfunctioning heart even if it pumps blood. In sum, an effective "pumping blood" must be contextually individuated. However, the operation* of the heart may have also been individuated only by reference to intrinsic properties of the heart. Of course, there are many actions that a heart may fulfill, but we are interested in that which contributes to the satisfaction of the capacity*. In this case although it would be more accurate to talk of the function as "pumping blood-like liquids", I think we can allow -for explanatory reasons- a certain looseness of specification and leave it as "pumping blood". Then in this case what we have is that a capacity* corresponds to an operation*, that is the system fulfils a capacity* with only one operation*. So much the better for the theorist, because he has gotten his explanation for a cheaper price. Yet, this is in fact only one of the many other possible cases of some capacity* explanatory undertakings. The important point is to see the different explanatory aims that each notion subserves.

There are still some further clarifications to be made. One can be formulated by the question "how an operation* that does not change is able to contribute to different capacities*?" One possibility is to see the problem as one of context-dependence. As we have seen in the previous chapter, the position of a given DNA sequence with respect to the rest of the genetic material is critical to its status as a gene; type-identical DNA sequences at different loci can play different hereditary roles - be different genes, if you like. So for a particular DNA sequence to be, say, a green-eye gene, it must be in an appropriate position on a particular chromosome. A DNA sequence can have a description as an object of the role (its operation*) it has in producing certain type of RNA. That description would describe the molecular mechanics of the sequence. Yet, that very same sequence could be also described as contributing to a role (its capacity*) in fabricating proteins that participate in some process that affects the phenotype of a system in such a way that it allows the survival of the individual, and of the species. The description of that capacity* would have to rely on the *position of the gene in the DNA sequence*, since it is what determines the contribution of the capacity to each specific function. Similarly for a given action of a computer's CPU, such as storing the contents of internal register A at the memory location whose address is contained in register X. Two instances of that very same operation* might, given different positions in a program, differ completely in terms of their functional properties at the capacity* level: At one place in a program, it might be "set the carry digit from the last addition", and at another, "add the new letter onto the current line of text".

Another possibility, however, there is another way to see the question of how an operation* can contribute to different capacities*. The fact is that the conceptual independence between operations* and capacities* imply, among other things, that there can be both a one-to-many relationship between a capacity* and the operations* that contribute to it, as well as a one-to-many relationship between an operation* and the capacities* to which it contributes. In this regard, operations* can be engaged, as we have seen, in many different, partial and/or redundant, ways to satisfy a given demand.

A final observation. As I have argued, we can attain a capacity* specification independently of the operations* that make it possible, and such a characterization can be ascribed to an individual without assuming the internalization of the *contents* of the theory that specifies the characterization. Yet, that does not preclude that an equivalence can be found between the capacity* and the operation*, so that it could happen, concerning for example the reasoning about objects, that we have a mechanism that instantiates a theory of dynamics. So much the better for the theorist. Again, the relevant point is that we *need not* have such equivalence. In the case of the vitamin C the intrinsic process could be to detect some property of vitamin C. Yet, the description of the ability in terms of *capacity** must be seen as "resultant"[41] of the *operations** of the system processes.

6.7.Naturalistic support: Bounded functionality

There is another issue that we must tackle to give the proposal a robust basis. We still need some naturalistic argument that can give support to the idea that the functional design of a biological system might not correspond to the functions it needs to fulfill. In short, if it is

[41] Resultant is *not* used here as one of the two elements within the opposition between emergent and resultant properties that has been proposed in the discussion about emergent properties. Resultant properties are simple, non-structural natural properties that are exemplified by objects or systems that attain the appropriate level and kind of organizational complexity but that don't exert causal influence on the behaviour of the possessor (like mass or shape) (cf. O'Connor 1994).

easier to implement a design to solve a problem, why would a biological system employ a more complicated way of doing so?

To get to the point, recall Grandpa's story once again. After having identified the *capacity* of a cognitive system, Grandpa goes on to assume that the explanation for a cognitive capacity should be revealed via functional analysis. The generally held answer is that what makes a given output right is that it is derivable via the characterizing inferential pattern that decomposes:

i) S is competent in T

ii) T as analysis t_1, t_2, t_3...t_n

iii) S implements t_1, t_2, t_3...t_n

I have been arguing all along that an alternative hypothesis is that functional compliance may be satisfied without the internalization of the analysis. As a matter of fact, we have seen how sometimes the brain avails itself from a variety of mechanisms to achieve its goals or capacities*. The brain gets the job done by "quick and dirty stratagems" in terms of Clark (1996). In short, the brain does not always implement a neat and lineal "functional analysis" to undertake its functional requirements. Some processes are engaged in different ways to achieve some functional requirements. And the way such mechanisms accord with a functional description may block the explanatory cascade. The satisfaction can be seen as an intersection of different analyses employed to attain functional description as a result. However, and that is the important point, the system seems to comply with the functional description. Why does the cognitive system not implement the analysis?

As we said above, the functional-analysis strategy is an *optimizing* story. It is true that what an optimizing story amounts to is not at all clear. By optimization is usually understood the theory (or theories) that try to give the best (*optimal*) solution to a specified problem, assuming that the best solution is the right one. However, nobody knows what it means to be the *best*: The most elegant? The most parsimonious? The most natural? In a sense, every theory could be seen as an optimizing theory, if we take it to look for the right solution. As Gilman has pointed out:

> (...) simply framing an optimization problem often is not so much a straightforward technical task, as a complex pragmatic, and even philosophical, problem of how to *characterize the function in question, and how to identify the relevant variables and constraints*. (Gilman 1996, p.304, my italics)

Yet, we can get a grip on the idea if we remember one of Grandpa's basic assumptions presented in Chapter 1, namely:

> A6: A biological system is designed to comply with its function

For Grandpa, this is the default assumption of functional analysis. Grandpa is confident that his analysis is implemented because he assumes that the brain must have found the very same solution as him. This was guaranteed by what I called:

> *Principle of Convergence:* Entirely independent design teams come up with virtually the same solution to a design problem.

Therefore, we grant that our solutions will be correct in virtue of the principle which every solution to a problem that the functional capacity has to meet will be similar. But this is the problem. Such a principle is not only conceptually unsustained, but empirically also. Even if we don't know what an optimizing system is, we have seen up to now examples that show how a solution might not accord with the cascade. The cognitive system is a system that looks for the *most available solution to a functional requirement with the resources at hand*. What we have seen until now should have already persuaded us of the following claim: The cognitive architecture could be seen to work, at least in some cases, as a multipurpose system, which bases its success in using sundry, sub-specialist and redundant mechanisms to achieve (in a non-unique manner) a particular capacity*.[42]

[42] A caveat: This approach is neutral on certain questions where that has pivoted much debate. For example, I will take the notion of *functional specialization* (Shallice 1988) conceptually equivalent (in what respects this discussion) with that of *modularity*, so that my regarding the system as "multipurpose" can accommodate the notion of functional specialization. Likewise, the proposal should not be taken as contradicting

To give naturalistic support for this assumption, I will resort to the idea that cognitive systems are subject to constraints that force them into certain functional paths, which belong to a contrast class of other possible functional paths. Instead of looking at functional requirements based on the idea that *a problem needs the best solution*, nature normally counts on the idea that *a problem needs to be solved.* As a matter of fact, biological systems use their resources to face demands and look not for the most elegant, parsimonious solutions of all, but for those available. I will call such dynamics **bounded functionality**, after the notion of bounded rationality advanced by Simon (1981). Bounded rationality is the dynamics that economic agents engage in when faced with optimizing insoluble problems and who reach acceptable decisions by a process of heuristic search.

Then, for us, bounded functionality accounts for the dynamics and scope of functional solutions that the resources of the system and the environment allow. Bounded functionality is the principle that constrains our cognitive systems in search of adaptations. A functional explanation would then have to take into account the available cognitive resources, the functional demand and the environment in order to give an account of cognitive capacities. The dynamics of bounded functionality impose a "functional history" that minimize optimal designers, regardless of the need to comply with an optimal solution to satisfy the functional requirement. The ability to deploy "virtual machines" out of operations* allows the brain to satisfy functional requirements without the need for implementing the "official" cascade. Accordingly, the explanation of a cognitive ability will resort to how such virtual machine has been satisfied.

Bounded functionality changes Grandpa's assumption of how we are endowed with our functional capacities. In this regard, we don't reach competence with respect to certain problems, it is rather we use (and evolve) our cognitive resources to be effective in the minimally evolutionary valuable proportion with respect to such problems. In other words, it might not be so crucial how we are efficient in doing something, but how we have come to be so provided with our basic cognitive endowment. In this sense Dehaene contends, for example, that there are epigenetic changes that account for the way a certain cerebral circuit

the *innateness* of certain operations*, since as part of the system machinery and abilities (control states, modes of organization) could be taken to be innate, while the whole proposal remain intact.

is used for purposes other than those for which it was designed for:

> The basis for such changes in the function of cerebral circuits is *neuronal plasticity*: the ability of nerve cells to rewire themselves, both in the course of normal development and learning, and following brain damage. Neuronal plasticity, however, is not unlimited. In the final analysis, the adult pattern of cerebral specialization must therefore result from a combination of genetic and epigenetic constraints. Certain regions of the visual cortex, initially involved in object or face recognition, progressively become specialized for reading when a child is raised in a visual universe dominated by printed characters. Patches of cortex entirely dedicated to digits and to letters emerge, perhaps by virtue of a general learning principle ensuring that neurons coding for similar properties will tend to group together on the cortical surface. (...) Learning never creates radically novel cerebral circuits,. But it can select, refine, and specialize preexisting circuits until them meaning and function depart considerably from those Mother Nature initially assigned them. (1997, p.203)

To sustain the bounded functionality proposal we need an evolutionary principle that accounts for it, and another that blocks the more widely held idea of optimal attribution of functions. The former could be called:

> **Principle of sufficient functional satisfaction**: A system only needs to find a partial solution to reach competence.

Whenever such partiality is "selectively sufficient" it will get caught in the net of adaptive properties. This might be called:

> **Law of the minimum evolutionary effort:** A system that minimally satisfies a function will not evolve to satisfy it further.

The bottom line is that the dynamics of the "functional satisfaction" has to be freed of designers, regardless of the need to comply with an optimal solution to satisfy the functional requirement. The power of basic processes and the ability to engage them in a "virtual machine" allows the cognitive system to achieve the functional requirements without the need for implementing the official cascade.[43]

Moreover, bounded functionality can help us understand why certain demands have limitation in their satisfaction. As Dehaene has eloquently put it:

> As humans, we are born with multiple intuitions concerning numbers, sets, continuous quantities, iteration, logic and the geometry of space. Mathematicians struggle to reformalize these intuitions and turn them into logically coherent systems of axioms, but there is no guarantee that this is at all possible. Indeed, the cerebral modules that underlie our intuitions have been independently shaped by evolution, which was more concerned with their efficiency in the real world than about their global coherence. This may be the reason why mathematicians differ in their choice of which intuitions to use as a foundation and which to relinquish. (1997, p.246)

6.8.Applications

How can we apply the distinction between capacity* and operation*? The empirical review presented in Chapter 4 can help us in this direction. Let us begin with Ullman's theory of visual cognition. We saw that there is set of spatial relations that our perceptive system seems to be competent in. We are able to detect quite efficiently complex relations such as inside/outside, above/under, etc. Then, this competence, the computation of such sort of spatial relations, could be accommodated in my proposal as the capacity* specification, the functionality that the system shows as an agent-in-an-environment. On the other hand, we also saw that in Ullman's characterization of visual cognition there is a set of basic operations that the visual system can assemble elementary operations in what he calls "visual routines". This subserves the goal of extracting shape properties and spatial relations from the visual representations. The elementary operations, such as shifting the processing focus, "colouring", boundary tracing and marking, are used by the cognitive system to perform a variety of spatial relations that contribute to fulfill the capacity*. This set of basic perceptive operations would count as operations* in my framework. Finally the notion of

[43] Ullman (1996) recalls an earlier experiment on pattern recognition in animals to clarify this point. In perceptual experiments, octopuses were trained to successfully distinguish squares from diamonds of different sizes and locations. As it turned out, however, the animals then responded to triangles as equivalent to diamonds. Apparently, what they actually used to make the distinction was the property of having a pointy top, while ignoring the rest of the shape. This property of having a pointy top depends on a small local region (at least for convex shapes), and it can be tested without analysing the entire shape. Such properties are easier to compute.

visual routines could be subsumed under my notion of virtual machine, the capacity* fulfilment, which corresponds to the conjunction of the different operations* to satisfy a capacity*, and the role of each operation* in the capacity* fulfilment counting as their functions*.

Similarly, in the case of mental arithmetic, the competence in performing arithmetical operations would correspond to the notion of capacity*, whereas the basic operations identified by Dehaene included quantification (counting and subitizing) and processing of quantities strategies would count as operations*. Finally, there is the case of reasoning about objects. It is my view that there is a way out to account for the paradox that humans seem to be employing kinematic principles to reason about objects, even if they must have a sensitivity to the notions of dynamics. On the one hand, we have the notion of *capacity**. In this example we can fill such a notion with the principles that must be employed by an individual to be able to undertake an efficient engagement with the physical world. These include, for the concept of a physical object, the notions of space, time and causation, provided they are fundamental to the very possibility of a mature grasp of the idea of a physical world. They are fundamental because, among other reasons, it is a fact that accordance with them is necessary to explain why humans have survived. For instance, a species *not* sensitive to dynamic principles might become extinct because its individuals wouldn't care if a rock fell on their heads, since they wouldn't attribute mass to stones.

On the other hand, we have the notion of *operation**. This notion can be filled here by attributing to the cognitive system the faculty of reasoning according to the principles of kinematics. In other words, the cognitive system has a subsystem (a module or processing mechanism) that has the property of interpreting the physical world in kinematic terms. This view also lends support to the "principle of minimal evolutionary effort" and that of "sufficient functional satisfaction". This is so by noting that the extra computational cost of taking dynamic variables into account, over and above kinematic variables, arguably outweighs the small increase in accuracy that would be gained from doing so. However, as I argued above, there will also be the possibility that a cognitive system uses external aids to accomplish an operation*, as for example what happens when we use an abacus, or a pen and paper. In our case this can include Peacocke's view on the requirement that for an individual to be said to apprehend an object as a material object she should be *at least* *sensitive* to dynamic principles. If we consider that an individual *reasons* according to kinematic principles, then it could well be, and be compatible with Peacocke's proposal, that the normal subject is sensitive to such a conception *just when she interacts with the physical world and perceives the consequences of dynamic principles such as "weight" or "inertia"*. These situations occur when we examine an object haptically, or when we feel the weight of an object. Hence, we could see such a "contextual sensitivity" to the dynamic principles as a surrogate operation*. To achieve that capacity*, then an individual must deploy a virtual machine whose components are:

(a) The operations* corresponding to the reasoning according to kinematic principles.

(b) The surrogate operations* corresponding to the sensitivity to dynamic principles by the interaction with the physical world

6.9.Advantages

The distinction between capacity* and operation* will help us in different ways. First, the capacity*/operation* distinction will help us in accounting for the *productivity* of cognitive functions. The thesis I have presented takes for granted the assumption that the resources of the brain -as a cognitive system- can be used in an open-ended number of ways to satisfy an indefinite number of functional requirements possibly posed to the individual (or the species). Indeed, we are apparently able to have an unbound cognitive functional capacity, yet we have a (small) finite computational capacity. The capacity*/operation* characterization can explain in this sense how a system that does not have intrinsic properties in some capacity, e.g. playing chess, can play chess: The engagement of a number of operations* in a specific way. Suppose that the *capacity** of some cognitive system is to "perform logical inferences" since in a certain environmental situation, say in a logical-world, it needs to perform logical inferences in order to get food. Say that such cognitive system does not have a "knowledge of logic", it has not internalized a theory of logic, though it is capable of complying with the logical inferences because it emulates a "vĩrtual logical machine". Say that this emulation is accomplished by the effective engagement of a number of basic intrinsic properties of the system, its statistical-reasoning operations*, that perform statistical calculations and certain operations* that filter the outcome of the first ones. Such a capacity* fulfilment is possible thanks to the productivity of basic operations* that perform the function* of deriving logical arguments.

Second, the notions of capacity* and operation* will help us in accounting for the *adaptability* of cognitive functions. Suppose that humans were space creatures, that is, a species that had appeared and evolved in zero-gravity space. Suppose further that Earth-humans were made out of the same stuff and in the same way as Space-humans. Then in the space-environmental situation the proper function of a system in dealing with objects would be "reasoning in accord with kinematic principles", so long as the notions force mass would be useless in zero gravity. Then in this case the capacity* of the system would correspond to the operation* of, say, the perceptual system provided that it "reasons in accord with kinematic principles". Suppose then that such Space-human species move to the Earth. Then in this new environmental situation the selective advantage is attained by "reasoning according to dynamic principles". Then the Space-humans will survive and only if they satisfy the requirement of "reasoning according to dynamic principles". Therefore, if they satisfy that requirement the capacity* of the system changes. However, and that is the relevant point, that need not require that the operations* of the system change. If the cognitive system is able to complement its "reasoning about objects according to kinematic principles" with some other operation* that "sufficiently" satisfies the functional requirement of "according to dynamic principles", then the system has adapted to a new situation without evolution. The operations* of the cognitive system remain the same, and the only thing it has to do is to account for the new capacity* out of the right combination of operations*.

Finally, the capacity*/operation* distinction will help us accommodate two aspects of the notion of biological function that seemed to be incompatible in recent discussions, etiology and disposition, that is, the opposition between Millikan proper functions and Cummins-functions that we saw in Chapter 5. The reconciliation has the following form. The contention presented to dispositional theories of function is basically that dispositional analysis is too liberal, since it cannot distinguish between "accidental" functiõns (causal results of structures that could well add survival propensity to the organism in which they are placed) and "proper" functions. The beating noise of the heart is the usual example; it increases our propensity to survive through its value for the diagnosis of heart diseases, but we do not contend it as a function of the heart. Dispositional analysis actually overgenerates functions. A related criticism is that the analysis has the opposite effect. Let us understand a disposition in the following way:

> For a system *s* of type *S* to have function *F* requires that system *s* satisfy structural conditions under which the system (relative to typically unspecified *ceteris paribus* conditions) achieves some specified state; for instance a given state of "equilibrium" with the system's environment. The "equilibrium" are the states that play the role of fitness in the biological endowment of the system.

Then, by including in the definition the "equilibrium state", whatever that is, it may become unintelligible to take it to be a function of the system in any substantive sense.

However, there are authors that deny that both notions need be incompatible or orthogonal. Godfrey-Smith (1994, 1996) argues, for example, that both the form of explanation discussed by Wright and Cummins exist in science. There are explanations about what a thing does, and there are explanations of how systems realize complex capacities via the capacities of their parts. Likewise, Griffiths (1993) argues that objections about the null contribution of Cummins-functions to the understanding of "proper functions" of biological items and human artifacts are correct but they miss the point, since the proper function of a biological trait is the function it is assigned in a Cummins-style functional explanation of the fitness of ancestral bearers of the trait. The adequacy of teleological explanations given using proper functions depends on the validity of these earlier functional explanations.

The fact is that, according to these authors, we can incorporate the etiological approach into the Cummins-function approach. The proper function of a biological trait is the function it is ascribed in a functional analysis of the ability to survive and reproduce (fitness) which has been displayed by animals with that feature. This means that a feature will have a proper function only if it is an *adaptation* for that function. The trait must have been selected because it performs that function.

For Griffiths, this allows the analyst to consider proper functions with two roles in biological explanation. First, the biological fitness of a type of organism can be explained by a Cummins-style functional analysis. An organism's fitness is a measure of its overall ability to survive and reproduce, relative to the capacities of competing types in the population. The analysis of this ability will reveal a number of 'fitness components'. Fitness components are those effects of traits which enhance the fitness of their bearers. They are Cummins-functions of those traits about the overall ability of the animal to survive and reproduce (fitness). The proper function of a trait is that effect of the trait which was a component of the fitness of ancestors. They are the effects in virtue of which the trait was selected, the effects for which it is an adaptation.

Accordingly, proper functions are characterized by their capacity to enter into the second kind of biological explanation, the teleological explanation of the presence of certain traits. Proper functions can be used in this second kind of explanation precisely because they figure in the first kind of explanation. In sum, the proper functions of traits are those effects for which they are adaptations. To explain a trait by alluding to its proper function is to explain it as the result of natural selection.

García-Carpintero (1996) completes this approach by noting that the contribution of etiological theories, such as those of Millikan and Neander, is the introduction of the *normativity* of functions. In that sense for him, "the normativity criticism has to do with the fact that something (as it will turn out, a certain explanatory force) is *missing* in the dispositional analysis, and is independent of whether the dispositional condition is to be kept or not." However, he contends that the important point is to see clearly that the normativity of proper functions is not explained by *lacking* a dispositional clause like Wright's, but entirely *by having* a historical condition. In that sense, he backs a modified version of Wright's proposal in which both the dispositional and historical accounts are taken into account. The maintenance of the dispositional account is necessary for it is implicit in any notion of function. García-Carpintero remarks in this regard:

> All she needs to claim is that it is incompatible with explaining the normativity of functions that functions be *current* dispositions; for it is a symptom of the normativity of functions that (as she insists so often) "malformed" organs have proper functions (this is why they are "dys*functional*"), although they currently lack the relevant dispositions (this is why they are "*dys*functional"). But she cannot be taken as denying that having a function entails that something else have or had a related disposition. In fact, her own account entails this. There are her denials that there are biological laws, psychological laws, etc. but, as we have seen, all that she is in fact denying is that the relevant laws are strict and deterministic. We can take it that, to the extent that there are non-strict laws, she would accept that teleo-functional ascriptions presuppose them, and they individuate the relevant (not necessarily current) dispositions. (ibid., p.20)

A disposition, for García-Carpintero, is constitutively specified by means of some conditional laws with subjunctive force ('if such and such, that would happen'). These laws hold *ceteris paribus*. The *cetera*, for García-Carpintero, that must be *paria* include at least, typically, that items having the disposition have (in actuality) a certain internal constitution (perhaps one among several) and that certain conditions are (in actuality) obtaining in the environment. Provided such a framework, he follows Schiffer (1991) in calling 'realizers' to the former elements and 'completers' to the latter. Accordingly, García-Carpintero believes that we can save Wright's version of function which would include both the dispositional and historical condition, namely, a token-structure *s* of type *S* has function *F* iff:

(i) Under conditions *ceteris paribus*, *S* causes *F*, and
(ii) *s* is there because under conditions *ceteris paribus*, *S* causes *F*.

This version includes the reference to dispositions, which includes then the dispositional account, and additionally it is a much more general account concerning the mechanism of fulfilling the historical condition: it can be either natural selection, as well as learning by operant conditioning, Lewisian conventions and conscious design. This version should comprise a modification of the dispositional condition, though, which in Wright's version would require that the condition obtains now, that *s* has the disposition to cause *F* now. The modified version is thus weaker, since the general conditions required for the attainment of the system's higher-order goals[44] and the "complementing" conditions necessary for the contribution to it by instances of *S* still obtain, even if there is a malfunction. That is, even though the "realizing" conditions necessary for the structure to perform its proper function may fail to obtain, actual conditions are still such that the structure's contribution to the main purposes of the system of which it is a part would have its normal effects. In that case we can still attribute the function of pumping blood to a diseased heart.

How can we relate this discussion with my proposal? Overall, my opinion is that the notion of *fitness* implicit in dispositional accounts needs an account of *why* a certain proper function was selected for or maintained, as well as it needs some intermediate specification above which we could have the historical condition and under which we have the mere dispositional account of a system's components. In this regard, I agree with García-Carpintero's assertion that Millikan proper functions need an appeal to dispositions in order to explain *why* a certain trait has been selected for or maintained. Specifically, we need to appeal to the notion of fitness that Millikan's account does not comprise. However, a dispositional account understood in terms of Cummins will not do. The problem in including the fitness requirements in Cummins-functions has been presented in the previous chapter. Cummins-functions are assigned to a system when analysing a complex capacity into a set of simpler capacities. The function of an item is its contribution to the overall capacity, which is explained in terms of the contributing capacities of parts of the system. Accordingly, dispositional theories of functions define the function of a mechanism or process in terms of its functional role, that is, in terms of its contribution to some capacity of the system to which the process or mechanism belongs. The key then is to note that this notion is actually concerned with the intrinsic properties of the system, whereas explanations that cite *fitness* are *environmentally-driven* explanations.

As a matter of fact, fitness accounts are generally defined relatively to some environmental object or feature. An organism's fitness is a measure of its overall ability to survive and reproduce, concerning the capacities of competing types in the population. The

[44] For García-Carpintero a system has proper function if and only if it has proper goals. The goals of a system can be arranged in a hierarchical system, lower-level goals being means serving higher-level ones. Typically, survival and reproduction will be higher-level goals of systems with functions established through natural selection.

function of the chameleon's skin is to make the chameleon the same colour as its immediate environment, etc. The existence of the feature is due to the fact that it meets certain environmental demands. Therefore, the fitness characteristics of a given cognitive mechanism derive from the environmental objects, properties or relations that are incorporated into that mechanism's capacities.

The characterization of fitness requires, in my opinion, the distinction that I have been drawing all along this chapter, namely, the separation of accounts of capacities understood in terms of environmentally determined demands from capacities intrinsic to the system. For this reason, capacities* are the necessary constraint to impose to dispositions -in terms of Cummins- for fitness accounts. Capacities*, for example, can change in the course of evolution in parallel with fitness, and regardless of whether the cognitive system has changed or not. In this sense, there might be capacities* that have been *lost* during evolution. In sum, if we restrain the account to the dispositions of the system in terms of the intrinsic operations* of a system's components, we will find ourselves caught in the mismatch between the outcome of such an analysis and the right characterization of the relevant capacities* of the system.

My view is that Cummins-functions should therefore be amenable to the distinction between capacity* and operations* presented all along. The distinction of the notions of capacity* and operation* gives content to the account of fitness that we are looking for dispositional accounts and, as we will see shortly, it is the necessary complementing condition for Millikan proper functions. The *relevant* dispositions should be constrained by the specification of capacities*. The capacity* specification provides the "completers" conditions of Schiffer (1991) and the "realizers" will be produced by the operation* specification.

Concerning proper functions, let us first recall that a proper function *F* of item *A* is the function of the item that originated as a 'reproduction' (to give one example, as a copy, or a copy of a copy) of some prior item or items that, due in part to possession of the reproduced properties, have actually performed *F* (or a partial contribution to *F*) in the past, and *A* exists because (causally, historically because) of this or these performances. The proper functions of that system are the functions it has performed in order to survive and reproduce (fitness) and which has been displayed by animals with that system. The proper functions of that system are those effects of the trait which were components of the fitness of ancestors; they are the effects by virtue of which the system was selected, the effects for which it is an adaptation.

Accordingly, proper functions are characterized by their capacity to enter into the teleological explanation of the presence of a certain system. We have seen that the crucial problem with this account is that proper functions are tied to etiology in a way that hinders explanations of *why* the system survives now or survived in the past. It is actually a requirement for fitness accounts that the characterization of the functional satisfaction of the system explain how the system meets the environmental demands, otherwise we won't explain why the system is selected for or maintained. Then, the notion of capacity* comes to our aid here, since it gives content to the notion fitness that proper functions require (Millikan opinions on this matter notwithstanding). Capacities* would provide the analysis of the environmentally constrained properties that allow the fitness of the organism, providing the characterization of *how the cognitive capacities* of the system helped the system to survive and reproduce in a given environment.* Hence, the proper function of a system (or subsystem) would be the capacity* whose satisfaction the system (or subsystem) that accounts for it was selected for or maintained, that is, the effects for which it is an adaptation. Proper functions can then be used as a full explanation precisely because they comprise an explanation for capacities*, which specifies the reasons behind the fitness of ancestors fulfilling the functional requirement.

In other words, proper functions are capacities* to which we have added the historical condition. Capacities* will thus have its place in the etiological accounts of functions, provided that the notion of capacity* specifies the class of possible etiologies. The class of capacities* (or the partial contributions to satisfy such capacities*) is the class of candidates for proper functions (in the sense of Millikan) of a system. Millikan proper functions are therefore a sub-class of the class of capacities*. It is in fact in the class of capacities* where we have to look for candidates to be proper functions. Therefore, the notion of capacity* could be seen as the meeting point for both Millikan proper functions and dispositional accounts. Capacities* are therefore useful insofar as they let us separate fitness from both etiology and the cognitive system's intrinsic "computational" properties. The former explains why one survives, and the latter explains how one can survive without being optimal. We could sum up the whole argumentation with the following characterization:

i) Operation* = Disposition + System

ii) Capacity* = Operations* + Environment

iii) Proper function = Function*(Capacity*) + History

Conclusion

The brain as a cognitive *bricoleur*

Now it is time to recapitulate the arguments developed in this dissertation and outline the conclusions. As I explained in the introduction, the -what we could say- *ultimate* aim of this work has been to clarify the idea that cognitive abilities might have nothing to do with their usual characterization, saving cognitive science with it. In other words, I set myself to make room for the possibility that a given system could show full-blooded accordance with a given cognitive theory, even though the system's causal structure would not match the structure of the theory nor would it draw upon the same information than that of the theory. Moreover, I contended that this could be done without dispensing with cognitive science research. How have I managed to accomplish these objectives?

First, I decided to set out the discussion with the issue of psychological reality. I did so because it is in attributions of psychological reality where the idea that there might be accordance without internalization hurts the most. In order to reveal the problems of functional attributions based on what I call Grandpa's explanatory framework, I examined what for me are two of the most complete and articulated proposals regarding the conception of psychological reality. Regardless of how such problems appear, my analysis of both proposals recognises nevertheless the difficulty of developing a robust notion of psychological reality.

In this sense, Davies' claim of what it is for a theory to be psychologically real seems to be a solid conception, though in need of revision since it leaves out certain conceptual, as well as empirically realized, possibilities that could be considered to be genuine attributions of tacit knowledge. Davies' proposal stands out nevertheless as a sound and coherent conception of psychological reality, since his account can confront Quine's challenge: There is a sense in which we can choose a theory among the class of its extensionally equivalent theories to sanction its psychological reality, which is the selection of the one that best describes the causal structure that accounts for the internalization of the theory by an ideal cognizer.

However, I presented serious objections concerning the development of the basic claim into a criterion to sanction the psychological reality of a given theory, what Davies labels *Mirror Constraint*. This constraint has been found to be too weak to account for the psychological reality of a specific theory since the Mirror Constraint compares a *theorist's structure* with a *mental element*, which is an epistemologically impossible comparison, so long as one is the mode of characterization of the other.

The second proposal I examined is that of Peacocke's. I tried to show that Peacocke gives a *necessary* condition of psychological reality for competence theories, but he comes short of giving a *sufficient* account. Any notion of psychological reality will have to accommodate a Criterion of Information, but it will need something else. If the examples presented in the course of the discussion are relevant, it should be clear by now that a pure informational account of psychological reality will have to face many problems that might turn it into a not very useful notion. As a matter of fact, a great deal of cognitive theories have a more or less important component directed at explaining *how the mind handles information*. The basic problem with Peacocke notion is that, in effect, all theories are informational, in the sense that they draw upon information, but it is not a necessary condition that the information that they describe as drawn upon is the causally significant description of the mechanism. I tentatively argued that the Peacocke's notion could be improved with a complementary condition, the specification of the cognitive architecture to which we attribute the theory, so that the criterion of psychological reality should be specified as *information-under-a-cognitive-architecture*.

In sum, we could say that, on the one hand, Davies' proposal gives neither necessary nor sufficient conditions to account for psychological reality, whereas, on the ôther hand, Peacocke's proposal doesn't provide sufficient conditions.

However, as I said above, even if we can try modifying both accounts, the attempt will clash with an underlying conundrum, which can be generalized from the particular Davies and Peacocke cases to the set of all functional attributions framed within Grandpa's explanatory framework. As we saw in Chapter 1, under Grandpa's framework explanations must be built according to certain substantive and methodological assumptions as well as to a certain explanatory strategy. Solid as it may seem, this framework breaks down before the evidence of certain empirical projects. The fact is that we can very well say that a theory is real if it describes the causal structure in the mind of the cognizers, or alternatively that the system draws upon the information stipulated by the theory.

The problem is that what might be relevant from a point of view of the theory may have nothing to do with the way in which the system manages to satisfy the theoretical description. For me the examples reviewed are well designed in order for us to see how such inconsistency is faced in cognitive science theorising. We saw that some plausible empirical projects point to the fact that there are ways in which a correct functional attribution might violate the inheritance of the superordinate in an explanatory cascade. Additionally, in some cases cognitive systems seem to satisfy a functional requirement by employing a number of strategies that cannot be interpreted through the functional analysis of the capacity.

Moreover, some of the processes employed by the cognitive brain can be seen to profit from strategies that are not specifically attributed to the properties of the cognitive architecture, but use the power of certain external devices such as, for example, a pen and paper. In other words, we saw how a system may accord with a certain functional description that is not implemented as a form of internal structure.

I interpreted these inconsistencies as meaning that the cascade sometimes fails, and a mismatch appears between the description of the faculty at one level, and the description employed at the lower level. This has relevant consequences for any notion of psychological reality. Indeed, if we complete a theory of a specific cognitive capacity which is to be right of a system but which is not to be "complied" or "implemented" or "discharged", then we face an inconsistency.

I said in the introduction that this could have devastating consequences for cognitive science. In what sense do I interpret this "devastation"? As I also expressed then, my claim is not that we have to wipe out all what has been done until now and start it all over again. Far from it. As a matter of fact, the finding does not undermine either cognitive science *research projects* or their *data*, not even some of their *explanations*. What the finding actually contests is the formula:

theory accordance equals theory *attribution*

What this actually means is that there is more distance to the goal of a complete cognitive explanation than meets Grandpa's eye: Even if we show that a system accords with a theory, the task of explaining a cognitive capacity is far from complete. Unfortunately for Grandpa, he has more homework to do than he thought. Indeed, explaining cognitive capacities might not just be the straightforward task it seems to be at first sight. Grandpa's comfortable assumption of the

Principle of transference: The individuation of a cognitive capacity is transferred from that of its behavioural efficiency

has to be abandoned. He can no longer rely on the functional specification of externally determined tasks to infer the operations going on in the mind of the cognizer to satisfy the requirement.

The way ahead is, in my opinion, through the distinction between the explanatory projects that I developed in Chapter 6. As a matter of fact, the kernel of my proposal is that what we need is to incorporate the research into the framework that separates the explanations of the system as an agent-in-an-environment from those operations the cognitive system actually performs, since these two perspectives actually correspond to a difference in functional explanatory activities. One corresponds to what I called the *capacity**, the explanation of the way in which cognitive agents comply with certain environmental demands, those that constitute the class of potential selectively relevant functions. This sort of explanation should account for the agent's behaviour in an environment, characterized according to a functional pattern that must be satisfied, and which, in turn, might be what (partially) accounts for why the system is there. On the other hand, there is the notion of *operation**, the project of explaining the intrinsic processes of the system that account for the satisfaction of the task.

As we saw, both types of explanation are needed to account for a given cognitive function, though both projects are *independent* and do not share the same explanatory cascade. One can happen without the other; they are not *necessarily* but *contingently* connected. The system might be performing a process which draws upon a different piece of information than the theory but which nevertheless subserves the completion of the task specified by the theory.

I completed the account with the introduction of the notion of *function** as the role that a certain operation* has in satisfying a given capacity*. I also provided some naturalistic support for that idea. I resorted to the idea that it might be that we don't achieve competence with respect to certain problems, and then such competence becomes *fossilized*, it is rather that we use (and evolve) our cognitive resources to be effective in the minimally evolutionary valuable proportion with respect to a number of problems. In other words, it might be not so crucial how we are efficient in doing something, but how we have come to be so provided with our basic cognitive resources. To back this proposal I presented an evolutionary principle that accounts for the proposal, and another that blocked the more widely held idea of optimal attribution of functions. The former was called the "principle of sufficient functional satisfaction": A system is faced with a problem and then finds a "partial" solution; whenever such partiality is "evolutionary sufficient" it will get caught in the net of evolutionary transmittable. The latter was called "law of the minimum evolutionary effort" which prescribes that a system that minimally satisfies a function will not evolve to satisfy it further.

It is my belief that this proposal adapts current cognitive theorising inconsistencies and makes them plausible. Furthermore, the proposal helps to understand *why* such inconsistencies appear. However, the proposal also presents a new agenda of cognitive phenomena to account for, which basically corresponds to the identification of the actual cognitive operations* that go on in the mind of cognizers. Some have been identified, and some need much more investigation, but the conceptual apparatus meets the explanatory requirements.

There is a price to pay, though. We have to accept a cognitive system which is sometimes a satisficing system, a system that looks for the *most available solution to a functional requirement with the resources at hand.* We must drop the idea that *any solution needs an optimal design*. Cognitive systems use their resources to face their problems: there is no designer that can analyse and look for the right decision. This is the basis of what I called bounded functionality. Bounded functionality is the principle that subserves our biological systems It is the dynamics by which the pair organism-environment is constrained by the conditions of the environment, the resources available and the process to reach a solution.

All this entails that the brain is best viewed as a "bricoleur" rather than a engineer. Under this metaphor, the cognitive system can be seen to solve functional requirements by availing itself to use the "materials" and "tools" it has at hand, rather than by analysing the problem and applying the best design. In this sense, we saw that some of the processes employed by the cognitive system can be seen to take advantage of mechanisms that are not specifically designed to solve the problem at hand, and it can even profit from the use of certain external mechanisms, such as for example memory tags, that might subserve certain difficult operations.

This does not mean that the solution attained by a cognitive system has to be *deficient*. The fact that the cognitive system is not suitably sensitive to discrete numerosities does not mean that we cannot *fully* comply with arithmetical operations. What this actually means is that even if the solution requires arithmetic, the cognitive system need not be designed to perform arithmetical operations; rather, it can comply with such a requirement by indirect, partial, redundant and external means. Accordingly, Dennett's "Principle of Convergence", which stipulates that entirely independent design teams come up with virtually the same solution to a design problem, and by which Dennett grants the soundness of the functional analysis strategy, should be downgraded. By this I mean that the principle can apply to *the solution to a problem*, such as for example the use of arithmetic to solve calculation problems, but not to the *design* that underlies its satisfaction, which can be quite different according to the different resources of the system that faces the functional demand. The fact is that naturally occurring environments hardly have good functional tools to design systems, so we have to expect evolved functional devices to be relatively efficient, robust in achieving solutions to their evolutive problems with the resources at hand.

The task undertaken in this dissertation is obviously unfinished. The way ahead should develop an important issue that I have hardly touched upon. This concerns how to identify capacities* and operations*, and how to distinguish the one from the other. This is a complex question for which I do not have a simple answer. Identifying capacities* is arguably the easy part of the problem here, since it does not differ from the way that cognitive science has been approaching cognitive capacities up to now. It is in fact the identification of operations* what poses the major problem. I would tentatively say that to achieve a good characterization of operations* we need not only good *computational* characterizations, but also good *evolutionary*, and *emergent* explanations. As a matter of fact, it is my view that the resources available in the natural history of the brain, as well as the ecological constraints with which our antecessors were confronted with are the stuff we should write the story of our present cognitive abilities. Such a story will consist of an explanation of the computational properties these resources have, and how such resources satisfy functional demands according to the principles of bounded functionality. Obviously, this is only a statement in search of empirical development, even though I believe that many cognitive scientists work under the same assumption.

In sum, the bottom line of this dissertation could be seen as just a warning about being overconfident in Grandpa's capacity to explain our cognitive abilities. Grandpa has drawn such a persuasive picture of mind that has driven us to believe that his theoretical and methodological tools are all that we need to carve cognitive capacities at its joints. Yet, as I hope to have made clear, Grandpa has overlooked some of ways in which biological systems satisfy cognitive tasks. This finding requires a different approach to cognitive theorising. Paradoxical as it may seem, it is because Grandpa is merely a sophisticated behaviourist, as Kim (1996) acknowledged, that such inconsistencies appear. As we have seen all along this work, Grandpa's principle of transference fails when it comes up against evolution's clever ways of dispensing with it. I have tried to present and explain such strategies. In the best of possible scenarios, I would have helped dissolving the paradox expressed by David Marr in the opening quotation.