THE PSYCHOLOGICAL REALITY OF COGNITIVE THEORIES:

CONCEPTUAL ROOM FOR THE BRAIN AS A FUNCTIONAL *BRICOLEUR*

Oscar Vilarroya

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

For my parents

I have no doubt that
when we do mental arithmetic
we are doing *something* well,
but it is not arithmetic.
David Marr

Acknowledgements

I owe many thanks to many people without whom this dissertation would not have been accomplished. First, I am thankful to my director, Professor Daniel Quesada, who devoted precious time, analysis and patience to this work. I am also indebted to my tutor, Professor Manuel García-Carpintero, who gave me the intellectual stimulus to undertake this project. I would also like to mention here the invaluable experience that I gained by working with Dr.Josep Mª Vendrell and with Professor Christine Temple, who allowed me to forge my current understanding about the brain and the mind. Bits and pieces of this work were read by Eva Juarros, David Casacuberta and Manolo Pérez Otero. I appreciate all of their comments. I would also like to show my gratitude to the members of Seminar directed by Professor Quesada and those of the GREC group whose comments allowed me to model much of what is in this dissertation. Last, but not least, I am tremendously indebted to a great friend of mine, Joe Hilferty, who has dedicated long hours to transform this work into a comprehensible text.

Contents

Introduction

The understanding, like the eye, whilst it makes us see and perceive all things, takes no notice of itself; and it requires art and pains to set it at a distance and make its own subject.
John Locke. *An Essay Concerning Human Understanding*

The seeds of this dissertation were sown when I was still an undergraduate. At that time I was a part-time assistant to Josep Mª Vendrell in his Neuropsychology Department at the Hospital de la Santa Creu i Sant Pau. Encouraged by him to read as much neuropsychological literature as I could, one day I came across a sentence that struck me, as I thought it was a deep and original insight into the nature of human cognitive capacities. I recall quite clearly that the article was a theoretical assessment on how to understand neuropsychological deficits of language competence. The sentence ran more or less as follows: "What underlies our linguistic capacities may have nothing to do with language". I say "more or less" because even though I thought I had made a copy of the article, I have never been able to trace it back again, regardless of painstaking efforts to do so. My belief was that the paper was written by the famous neuropsychologist Michael Gazzaniga, and I say "was" because I haven't found the sentence in any of his papers that I have consulted. Moreover, I now doubt that he could have ever said something like this. Fortunately, I later found a very similar idea in a paper by another famous researcher, David Marr, which is reflected in the quotation that opens this dissertation. It is my belief that under the apparent innocence of this sentence there is a devastating thesis for cognitive science. Indeed, if what accounts for our cognitive abilities has nothing to do with their usual characterization, then we might as well think about it; otherwise we would be wasting our time. My aim here is to make room for such an idea, since I contend that, even though there are many researchers and theorists that work under this assumption, it still needs to be clarified. As it will become clear in due time, the "having nothing to do" idea is not a metaphysical claim about, for instance, the hopelessness of folk psychology for cognitive science; it is rather a methodological contention that, nevertheless, implies far-reaching consequences for cognitive-science theorising.

However, the "devastating" effects of the thesis are not to be taken as implying that we should dispense with all cognitive science that has been undertaken until now. Far from it. The thesis aims at redressing *what* cognitive science is actually explaining and *what* remains to be explained. In case of being successful the *only* thing that the thesis would require is a different interpretation of past and present research, and the amendment of the explanatory framework.

I will approach the dissertation from a disputed and uneasy notion, that of "psychological reality", since I believe that it is in considerations of psychological-reality where the idea that cognitive abilities might have nothing to do with their usual characterization hurts the most. This is for me because the notion of psychological reality is a central tenet of cognitive science. Basically, it has been proposed as a way to legitimize explanations of cognitive functions, so a clear view on what we mean by "the psychological reality of a theory" is a good way to look at the causal implications of cognitive theories and, second, it can provide the theoretical and empirical criteria to sanction specific theories.

The contemporary debate about psychological reality stems from considerations made by Chomsky (1965, 1976, 1980a; Chomsky and Katz 1974; Fodor 1981a) and Quine (1972) regarding the attribution of linguistic grammars. Chomsky advanced the proposal that the attribution of a grammar to a speaker entails the status of psychological reality, and Quine challenged such a view with the argument that there is no evidence that could make one favour one of two equivalent extensional theories. So far, this discussion has been mainly concerned with the epistemological questions of whether inference to the best explanation grants the status of psychological reality, as well as whether evidence that counts towards the psychological reality of cognitive theories is different from evidence that counts towards the truth of that theory. For example:

> The question is: what is "psychological reality", as distinct from "truth, in a certain domain"? (...) I am not convinced that there is such distinction. (...) the evidence available in principle falls into two epistemological categories: some is labelled "evidence for psychological reality", and some merely counts as evidence for a good theory. Surely, this position makes absolutely no sense, but it remains implicit in discussion of the matter by psychologists and linguists to the present. (...) What we should say, in all these cases, is that any theory of language, grammar, or whatever, carries a truth claim if it is serious - though the supporting argument is, and must be, inconclusive. We will always search for more evidence and for deeper understanding given evidence, which also may lead to change of theory. What the best evidence is depends on the state of the field. (...) There is no distinction of epistemological category. In each case we have evidence -good or bad, convincing or not- as to the truth of the theories we are construction or, if one prefers, as to their "psychological reality", though this term is best abandoned, as seriously misleading (Chomsky 1980b p.12).

However, these discussions take for granted two issues that seem to be relevant for the debate. One concerns the matter of what psychological reality is, and the other has to do with what it is for a theory (or its constructs) to be psychologically real. The former has been discussed by certain authors (i.e., Searle 1990, 1996; Davies 1989a), who have focused on the metaphysical status of psychological states. They have been especially concerned with the sort of states that can be accepted as being psychological, as well as the relation between individuals and such states. I will not be concerned with issues of this sort. Rather, I will be concerned with the issue of what it means for a theory to be psychologically real, especially those proposals that claim to constrain what is it for a theory to be true of an individual.

Admittedly, most of the discussions between Chomsky's advocates and Quine's defenders take this last issue for granted, implicitly referring to different conceptions of psychological reality. Therefore, it is not clear which psychological-reality commitment theorists make when proposing a given theory. This has even driven some authors to consider that once we clarify such conceptions the very same positions of Quine and

Chomsky could be reconciled (George 1986). In any event, there have not been many attempts to tackle the notion of psychological reality in the context of cognitive science and those that have have taken one of two positions. Roughly, the first position could be described as follows: If empirical theories are correct, then they are real. This is the position of all those who adhere to the Realist principle, which contends that one should accept the ontology that the best explanation presupposes (cf. Fodor 1981a). Obviously, this is akin to embracing the principle of inference-to-the-best-explanation as a basic epistemological assumption. Therefore, for these theorists any explanatorily correct theory about what happens in the mind of an individual must be real; the sole strong restriction is for it to be explicatively adequate.

A second reading of psychological reality states that if a theory is correct, then it is isomorphic with the mental representations underlying the capacity (Bresnan and Kaplan 1982; Fodor, Bever and Garrett 1974; Pylyshyn 1984). Here there are its two main positions:

> **Propositional psychological realism** (Clark 1989) or **strong competence equivalence** (Bresnan and Kaplan 1982; Pylyshyn 1984): Basically, this position claims that if *a* is a competent cognizer, then *a*'s competence is causally explained by unconscious knowledge of the rules that account for the capacity. These rules are internally represented by structures in *a*'s mind and have the syntax of the natural language sentences describing the rules. In terms of Bresnan and Kaplan, a model satisfies the *strong competence hypothesis* if and only if its representational basis is isomorphic to the competence grammar.
>
> **Structural psychological realism:** On this view, if *a* is a competent cognizer, then *a*'s competence in some capacity is causally explained by the fact that *a*'s information-processing capacities are internalized according to the theory of the capacity (Clark 1989, p.155). Some authors provide a different but nevertheless compatible account based on the notion of informational content (Peacocke 1986). For such authors, a cognitive theory is psychologically real insofar as the in-the-head processings appeal to the very same informational contents than the theory.

The Realist reading seems to offer little more than a stipulative notion. Hence, we can agree or disagree with the way in which the term is used or with the assumptions made, such as whether inference-to-the-best-explanation is the best we can do to account for the psychological reality of a given theory. I will not be concerned with such a position. Regarding the second reading, I believe that the *propositional psychological realism* hypothesis is only supported by such conceptually candid theorists who embrace all the consequences of a good argument, intuitions notwithstanding. In any case, I think there have been a few successful criticisms of such a view (Clark 1989; Matthews 1991; Stabler 1983).

The structure of the thesis takes the following shape: I will first present the framework that establishes the conceptual and methodological assumptions on which I intend to base my discussion. I do so in order to focus the discussion on the conditions that a theory must satisfy to be considered psychologically real. This framework, which I will call Grandpa's framework, will be taken as an undisputed fact of all the proposals considered to belong to a specific theoretical framework that informs cognitive science in general.

Two proposals concerning the notion of psychological reality for cognitive theories will then be examined. The first one will be that of Martin Davies. Davies' notion of psychological reality will be seen to point to the right constraint for a cognitive theory: The existence of a causal structure in the mind of the cognizer. However, I will argue that the criterion that Davies presents to sanction the psychological reality of some specific competence theory is not suited to the task, because it is too weak as a conceptual notion and it is a too strong as an empirical claim. The ensuing chapter will be concerned with Christopher Peacocke's proposal about the psychological reality of cognitive theories. I will try to show that Peacocke gives a *necessary* condition for the psychological reality of competence theories, but he comes short of giving a *sufficient* account. At this point, I will argue that even though we could try to modify either account, both assume an underlying framework about functional ascriptions that compromise their revision. This could be succinctly put as the idea that functional attributions *must* be framed according to two basic explanatory assumptions. The first prescribes that psychological explanations should be developed in what has been called the "classical cascade" (Franks 1995), namely, the idea that a psychological explanation can be formulated at different levels. Marr (1982) is the usual reference in this regard. He proposed three distinct, but interrelated, levels, which he called the *computational* (which I will call task level), *algorithmic* and *implementation* levels. The second assumption establishes that psychological explanations must be unfolded according to what has been called the "functional analysis" (Cummins 1983), i.e., the explanatory strategy of analysing a given disposition into a number of less problematic dispositions such that the programmed manifestation of these analysing capacities amounts to a manifestation of the analysed disposition.

In order to show where the framework breaks down, I will review some empirical work in cognitive science. We will see that some of this work points to the fact that there are ways in which a correct functional attribution might violate Grandpa's explanatory framework. Specifically, a system may accord with a certain competence theory at the task level which does not describe the internalized cognitive structure that accounts for it. The empirical review will show that sometimes the cognitive processes responsible for some function may be satisfied by means that have little to do with the function itself. We have then a "paradox": a system that seems to comply with a given functional analysis though does not internalize such an analysis. Specifically, the system does not obtain its goals *by virtue of* executing the functional analysis, but by engaging different mechanisms that satisfy partially or redundantly the functional requirement. This will have relevant consequences for any notion of psychological reality. Indeed, if we construe a theory for a specific cognitive capacity as being true of a system but which is not "complied" or "implemented" or "discharged", then we face an inconsistency.

I will then make room for the possibility of a correct functional accordance without internalization. The way in which the tension will be resolved is by appealing to a distinction made in the discussion of naturalistic notions of functions. This distinction will be adapted to allow a distinction in two different explanatory perspectives of a unique cognitive phenomenon. My claim will be that it is possible to provide two explanatory accounts of a single cognitive capacity. In this sense, the paradox should be properly construed as two separate aspects of the faculty for understanding and producing the function in question. Clearly, both *explanatory projects* must be taken into account in cognitive science theorising. One corresponds to what we could call the *agent-in-an-environment level*, the explanation of the way in which cognitive agents comply with certain demands, those that constitute the class of potential selectively relevant functions. This sort of explanation must account for the agent's behaviour in a given environment, for which it may have been selected. On the other hand, there is the project of explaining the *intrinsic processes* of the system that account for the satisfaction of the task. Both projects are conceptually *independent* and do not share the same explanatory cascade. One can happen without the other; they are not *necessarily* but *contingently* connected.

In the last chapter the conceptual apparatus necessary to develop this idea will be presented. I will introduce the notions and naturalistic support that each explanatory project requires in order to subsequently sketch how they can account for the empirical findings that I review, and for any other cognitive explanation for that matter. I will then show how the proposal helps in accommodating the paradox of theory accordance without internalization, as well as in explaining why it occurs. This will be based on a naturalistic argument that portrays the brain as a biological system that, instead of being designed according to the problems it has to solve, it solves (selectively relevant) problems with the tools it has at hand. Finally, I will argue that the proposal is the natural meeting point for two notions of function normally considered to be orthogonal or incompatible: Millikan proper functions and Cummins-functions.

Chapter 1
Grandpa's explanatory framework

In this chapter I will outline the theoretical framework on which this dissertation is going to base its discourse, as well as to indicate those issues about which I am not going to be directly concerned, even if they have some weight in the discussion. First, I present the framework that fixes landmarks of the cognitive-science theorising I am interested in. I shall label this framework *Grandpa*'s model, as a logical addition to the use that Fodor (i.e. Fodor 1987) has made of the figures of Granny and Aunty in his papers. Fodor has referred to Granny as the voice of common sense; she is an Intentional Realist, that is, she believes that we have minds, beliefs and other sort of intentional states. Aunty, on the other hand, speaks as the voice of the establishment within psychology, and lately is reputed to be a Connectionist (having been a behaviourist in her younger years). Fodor invites Granny or Aunty when he wants to give an 'authorized' opinion. However, we lack a character for the object of their opinions, namely, the enterprise of cognitive science. Taking the Addams Family as a model, I think that the best characterization for cognitive science is that of Grandpa, an eccentric and boisterous self-granted scientist with foolish but colourful ideas about the mind. The choice is of course personal and innocently contentious.

My task here then is to present the basic assumptions and supporting considerations that make up cognitive science, i.e., Grandpa's, framework. Even though the notion of a coherent framework is an idealization, there are some common elements of the whole enterprise that can be extracted and used in our discussion. I will not give a full account of the model; rather, what I wish to do is to present the central assumptions that will concern us here.

On the other hand, I shall sketch some issues whose discussion I will omit. These correspond to some questions related to the notion of psychologically reality for cognitive theories. One concerns the nature of the knowledge to be attributed to a person; specifically, the contentions that face the question "do the principles that underlie the execution of our cognitive capacities correspond to a theory of the domain?" Another issue focuses on the relation between the person and the knowledge attributed: "Is the relation between the person and principles that underlie our cognitive capacities a belief-like relation?" All these issues are contentious though they do not really affect the main points of this dissertation. Even if the proposals that I examine do require the resolution of these problems, the points I wish to make are not really affected by the outcome of the discussion. Therefore, once presented they will be assumed for the remainder of the discussion.

1.1.Grandpa's framework

The ground over that this dissertation is intended to cover is offered by what is known as *cognitive science*, the experimental and theoretical paradigm that rose from the limitations of *behaviourism* and which has dominated psychology since the early sixties. It is obviously possible to be skeptical about whether cognitive science exists as a coherent enterprise. Some have voiced concerns that there is no agreed-upon research paradigm among the fields that comprise cognitive science (Gardner 1985; Miller, Polson and Kintsch 1984). However, I support Von Eckardt (1993) on the view that there is far more implicit agreement among cognitive scientists (in all disciplines) as to their goals and their basic assumptions than the sceptics believe. Grandpa's model could be described as a fairly coherent, transdisciplinary framework of shared commitments and this reconstructed set of commitments is substantially in accordance with what everyone considers to be the clear cases of cognitive-science research. Specifically, what all cognitive scientists have shared over the years is a sort of commitment to an approach to the study of mind, rather than some specific set of theories, explanations or laws. It is not that they do not exist. The point is that commitment to such theories, explanations or laws varies. The basic assumption about cognitive-science enterprise that I will sustain is that there is such a thing as a community of cognitive science, that this community shares certain commitments and that these commitments function as a framework for research. Taken together, these factors contribute to the coherence of the scientific activities that fall under the label of cognitive science.

I will now present the basic assumptions that underlie the notion of the cognitive-science framework, fundamentally as Von Eckardt has presented them. Following Von Eckardt the framework could be said to have four elements: a set of assumptions that provide a pretheoretic specification of the domain under study; a set of basic empirical research questions; a set of substantive assumptions that embody the approach being taken in answering the basic questions, and that constrain possible answers to those questions; and finally a set of methodological assumptions. I have included one more element in this account of the cognitive-science framework, the explanatory strategy employed in cognitive science that constrains the way in which the answers are formulated. I have also added and modified some of the substantive and methodological assumptions.

1.1.1.Domain-specifying assumptions

What is the object of study of cognitive science? According to Von Eckardt, the object of study of cognitive science is "the study of the human adult's normal, typical cognition". However, I think that we can easily include in the paradigm the areas of Artificial Intelligence, the study of the cognition of infants, that of nonhuman animals, and the study of abnormal cognition. Basically, cognitive scientists want to know how cognition typically works in humans, how it varies from individual to individual, how it varies in different populations, how it varies across cultures, how it develops, how it goes wrong in neurologically impaired patients, and how it is realized in the brain.

More concretely, the domain of cognitive science consists of cognitive capacities. Although it is not clear what a cognitive capacity is, there does seem to be fairly widespread agreement on what constitutes the clear cases within the general class. They include our capacity to use language (perceive it, comprehend it, produce it, translate it, communicate with it, etc.), to perceive using the visual system, to apprehend music, to learn to solve problems (to reason, draw inferences), to plan action, to act intentionally, to remember and to imagine. Each of these subclasses can be broken down into far more specific capacities; for example, the capacity to remember includes the capacity to remember faces, episodes from the past, lists of nonsense syllables, facts, concepts and so on. There are also several borderline phenomena whose membership in the domain of cognitive science is, at present, unclear. These include our capacity to acquire skills with a significant motor component and our capacity for nonlinguistic auditory, tactile, olfactory and gustatory perception.

Conceived pretheoretically, these human cognitive capacities have a number of important properties. These include the fact that each capacity is *intentional*, that is, it involves states that have content or are about some object, property or relation of the world. Second, virtually all of the capacities are *pragmatically evaluable*; that is, they can be exercised with varying degrees of success. Third, when successfully exercised, each of the evaluable capacities has some *coherence* or cogency. Fourth, most of the evaluable capacities are *reliable*; that is, typically, they are exercised successfully (at least to some degree) rather than unsuccessfully. Finally, most of the capacities are *productive*; that is, once a person has the capacity in question, she is typically in a position to manifest it in a practically unlimited number of novel ways.

The cognitive capacities make up a theoretically coherent set of phenomena, or what Von Eckardt labels a *system*. This means that, with sufficient research, it is possible to achieve a set of answers to the basic questions of the research framework thus constitute a unified theory that is empirically and conceptually acceptable.

1.1.2.The basic questions

I endorse Von Eckardt's view that a research framework gets its direction from the questions it attempts to answer. The questions that cognitive scientists wish to answer about cognition could be formulated in four basic and distinct ways. Questions of each kind can be raised for each of the cognitive capacities included in its domain. The questions leave blank what is to be filled in with an expression referring to some particular cognitive capacity (such as 'recognize words' or 'perceive a scene'). The first basic question is:

> Q1 For the (normal, abnormal adult/infant/animal/artifact) what is precisely the capacity to __________?

At the beginning of the research process, we can describe the cognitive capacities of interest in a fairly determined manner. However, we do not know precisely what each capacity is, either empirically or theoretically. Deeper understanding can be gained in four ways. First there is the help of philosophical reflection on our common-sense conception of the capacity. For example, if the object of investigation is the capacity to perceive images, we can ask the following about this specific capacity: In what respects is it intentional, pragmatically evaluable, coherent, reliable and productive? Second, we have the way of seeking greater empirical understanding of the capacity. How does such capacity manifest itself in particular circumstances? Third, we can gain deeper understanding of the capacity by reconceptualizing it in terms adopted in the research framework; for example we can take the use of mental images in information-processing terms. Finally, we can investigate the scope and the limits of the capacity. When a person fails to exercise a capacity successfully what form does this failure take?

The second question takes the following form:

> Q2 In virtue of what does a (normal, abnormal adult/infant/animal/artifact) have the capacity to _______ (such that this capacity is intentional, pragmatically evaluable, coherent, reliable and productive)?

As Von Eckardt notes, people do not have their cognitive capacities through magic. There is something about the mind/brain in virtue of which they have the capacities they do. This question addresses the issue of the psychological resources that make any given capacity possible.

> Q3 How does a (normal, abnormal adult/infant/animal/artifact) typically (exercise his/her/its capacity to) _______?

The difference with Q2 is the fact that Q3 seeks a dynamic account of the same answer. How are the mental resources described in answer to Q2 actually deployed when a person *exercises* the capacity in question? What stages or steps does a cognitive system typically go through when the capacity is exercised successfully?

> Q4 How does the (normal, abnormal adult's/infant's/animal's/artifact's) capacity to ______ interact with the rest of his/her or its cognitive capacities?

So far each of the capacities has been treated in isolation. This question is a request for an integrated account. How does our capacity to perceive interact with our capacities to remember?

1.1.3.Substantive assumptions

What distinguishes cognitive science is that it is interested in answering the previous questions in a certain way. And this way is what the substantive assumptions specify. There is substantial agreement among cognitive scientists that the research framework of cognitive science in general is committed to what has been called the computational metaphor. As a matter of fact, Grandpa's paradigm stems directly from the computational approach to the mind. The advent of computers provided a piece of invaluable insight to psychologists. Computer science showed that it was possible to explain the intelligence behaviour of a complex system without presupposing the intelligence of its components. This provided the idea that cognitive processes could be explained by a computer metaphor, comparing mental processes with the sorts of informational processes carried out by computers. The mind began to be treated essentially as an informational processing system. The *mind* could be *objectively* studied by providing an indirect account of it. And the two Cartesian birds, mind and body, mated. Physicalism was respected as long as mental representations were analyzed as *functionally interpreted* states of physical systems, and mentalism was also respected, by conjecturing that the mind could be objectively studied by reconstructing its structure and processes as an informational processing system. This approach had invaluable properties regarding the rationalist or materialist excesses of the past: it avoided dualism, by taking the mind to be an interpreted physical system and it avoided the reduction of the mental to the physical, by granting explanation at the level of representation. In sum, the computational approach to the mind helped explain mentality.

The following presentation is sketchy and partial, but my aim is only to outline the theoretical landmarks of my discussion. Therefore, even if it is an incorrect picture for some, it constitutes the rules by which I want to play. Additionally there is widely held agreement in the formulation of the assumptions, though there is much less accord on what is being assumed. The problem is that the relevant concepts are quite open-ended and vague, as Von Eckardt points out (ibid., p.9). Thus, the presentation should be considered to state the basics.

The first of the two basic tenets of Grandpa's framework is:

A1: Cognitive states are individuated by their causal role.

As a first approximation, to say that something has causal powers is to say that it has the capacity to cause one or more effects. In other words, X has the capacity to cause an effect (say, Y's having a certain effect property EP) if and only if X has some causal property CP such that there are nomologically possible conditions C under which X's having CP conjoined with the presence of C would cause Y to have EP. What are the causal powers? Suppose that X has three distinct causal capacities: The capacity to cause Y's having EP1, the capacity to cause Z's having EP2, and the capacity to cause W's having EP2. It is natural to say that X's causal powers simply are X's capacities (in this case three) to cause certain individual effects. Under what conditions will two things differ in their causal powers? The identification of causal powers with causal capacities suggests that two things X and Y will be different in their causal powers only if X and Y differ in their capacity to cause one or more individual effects, that is, only if X has a capacity to cause an effect that capacity Y does not share (or vice versa).

This property of individuating by the causal powers explains why Grandpa is a functionalist, that is, somebody who believes that the various states of some cognitive system can be understood in terms of their functional or causal role with respect to inputs to the system, the outputs from that system, and the other states within the system causally connected to the states to be explained.

> A2: Cognitive processes are formal and representational at the same time.

According to the formality condition put forward by Fodor (1981b) we can say that the computational processes apply to representations in virtue of the formal properties (the syntax) of representations. Computation by (the form of) mental representations is essentially truth preserving. In other words, computation is effected according to the formal properties of mental representations but meaning is preserved across transformation so long as mental representations preserve its contents. This property of assuming the dual character of mental representations makes Grandpa a proponent of the representational theory of mind. According to this theory, cognitive processes are to be viewed as the manipulation of representations, which are representations in virtue of standing for some object, property or relation of the world. On the other hand, mental representations have a formal character, a syntax, and it is in virtue of this character that they have mental causal powers.

The remaining substantive assumptions supporting cognitive science derive from this scheme and from applying concepts of the computer metaphor. Basically, the logical development is extended by the following assumptions:

> A3: Psychological states are (to be known as)[1] symbols, that is, structured, semantically interpretable objects.

> A4: Cognition is (to be known as) computation, that is, an effective procedure by which symbols are transformed according to a specified set of rules or instructions, and cognitive psychology is the study of the various computational processes whereby mental representations are constructed, organized, interpreted, and transformed.

[1] The parenthetical qualification is included to accommodate both a strong ontological view about what psychological states are (Pylyshyn [1984] takes, for example, the computer metaphor as a realist hypothesis rather than as a metaphor) and as an epistemological thesis (Hardcastle [1996, p.74] argues that the best we can hope to say about any psychological states is that we understand it in virtue of a computational model).

Cognitive abilities are then conceptualized as a function identified in informational-processing terms. In this sense, it is customary to understand computational satisfaction -of a system with respect to a task- by interpreting physical systems as instantiating some mathematical function. For example, a physical system satisfies the identity function if there is an interpretation of the function that maps the physical processes of the system onto the identity function. More formally, this can be stated as in Hardcastle (1996):

A5: A cognitive function is individuated as an input-output mapping:

(i) A system satisfies a function by transferring inputs (*I*) into outputs (*O*)

(ii) There is a process that relates the inputs with the outputs in a given way for that system ($P{:}I \rightarrow O$)

(iii) There is an interpretation function that maps the inputs, processes and outputs onto a pattern ($P(i,o)=f(A)$)

A cognitive system that accords with these assumptions is considered to be an information-processing system, a system that trades with symbols and their transformations. Cognitive capacities are then understood as a certain way of transforming symbols, and we count such transformations leading from inputs to outputs as *computations* -i.e. as stages in step-by-step transformations of the symbols interpreting inputs into the symbols interpreting outputs. Cognitive science merely tries to use these theoretical tools to explain cognition, and this explanation is applied to cognitive systems assuming a structure in levels of explanation.

1.1.4.Explanatory strategy

According to the previous substantive assumptions, the cognitive capacities consist, to a large extent, of a system of computational and representational (i.e. information-processing) capacities. These assumptions constrain what counts as a possible answer to each of the basic questions. Thus, in endorsing the substantive assumptions, cognitive scientists limit themselves to entertaining only answers to the basic questions that are formulated in information-processing terms. However, there are constraints on how the substantive assumptions are brought to bear to answer the basic questions of the field, and this is provided by the *explanatory strategy* with which the cognitive science paradigm accords. This notion has two legs, as it has two words. One is obtained by the application of the questions to the model of levels of explanation, and provides the sort of *explanation* offered in cognitive science. The other, what is known as functional analysis, provides the *strategy*.

1.1.4.1.Levels of explanation. It is a central assumption of contemporary philosophy of science that complex systems are to be seen as typically having multiple levels of organization and explanation. The standard model of the multiple levels of a complex system is 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. We thus often have explanations of a system's behaviour at higher (coarser-grained) and lower (finer-grained) levels. The behaviour of a complex system might then be explained at various levels of organization, including (but not restricted to) ones which are biochemical, cellular, and psychological. Similarly, a given computer can be analyzed and its behaviour explained by characterizing it in terms of the structure of its component logic gates, the machine language program it is running, the accounting task it is performing, and so on.

Higher-level explanations allow us to explain -as a natural class- things with different underlying physical structures -that is, types which are multiply realizable (see, e.g., Fodor [1974], Pylyshyn [1984], esp. chapter 1, and Kitcher [1984], esp. pp.343-6, for discussions of this central concept). Thus, we can explain generically how transistors, resistors, capacitors, and power sources interact to form a kind of amplifier independent of considerations about the various kinds of materials composing these parts, or account for the relatively independent assortment of genes at meiosis without concerning ourselves with the exact underlying chemical mechanisms. Similar points can be made for indefinitely many cases: how an adding machine works, an internal combustion engine, a four-chambered heart, and so on.

This strength of capturing generalizations has many facets. One is, of course, that higher-level explanations typically allow for reasonable explanations and predictions on the basis of far different, and often, far less detailed information about the system. Thus, 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 to an arithmetic problem while remaining ignorant of the electrical properties of semiconductors. What is critical here is not so much the fact of whether a given higher-level phenomenon is actually implemented in the world in different physical ways. Rather, it is the indifference to the particularities of lower-level realization that is critical. To say that the higher-level determination of process is indifferent to implementation is roughly to say that, if the higher-level processes occurred, regardless of implementation, this would account for the behaviours under consideration.

In the case of psychology, or cognitive science, many authors have argued that an adequate psychology will comprise explanations at different levels. More precisely, any explanation in cognitive science should accord to what we have called the "classical cascade" (Franks 1995), an explanation that *cuts across* the different levels that are normally proposed. Marr (1982) is the usual reference in this regard. He proposed three distinct, but interrelated, levels, which he called the *computational*, *algorithmic* and *implementation* levels. This terminology has been disputed, though generally accepted; thus, I will change the label of the computational one by that of the task level.

> *Task Level.* It is the top level of an explanatory cascade. It identifies what sort of *function* the system is performing. It has been also called, the *why* level, since it should account for the goals and logic of the system's behaviour. Some see it as providing an abstract formulation of the information-processing task which defines a given psychological ability, together with a specification of the basic computational constraints involved. Others see it as only a specification of a function in extension (Peacocke 1986), that is, giving the semantics of the function (Franks 1995). If we take the task level as specifying the function in extension, or the semantics of the function, which, following Clark (1990) might be termed "official dogma", then we can have a neat differentiation with the level of the algorithm. However, we could take level 1 to specify the information-processing

task, that is, adding the details of the informational contents involved in the functional satisfaction The distinction between this characterization, which I will call it the 'function-composition view', and that of the algorithm is sometimes tricky Be it as it may, the level should provide the necessary and sufficient computational basis for any creature faced with a given cognitive task

Algorithm level. This is the *how* level It establishes the mechanisms of the cognitive system and its transformations, i e , representations and symbolic processes It takes into account the computational constraints in specifying the psychological processes, or computations, by which the task is actually performed, which may differ in different creatures These processes are defined in terms of a particular system of representation, which can be shown to be reliable with reference to the top-level constraints There is a one-to-many relationship between any function and the set of algorithms that can compute it There is also a one-to-many relationship between any algorithm and the set of physical implementations (at level 3) that can realize an algorithm [2]

Implementation level. The study of the way in which computation and intention are

[2] There is however a problem in identifying what is an algorithm-in-a-cognitive-structure According to Franks (1995) there are two approaches The first is to provide a formal specification of the algorithm An algorithm is finite (it terminates after a finite number of steps), definite (each step in the sequence must be rigorously defined) and effective (all of the operations performed must be 'sufficiently basic so that they can, in principle, be done exactly and in a finite length of time by a man using a pencil and paper [Knuth 1973, p 6]) In particular there is no commitment that the function be computable within a limited time by an automaton with limited memory resources, nor any commitment to the psychological plausibility of the algorithm The second approach, makes just such commitments An algorithm in the first sense is a completely specified routine or procedure that can be carried out in a finite number of steps to solve a problem The problem is how to specify such a notion Some see it as a direct specification of the steps a system follows Others see it differently Pylyshyn (1984, pp 89-90) identifies a more specific level for such specifications, the *program*, which for him is the encoding of a particular algorithm in some programming language An *algorithm* is a more abstract notion than *program* in a variety of ways, and therefore it is possible as well to have different programs in the same language for a particular algorithm In this case, programs are viewed as differing in unessential respects, for example, they may differ in the order in which they do certain minor operations Then, for Pylyshyn, an algorithm is related to a program approximately as a proposition is related to a sentence Moreover, an algorithm might not be constrained by being of a sequential nature, since there are connectionist algorithms that should be explained avoiding the notion of sequence

instantiated in a physical system.

As an illustration, Marr applies this distinction to the levels of theorizing about a well-understood device: a cash register. At the computational level, "the level of what the device does and why", Marr tells us that "what it does is arithmetic, so our first task is to master the theory of addition" (1982, p. 22). Yet, at the level of representation and algorithm, "we might choose Arabic numerals for the representations, and for the algorithm we could follow the usual rules about adding the least significant digits first and `carrying' if the sum exceeds 9" (1982, p. 23). At the implementational level, we face the question of how those symbols and processes are actually physically implemented.

There are other versions of the levels of explanation. Zenon Pylyshyn's variant of the "three levels" view (Pylyshyn 1984) is similar in many respects. As he says, the "main thesis" or "basic working hypothesis" of his book *Computation and Cognition* is that within the study of cognition, "...the principle generalizations covering behaviour occur at three autonomous levels of description, each conforming to different principles. These principles are referred to [here] as the biological (or physical) level, the symbolic (or syntactic or sometimes the functional) level, and the semantic (or intentional) level" (ibid, p.259). Or again: "...we will see that there are actually two distinct levels above the physical or neurophysiological level - a representational or semantical level and a symbol-processing level" (ibid., p.24). Thus we have as levels the biological (Marr's implementational), the symbolic or syntactic (Marr's algorithmic), and the semantic (Marr's computational). As for Marr, the distinction of levels is seen as having some of its standard roles -- e.g. being used "to account for certain kinds of generalization and to give a principled account of certain constraints and capacities" (ibid., p. 39).

Simon (1981) provides a proposal that is couched explicitly in terms of the rationality of systems. For him, substantive rationality is a matter of the fit between a system's goals and its environment; an account of substantive rationality will characterize *what* the system *does* with respect to the environment. A system is substantively rational in the case it manages to act so as to achieve the satisfaction of its goals, as much as possible, although *how* it manages to do so is left open. This is exactly the opening that an account of *procedural* rationality fills, an account of the *procedures* by which that process of fitting actually works. Therefore, whereas an account of substantive rationality will characterize adaptively beneficial behaviour for a system with particular goals in a particular environment, an account of that system's *procedural* rationality would instead characterize how it manages to implement such adaptive behaviour.

There is also the distinction made by Chomsky between competence and performance concerning our linguistic abilities. Even if the notions are precisely applied in the linguistic domain, many authors have extended the distinction to other domains within cognitive science. The central idea of a competence theory is that of an idealization about the systematic behaviour of the organism under idealized circumstances. The competence theory of a domain of behaviour is seen as a formalization of the behaviour via "a system of rules that in some explicit and well-defined way assigns structural descriptions to sentences "(Chomsky 1965, p.8-9). As with other distinctions, there is an independence from considerations about actual production or what Chomsky calls *performance*. However, the distinction is rather peculiar, since Chomsky seems sometimes to appeal to competence as if talking about some sort of *optimality*, abstracting away problems of incomplete information, real-time constraints, memory load etc. This might be a non-trivial difference with other accounts.

There is another version that approaches my perspective, the *task/process* distinction offered by McClamrock (1995). For McClamrock there is an abstract and idealized explanatory level where the theorist explains the behaviour that a system must achieve; the specification where the system is "doing what it is for", or as he calls it, the *task* to be handled by the system to achieve its goals. Such an account is seen as compatible with various *processes* accounts.

1.1.4.2.Functional Analysis. We have seen one of the two legs necessary for a cognitive explanation. In this sense, the levels-of-explanation strategy defines the structure of a cognitive explanation. A functional analysis provides, on the other hand, the methodological strategy to obtain the explanation. Succinctly, a functional analysis is a type of explanation in which some system is decomposed into its component parts and the workings of the system are explained in terms of the capacities of the parts and the way parts are integrated with one another.

Cummins is one of the authors who has developed such a strategy. He argues that the characteristics of the answers to cognitive-psychology questions correspond to what he labels a property theory. A property theory explains the properties of a system not in the sense in which this means 'Why did S acquire P?' or 'what caused S to acquire P?' , but rather, 'What is it for S to instantiate P?' or, 'In virtue of what does S have P?' He contrasts the property theories with transition theories. The characteristic question answered by a transition theory is 'Why does system S change states form s-1 to s-2?', whereas the characteristic question answered by a property theory is: What is it for system S to have property P?

For Cummins the usual strategy for answering types of question such as 'In virtue of what S has P' is to construct an analysis of S that explains S's possession of P. This is done by appealing to the properties of S's components and their mode of organization. The process often has a preliminary analysis of P itself into properties of S or S's components, but that does not matter here. This sort of analysis is recursive, since a given characterization may appeal to properties or components that require analysis themselves. The analysis of a *system* in components is called compositional analysis by Cummins, to distinguish it from analysis of a *property*, which he calls functional analysis when the property is dispositional and property analysis when the property is not dispositional.

Functional analyses consist then in analysing a disposition into a number of less problematic dispositions in such a way that programmed manifestation of these analysing capacities amounts to a manifestation of the analyzed disposition. The analysis of dispositions goes together with the componential analysis of the disposed system, analysing dispositions being capacities of system components. Componential analysis of computers, and probably brains will typically yield components with capacities that do not figure in the analysis of capacities of the whole system. (Cummins 1983, p.29) Finally, a complete property theory for a dispositional property must exhibit the details of the target property's instantiation in the system (or system type). *Analysis* of the disposition is the first step; *instantiation* is the second (Cummins 1983, p.31).

Any functional analysis can be expressed in flow chart form, as some sort of program, the elementary instructions specifying the analysing capacities and the input-output properties of the whole program specifying the analyzed capacity. Thus rather than say that an analysis is instantiated in a system, we can say, equivalently (and this is basic for Grandpa's model) that the system executes the (or a) program expressing that analysis. Therefore we can say that system S cannot execute program P unless S is so structured as to ensure the transactions specified in P. From this perspective we can see that when we show how an analysis is instantiated in a system S, what we come to understand is how it is able to execute the program specifying the analysis.

Hence, for 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 capacity of a containing system. When a capacity of a containing system is appropriately explained via a functional analysis, the capacities emerge as functions. More concretely, we explain cognitive capacities whose inputs and outputs are specified via their semantic interpretations. The capacity to add, for example, is the capacity to produce as output 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 (or outputs) count as the same -i.e. as tokens of the same type- in 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 the inputs. Then the outputs must be inferable from the inputs in a inferentially manner (the law specifying the capacity is a rule of inference).

Functional analysis is in a sense what Dennett (1995) has termed the reverse-engineering strategy. According to Dennett, functional analysis is a form of "reverse engineering" so long as we are trying to explain cognitive capacities by building (or explaining the functional principles of) systems that have such capacities. Reverse engineering must discover the functional principles of systems that have already been designed and built by nature -plants, animals, people- by attempting to design and/or build systems with equivalent functional capacities. Ordinary direct engineering, on the other hand, applies the laws of nature and the principles of engineering to the design and building of brand new systems with certain specified functional capacities that we find useful: bridges, furnaces, airplanes. Such is the logical strategy of, for example, Artificial Intelligence, whose goal is to create intelligent machines. It is quite obvious that standard engineering principles should guide the research activity. First one tries to describe, as generally as possible, the capacities or competences one wants to design, and then one tries to specify, at an abstract level, how one would implement these capacities and, finally, with these design parameters tentatively or defeasible fixed, one proceeds to the physical realization. For Dennett this methodology is a straightforward application of standard "forward" engineering to the goal of creating artificial intelligences. This is how one designs and builds a clock, a car or a camera. It is a top-down design process, although there are revisions from bottom up:

> *Dennett-Forward engineering:* The idea is that one starts with the ideal specification of an agent in terms of what the agent should know or believe, and want, what informational powers it should have and what capacities. It then becomes an engineering task to design such an intentional system, typically by breaking it up into organized teams of subagents, smaller, more stupid homunculi, until finally all the homunculi have been discharged, replaced by machines.

This is, according to Dennett, the central strategy of research in AI. Reverse-engineering can be specified, on the other hand:

> *Dennett-Reverse engineering:* The interpretation of an already existing intelligent artifact or system by an analysis of the design considerations that must have governed its creation.

Reverse-engineering takes a cognitive system to be a designed system. When confronted with a capacity we tend to infer that it is obtained by a special mechanism designed to obtain the right solution. Therefore, a cognitive system must be explained by going *backwards*, that is, specifying first the capacity and then hypothesizing how such capacity could have been implemented. In other words, reverse-engineering prescribes that the way to find a solution is to find a design to satisfy the solution:

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

In this regard, for Grandpa, optimality should be the default assumption of cognitive analysis. The fact is that if cognitive scientists cannot assume that there is a good rationale for the features they observe in cognizers, they cannot even begin their task (Dennett 1995). This strategy normally over-idealizes the design problem, by presupposing first that one could specify the function of some system and second that this function is optimally executed by the cognitive machinery. However, even if optimality is the default assumption, cognitive scientists know, as Dennett has put it, that Mother Nature is not an optimal engineer. Therefore, we need something more for this principle to work, namely, a guarantee that the system has to find the same solution that theorists do. This guarantee is provided by robustly believed empirical regularity, what we could call the:

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

Therefore, we can trust that our solutions will be correct in virtue of the law by which every solution to a problem that the functional capacity has to meet will be similar.

An alternative hypothesis would be that the brain is not an optimizing system. The brain could in this case be an effectively "satisficing" system (cf. Simon 1981), a system that finds the *most available solution to a functional requirement with the resources at hand.* In my dissertation I will try to show that there is certain empirical evidence from different sciences dealing with the brain and cognition that are fairly persuasive 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 on using sundry, sub-specialist (unmotivated by the function they subserve) and redundant mechanisms to obtain (in a non-unique manner) a particular goal.

Dennett argues that we can apply the reverse engineering strategy to both natural systems as well as artificial ones. However, he believes that its application to natural systems carries a degree of complexity and is must be supplemented by what he calls a bottom-up reverse engineering:

> *Dennett bottom-up reverse engineering:* One starts with a specification of the local dynamics of basic elements and then tries to move towards a description of the behaviour of the larger ensembles.

For Dennett the paradigm of such a strategy is the enterprise of Artificial Life.[3] I think that Dennett is wrong in labelling the strategy that pervades cognitive science *reverse engineering*. One thing is engineering, another is modelling. Engineering is designing and constructing artifacts that accomplish a mission. As it is defined in engineering, reverse engineering is "the process of duplicating an item, functionally, by analysing it as a complete artifact". Therefore, until the artifact is not built we cannot talk of engineering. This is not a mere terminological contention. To cross the bridge between modelling and engineering we need an important step: adequacy of the analysis. I propose to call such a strategy:

> *Reverse modelling:* The interpretation of an already existing intelligent artifact or system by analysis of the design considerations that must have governed its creation.

On the other hand, I believe that it is also misleading to call the strategy that underlies Artificial Life *bottom-up reverse engineering*, for in this case the problem is not the *engineering*, but the use of the *reverse* term. As a matter of fact, the research that the Artificial Life labs undertake establishes an engineering framework as a point of departure, with local rules, and then the artifact is allowed to search for solutions to problems. Such dynamics is anything but reverse; it is perhaps rather more "forward" than the direct engineering to which the term forward is normally applied. Given that it would be confusing to use the term in the case of Artificial Life, I will use the following term:

[3] Artificial Life comprehends various lines of research which look for the emergence of top-level functions by the implementation of local rules and mechanisms in artifacts.

> *Emergent engineering:* The obtainment of engineering solutions by the specification of local dynamics of basic elements, and self-organizing ensembles.

Apart from these strategies, there is one more that we will use in later chapters but that should be defined here; otherwise it will not be well understood. The idea is that while we have defined the cognitive strategy as a sort of modelling, namely, reverse modelling, there might also be another modelling that could be considered and that will be basic for my proposals. I refer to the strategy that once the dynamics of basic natural elements is revealed, we proceed by a modelling of natural evolution to reveal how the system avails itself to find solutions:

> *Emergent modelling:* The specification of the dynamics of basic elements and the elucidation of how the system finds solutions to problems.

1.1.5.Methodological assumptions

A framework of shared commitments typically contains a methodological component. Often the methodological assumptions of a research framework concern which *specific* methods are appropriate for conducting research. In fact, many of the methodological assumptions associated with the subdisciplines of cognitive science are probably of this sort. However, because cognitive science itself is an umbrella term encompassing these various distinct disciplines, its methodological assumptions have a more general character. The first assumption I wish to consider is a basic methodological principle of cognitive science and which can be stated in the following form:

> M1: Behaviour is the standard for individuating the mechanism that accounts for it.

The assumption of Grandpa's model is that a behavioural efficiency identifies and individuates an internal capacity in the mind/brain. In other words, Grandpa's model confers upon behavioural descriptions the role of identifying internal capacities:

> M2: There exists a partitioning of cognition in general into individual cognitive capacities such that each of these individual capacities can, to a large extent, be successfully studied in isolation from each of the others.
>
> M3: Functional descriptions identify genuine cognitive capacities.

Once we have already identified the function, then Grandpa's model goes on to assume that functional analysis is the right way to identify the cognitive capacity in isolation. 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

Then, the next two methodological assumptions claim that cognitive science focus on individual cognition:

> M4: Human cognition can be successfully studied by focusing exclusively on the individual and her place in the natural environment. The influence of society or culture on individual cognition can always be explained by appealing to the fact that this influence is mediated through individual perception and representation.
>
> M5: Human cognitive capacities are sufficiently autonomous from other aspects of mind so as to be easily studied in isolation.

M4 and M5 assert thus that it is acceptable to study the individual cognition in isolation from social context and other aspects of mind:

> M6: Although there is considerable variation in how adult human beings exercise their cognitive capacities, it is meaningful to distinguish normal from abnormal cognition.

> M7: Although there is considerable variation in how adult human beings exercise their cognitive capacities, adults are sufficiently alike when they cognize that it is meaningful to talk about a typical adult cognizer and it is possible to arrive at generalizations about cognition that hold for all normal adults.

M6 and M7 allow a distinction between the research program directed at the normal, typical adult cognition and that aimed at studying abnormal cognition. This approach to the study of cognitive systems is largely determined by the fact that cognitive scientists seek to explain cognition in information-processing terms, but there are other features to the approach. For one thing, cognitive science is situated within the scientific enterprise as a whole:

> M8: In choosing among alternative hypothesized answers to the basic questions of the research framework one should invoke the usual canons of scientific methodology. Thus, all answers must be empirically justified.

Additionally, a complete understanding of cognition will require the conceptual and methodological resources of the different subfields of cognitive science:

> M9: A complete theory of human cognition will not be possible without a substantial contribution from each of the subdisciplines of cognitive science.

These include, following Gardner (1985), Psychology, Linguistics, Computer Science/Artificial Intelligence, Philosophy and Neuroscience.

1.2.What this dissertation is not about

In this dissertation, I am concerned with the notion of psychological reality, especially those issues that are directly concerned with the question of "what is it for a theory to be psychologically real". Specifically, I will be concerned with what we have called Quine's challenge: how can it make empirical sense to suppose that an ordinary cognizer has internalized one set of axioms, rather than an alternative extensionally equivalent set? I shall avoid issues concerning the nature of the relation between the subject and the attribution of knowledge, since this concerns the requirement for knowledge to be conscious or consciously available.

1.2.1.Nature of the theory attributed

Generally, the dominant explanatory strategy in cognitive science is to posit an internally represented knowledge structure (Stich and Nichols 1992), which is a set of rules or principles or propositions that allows the individual to execute the capacity to be explained. The rules or principles or propositions are described as the agent's theory of the capacity. However, this raises some problems that have not been fully resolved or consensuated. The basic question is whether to take seriously the idea that attributing a theory implies that it should be internalized as a set of axioms and theorems, or whether it is just a description of the practice or ability to be explained. Do we have to think of the body of knowledge as structured this way? It seems that the majority of cognitive scientists work on the hypothesis that knowledge might be structured in the mind of cognizers as a theory. The idea has two modern roots, one in philosophy and the other in cognitive psychology. The philosophical origin of considering a body of common sense knowledge as a theory appears in Sellars (1956) in the sense that our common-sense knowledge is theory-like in that the concepts are embedded in a framework of laws, at least tacitly appreciated. Its origin in cognitive psychology goes back to the seminal paper of Woodruff and Premack (1978) with their proposal that chimpanzees could have what they called a «theory of mind».

The problem has been to specify clearly what an attribution of a theory amounts to. For example, Woodruff and Premack's criteria of theoreticity consisted in a "System of inferences" that goes beyond empirical generalizations of directly observable aspects, with: a) states that are not directly observable, and b) the system can be used to make predictions. A simple proposal has been advanced by Fodor (1992), which states that a theory is individuated by:

a) An ontological inventory.

b) Certain empirical generalizations ("covering laws").

The individuals recognized in *a* provide the domain in which *b* is claimed to hold. Hence, by theory we could understand a list of elements each of which either makes a claim of a kind expressible in public language or expresses a rule of inference. However, for many theorists, a theory, or more precisely, a psychological theory should be more than a conjunction of such individual elements. Indeed, the idea could be open to criticism since we can acquire knowledge or pursue speculation by making inferences, so it is not the case that any instance of 'inferring' amounts to 'theorizing'. In this sense, a theory would embody information not as a mere list but in some articulated form. Heal (1996) makes a contrast that may be interesting here. Consider two schematic types of medical knowledge about diseases and their likely developments. We can imagine a wise woman who is able to offer some general remarks about symptoms and their severity ('A high fever is often dangerous', 'Laboured breathing is generally a bad sign', 'Many skin rashes are trivial') and can also, surveying a patient select from among the visible symptoms the ones which are, in fact, important in the particular case. Thus she can correctly say, 'this patient will recover, because the fever has broken sweat' or in another case, 'this patient will not recover, because the breathing is now very laboured'. Our imagined wise woman however, is unable to say why the breathing is not as severe in that case. We can also imagine another practitioner, say a doctor with modern training, who possesses what the wise woman lacks, namely a framework within which the various symptoms are listed and systematically related to each other; she cannot only predict what will happen in a particular case, but can distinguish that case from among other possible ones. Some theorists (Wellman 1990; Boterill 1996; Heal, Gopnik and Wellman 1992) would only attribute the notion of theory in this second case. According to Botterill, this idea would comprise the following:

Prediction, explanation and interpretation. A theory makes predictions about a wide variety of evidence, including evidence that plays no role in the theory's initial construction. It also produces interpretations of evidence, not simply descriptions of evidence and generalizations about it. We use the theoretical knowledge to explain and predict actions and reactions of people or objects. In other words, it provides a separate causal-explanatory level of analysis that accounts for evidential phenomena.

Implicit or explicit ontological commitments. We might grasp concepts such as belief or impetus after the assimilation of a given theory.

Abstraction, unobservables. Theoretical constructs are postulated abstract entities which differ from evidential vocabulary. Theoretical constructs need not be definitively unobservable, but they must appeal to a set of entities removed from, and underlying, the evidential phenomena themselves. Thus they have abstractness, but they do not restate the data.

Counterfactual support. Theoretical knowledge distinguishes genuinely law-like (or nomic) principles from generalizations which are true by mere contingency.

Systematization-Coherence. Theoretical constructs do not work independently, they work together in systems characterized by a certain structure or coherence. The power of knowledge systems high in theoreticity is to reduce the number of laws and principles needed to account for the data, replacing a large class of narrow-scope principles with a smaller class of more general ones (Collin 1985, p.61). The body of knowledge works by addition rather than projection.

Some theorists, however, do not constrain the attribution so much as to equate it to a theory in the just presented sense. The point has been interestingly sketched by Blackburn (1992). He suggests that if we are good at something then we can be thought of as making tacit use of some set of principles that could, in principle, provide a description of a device, or possibly a program for the construction of a device, that is also good at it. This will be true whether the skill is understanding others, recognizing syntactic correctness, perceiving spatial objects, or riding a bicycle. If this conception of theory is on offer, then for Blackburn, it will be difficult to avoid describing our linguistic understanding as theoretical, although it will not necessarily be we who theorise. The existence of theoretical principles, thought of like this, is, for Blackburn, simply a consequence of first the skill, and secondly of the possibility of describing a device that has such a skill. Goldman (1992) has also suggested that a great deal of cognitive science fits the *knowledge-rich* paradigm stressed by Stich and Nichols (1992). This, however, is not the sole paradigm in cognitive science. There is also a substantial tradition that posits *knowledge-poor* procedures, *i.e.*, processes or heuristics that are relatively simple and do not depend on quite so rich a set of rules or so complex a knowledge base. Instead of hypothesizing that naïve cognizers have rules or knowledge structures comparable in complexity and sophistication to probability calculus, these psychologists conjecture that cognizers have simple procedures for making probability judgments. Hobson (1991), on the other hand, argues that it is not appropriate to maintain that we are dealing with a theory level and mode of conceptual organization when this precludes so much of what is normally involved in 'theorizing'. Suppose a particular 'global theory' that drew upon non-theory-like sources of conceptual coherence and ontological distinctions. Suppose that such a theory entailed an individual in conceptual commitments that were more like those entailed by perception or knowledge than theory, reflecting non-theoretical modes of perceptual and/or conceptual understanding. Hobson wonders, would *this* kind of framework for 'specific theories' and/or for knowledge still count as a theory? For others, theory and knowledge are alike insofar as they provide underlying principles for conceptual organization (cf. Murphy and Medin 1985). Concepts are coherent to the extent that they fit people's background knowledge *or* naïve theories about the world. In sum, there is no consensus within cognitive science on what a theoretical attribution amounts to.

1.2.2.Relation between the individual and the elements of the theory

As we have seen, generally it seems as if certain capacities do require the internalization of a set of rules which enable us to perform the task at hand. The claim is that although we might not have explicit knowledge of the rules, we can still have the distinctive marks of a rule-following activity, as opposed to operating on mere regularity, or mere physiological processes, since some knowledge is located at a deeper level, one that is not available to introspection and therefore not available to consciousness. This kind of knowledge manifests itself in the actions we perform; for example, in those linguistic actions which are the uttering and understanding of a wide variety of sentences that are constructed in accordance with the rules of grammar. The cognitivist claim is that the *contents* of such rules might be causally implicated in the production of such understanding. This is why we can characterize it as tacit knowledge. But of what nature is that "tacit knowledge"? Any characterization of it involves two problems. First, tacit knowledge needs to be distinguished from the everyday kind of knowledge we have, some of which we do not consciously entertain. This requires making a principled distinction between representational states which are tacitly known and those representational states which seem to be available to the consciousness of agents who act on the contents of those states. States of tacit knowledge are states that have semantic contents and figure in causal explanations. Yet, are they like or unlike the more familiar psychological states which have the properties of being propositional-attitude states, that is, states like beliefs, desires, intentions hopes, wishes and the rest?[4] Secondly, we must distinguish these representational states from brute neurophysiological states and patterns. The advocate of tacit knowledge of representational states faces the problem that insofar as our cognitive relation to the contents of these states is in principle different from that relating us to ordinary beliefs, the question arises as to why we need to characterize these states as representational at all.

Concerning the first problem, for example, some researchers constrain psychological reality to intentional states, namely, propositional attitudes: beliefs, desires and the like. These are characterized by the following points (Davies 1987):

a) To be conscious or to be able to be conscious.

[4] Specifically, propositional attitudes are relations that persons have toward propositions. If a person believes that *p*, then her attitude is one of believing that *p*.

b) To be inferentially integrated with other beliefs (Stich 1978).

c) To conform to the Generality Constraint (Evans 1982): The fact that we can think that *a* is *F* and that *b* is *G* must imply that we can think that *a* is *G* and that *b* is *F*. Our thoughts are entrenched in the possession of the thoughts that use them. We are not punctate minds.

In this sense, some authors, such as Chomsky (1986), suggest that tacitly known states are like beliefs except that they are inaccessible to consciousness. A second differentiating feature, suggested by Stich (1978), is that these states are not inferentially integrated with beliefs, and they are what he calls *subdoxastic*. Many theorists have contended that these are the right level of constraint for psychological states. Indeed, there are some states that do not conform with the criteria of being beliefs and yet they are incorporated in full-blooded cognitive theories. In fact, psychologists propose models of certain cognitive abilities in a way that attributes states that will never be conscious or that need not be conceptual. If such states are explanatory, then they must be considered "real" and therefore psychologically real. Some authors, though, argue (Stone and Davies 1996) that folk theories need not appeal to such processes.

A subdoxastic state has some interesting properties. Beliefs can be put to use by being the basis upon which we draw inferences, thereby acquiring new beliefs. States of which we are only tacitly aware are not combined with beliefs to form new beliefs, nor do beliefs combine with subdoxastic states to form new subdoxastic states. This inferential isolation suggests that the subdoxastic states exist in special purpose separate sub-systems, a suggestion which is in consonance with Fodor's (1983) modular systems.

Davies (1987, 1989a) suggests a third differentiating feature of the states that we have only tacit knowledge with a notion of Evans (1982). With propositional states like beliefs we constrain the semantic contents of these states by describing them using only concepts available to the believer. The concepts used must be grasped by the person to whom the state is attributed; this is not true for states of tacit knowledge. With these states the content is not so constrained. A believer is said to grasp a concept only if that person can use the concept in a variety of situations. Concepts must have a general use-value. This generality constraint *necessarily* applies to the concepts employed in the attribution of propositional attitudes; it *may* be the case that in an information-processing sub-system we could have new informational states emerge on the basis of a recombination of the components of other states, but this ability to recombine is not essential to information states. In other words, it is possible for a pair of such states, say F(a) and G(b) to be followed by F(b). This does not count against the proposal that the necessary generalizability of concepts distinguishes propositional-attitude content from the content of other representational states. The idea is that if we came across a person who appeared to believe both F(a) and G(b), but could not grasp the proposition (Fb), then this would count as evidence against the attribution of the initial belief states (Fa) and (Gb). On the other hand, if one found a person to whom the information states (Fa) and (Gb) could be attributed, but found no reason to attribute the information state (Fb), then this would have no bearing on the initial assignment of information states.

There are other authors that have taken an opposing view on this issue. Searle (1996), for example, has formulated criteria of his own about the states that can be candidate to the attribution of knowledge. These states must have the following properties:

(i) To be causally responsible of behaviour
(ii) To have propositional content
(iii) To have a normative character
(iv) To be conscious or ably conscious
(v) To have semantic contents
(vi) To be voluntary
(vii) To be subject to interpretation
(viii) To be graspable in real time

In fact, Searle (1990) argues that there is an intimate connection between the mentality of a state and its availability to consciousness, a connection which makes the whole notion of an unconscious mental state problematic. The only way out is if that unconscious mental state is in principle accessible to consciousness. This is what is known as Searle's *connection principle*. Searle insists on the mental accessibility to consciousness: it is essential to the *aspectual* character of a mental state that it be so accessible. By this, Searle means that the content of a mental state must portray the subjective point of view of the person whose state it is, and this point of view is simply that person's consciousness of the world. Similarly, the fact that a person must grasp the concepts that are constituents of the contents of their mental states is explained by that person's (subjective) view of the world. We are constrained in our attributions of such contents to intentional states by the understanding the person has of the world around her.

Searle distinguishes four types of mental states that may be said to be inaccessible to consciousness, and thus 'unconscious'. The first is not quite a mental state. It involves mentality in a metaphorical way, such as when we attribute to a plant the desire to reach the sunlight. The second is an everyday notion, where we have beliefs that are not occurrent, not being consciously thought at a particular time, but which can be called into consciousness without effort. Thirdly, there are the deeper unconscious states postulated by psychoanalytic theory, which Searle prefers to call the repressed conscious states, since their inaccessibility to consciousness is dependent on their repression. Given the way repression operates, such states are always potentially accessible to consciousness: eradicate the repression and the content of the state will become manifest. It is the fourth type of unconscious state that Searle decries, those deeply unconscious states that are inaccessible in principle to consciousness. This inaccessibility deprives us of the means whereby we can determine the aspectual shape of the state. Searle argues that this sort of attribution is a way to illegitimately anthropomorphize the brain much as pre-Darwinian biologists attributed purposes to the states of organisms. What Darwin taught us, Searle notes, is that we can do without such purposes, so that what we need to do is to bring cognitive science into the post-Darwinian age by eliminating states and rules that are said to tacitly know. Such an elimination would leave us with the simple view of what it is that we are conscious of, plus a hardware explanation of what underlies our awareness. In visual-illusion cases -for

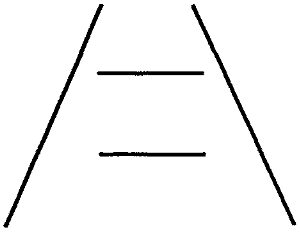

Figure 1.1. Ponzo illusion. The two parallel lines have the same length but the upper one looks longer.

example, the Ponzo illusion (see figure 1.1.)- it is said that our awareness of the top line as being larger and farther away is partially due to the rules of perspective which we unconsciously follow. Searle replaces this by 'We consciously see the top line as farther away'- and that is all there is to the intentionalist account. Any additional explanation cannot be intentional; it must be functional or neurophysiological, or both, and as far as functional explanations go. Searle is actually an instrumentalist about the last sort of mental state.

In sum, there are different and incompatible approaches about the issue of what the relation is between a person and the states that comprise her knowledge associated with some ability. Far from having been elucidated, it is possible that the discussion will maintain its vigour for a long time. My concern has been, nevertheless, to sketch which issues fall within the scope of my discussion, and to clarify what this dissertation is not about.

Chapter 2

Davies' account of psychological reality

Martin Davies (1986, 1987, 1989a, 1989b) is one of the few authors who has met the challenge of giving a sophisticated account of what I have called structural psychological realism. Specifically, he has proposed that a psychologically real theory is one that complies with the conditions for tacit knowledge:

> For a grammar to be a correct theory of I-language is for it to be psychologically real, or tacitly known. (1989b, p.545)

The notion of tacit knowledge he espouses is in line with that of Evans (1981). In essence, Davies' proposal is that someone has tacit knowledge of a theory when she has internalized it as a form of causal-explanatory structure. Davies undertakes the task of giving an account of psychological reality precisely to meet what we have called Quine's challenge, namely:

> There will always be extensionally equivalent theories. (...) Given that fact, does it make any empirical sense to suppose that an ordinary speaker tacitly knows, or cognizes, or has internalized, one set of axioms, rather than an alternative set from which just the same theorems of the relevant kind can be derived? Does it make any sense to suppose that one theory is psychologically real, rather than another extensionally equivalent theory? This is essentially Quine's challenge (1972) to the empirical credentials of the notion of tacit knowledge. Following a suggestion of Evans (1981), I would aim to respond to this challenge by construing tacit knowledge as a certain kind of causal explanatory structure. (1989b, p.542)

The first section of this chapter will be dedicated to the presentation of Davies' account; the second will try to clarify some possible confusions about the basic notions of the account; the third will develop some objections about the criterion proposed by Davies to sanction the reality of one theory over another; and in the final section I will present some objections, in the form of counterexamples, that question the scope of Davies' notion of psychological realism. On the one hand, my analysis yields a view that recognises Davies' account as a robust conception of psychological reality for cognitive theories, though in need of possible extension while, on the other, it presents serious objections concerning the specific criterion proposed to legitimize the psychological reality of a specific theory (what Davies labels the *Mirror Constraint*).

2.1.Davies first account

Davies' notion of psychological reality is based on the realist idea that cognitive theories are committed to describing a causal structure in the mind of the cognizer. As noted, Davies builds his proposal from intuitions of Evans (1981) about the knowledge a subject must posses, in the form of causal-explanatory structure, in order to be attributed semantic knowledge of whole sentences. Davies' proposal is, succinctly, that we can attribute tacit knowledge of a semantic theory if the structure of the speaker's competence *mirrors* the structure of the semantic theory. The notion of theory that Davies takes into consideration concerns semantic theories based on truth-conditions. Such theories can be described as a set of axioms from which we can derive theorems that state facts about what various sentences of a language mean. Davies extends this account to any sort of theory that concern knowledge attributions to cognizers. The basic idea of tacit knowledge would be applied then to the notion of *competence* proposed by Chomsky (e.g., 1976, pp.164-165), which argues that ordinary speakers tacitly know such axioms. In this sense, ordinary speakers know and cognize the facts stated by the theorems of the theory, and they also cognize -even if they do not know it in the ordinary sense- the facts stated by the axioms from which the theorems are derived.

Davies builds his proposal establishing the notion of theory, on the one hand, and a certain causal structure, on the other. Then the notion of theory is framed following a distinction between two abstract conceptions of semantic theories (say A1 and A2). Even though the presentation is focused on semantic theories, Davies intends the notion of tacit knowledge to apply to any cognitive theory. The distinctions can be stated as follows:

> (A1) One notion of an abstract structure in a language is that of a systematization (...) of the facts about the meanings of sentences in the language. (1986, p.131)

> (A2) There is however a different, more interesting though still abstract notion of semantic structure. This is the notion of a structure that an ideally rational creature would recognize. It is an abstract, rather than a psychological, construal because there is no assumption that the structure is realized in the actual speakers of the language. (Ibid., p.132)

This second notion is the one that Davies uses in his account of psychological realism, which he labels as *a priori semantic theory* (1986, p.135). Specifically, Davies considers, after Evans, that salient structural facts about a semantic theory are, for speakers who understand sentences s_1, s_2,...,s_n and are able without further training to understand sentence s, of the following form:

> (i) The resources used in derivations of the meaning specifications for s_1, s_2,...,s_n are jointly sufficient for the derivation of the meaning specification for s.

Since there may be more than one theory with the requisite structure, Davies introduces the notion of *derivational structure* which will enable him to define an equivalence relation for the class of theories extensionally equivalent. In principle, for Davies two theories are equivalent in relation to their derivational structure if just the same salient structural facts hold of both, that is, if their derivational structures match the same competence structures in speakers (1987 p.446). However, Davies himself recognises that the notion of derivational structure is slippery and needs a more precise specification. To begin with, the notion requires a precise position regarding what should count as a theory if we want a robust criterion about what to count as a derivational equivalent. On the one hand, Davies relies on a very specific consideration, namely:

> Officially, after all, a theory is just a set of sentences closed under deduction. Two sets of sentences -not necessarily closed under deduction- are logically equivalent if they deductively generate the same theory. And there may be many different routes for deriving a particular theorem from a given set of sentences. (1987, p.449)

However, as he further comments, a semantic theory is not only a theory in this official sense. A semantic theory involves a set of proper axioms, from which, given some background logic, certain 'target' sentences- the meaning specifications for whole sentences of L- are derivable as theorems. But, since those target sentences may be derivable from the axioms in many different ways, Davies assumes that it is possible to distinguish a 'canonical' proof procedure for deriving the target sentences from the axioms. In this sense, the salient structural facts about semantic theories should really be cast in terms of *canonical* derivations. And the definition of derivational equivalence should be adjusted accordingly.

Specifying a certain canonical derivation poses problems, however, since in general the attribution of tacit knowledge is fairly coarse grained. Sometimes attributions will be made when canonical constraints are not met, and some other times will be difficult to establish them. Then, to a certain extent, we will have to rely on an intuitive notion of structure. In any case, we can go along with Davies' intuitions on this matter, only recognising that there seems to be a tension between such intuitions and thorough specification.[5]

[5] There is a related issue about the attribution of a theory which has to do with the *existential commitments* of the theory. In determining the existential commitments of a theory, we must distinguish the *theoretical magnitudes* to which the theory *is* existentially committed from the representational constructs to which the theory is *not* existentially committed and which serve only to specify the theoretical magnitudes (Matthews 1991). This addresses the question whether the commitments of the theory are all assumed by the instantiation in the cognitive system or not.. Harman (1980) describes the issue in the following way:

> (...) given any theory we take to be true, we can always ask what aspects of the theory correspond to reality and what aspects are mere artifacts of our notation. (p.21)

Harman acknowledges that one linguistic theory may be a "notational variant" of another. In this sense, aspects of a true theory not shared by its notational variants should not taken, according to him, to have psychological reality, leaving unresolved the issue of what aspects of the theory correspond to reality and what aspects are artifacts of notation. Chomsky replies:

On the other hand, we have the notion of causal-explanatory structure. Davies specifies the psychological notion of semantic structure in the following form:

> (P1) To conceive of semantic structure as psychological, rather than abstract, is to conceive it as the causal-explanatory structure of the semantic ability of actual speakers, [i.e.,] normal or optimally functioning actual speakers. (1986, p.132-133)

> (P2) [W]e can consider [the causal-explanatory structure of the semantic ability of] an individual actual speaker, who may be abnormal and less optimally functioning. (1986, p.133)

The rationale behind the distinction between these two psychological notions concerns the fact that the first notion refers to the cognitive structure of a normal (ideal) speaker, and the second to that of a particular speaker, in which the structure could not have been internalized optimally and therefore could be different from the attribution of the ideal psychological structure. P1 would be, in that sense, the causal-explanatory structure to be used in the criterion of psychological reality. Specifically, the salient structural facts about the competence of speakers are explained, for speakers who understand sentences $s_1, s_2,...,s_n$ and are able without further training to understand sentence *s*, as follows:

> (ii) The operative states implicated in the causal explanation of speaker's beliefs about the meaning of sentences $s_1, s_2,...,s_n$ are jointly sufficient for a causal explanation of his belief about the meaning of sentence *s*.

Davies backs Evans in pointing out that the constraint on semantic theories is just that (i) and (ii) should match. Where there is, in the theory a common factor -for example, a common axiom- used in the derivations of several theorems, there should be, in the speaker, a causal common factor implicated in the explanations of the several corresponding pieces of knowledge about whole sentences. The way then to meet Quine's challenge, that is, to

[Harman] correctly points out an error in my formulation: there is a question of physical (or psychological) reality apart from truth in a certain domain. (Chomsky 1980b, p.45-6) distinguish which one of two extensionally equivalent theories holds for a speaker, is to use the notions of derivational structure and causal-explanatory structure to give an account of tacit knowledge and of psychological reality. Davies proposes then that we have to specify on the one hand the:

Derivational Structure: The structural configuration of the theory to be attributed,

and on the other:

Causal Structure: The operative states implicated in the causal explanation of the knowledge of the theory.

Then, we are entitled to attribute tacit knowledge and psychologically reality[6] to a theory if the following condition obtains:

Mirror Constraint: The derivational structure of an abstract theory matches the causal structure in the cognizer's mind.

In order to make the idea clear, Davies asks us to imagine that we have three semantic theories T_1, T_2 and T_3 about a finite language *L* of one hundred sentences. Each sentence is made up of a subject (a name) and a predicate. The names are *'a', 'b, ..., 'j'*. The predicates are *'F', 'G',..., 'O'*, The sentences are *'Fa', 'Fb', ..., 'Fj', 'Ga', Gb', ...'Oj'*. The sentences have meanings which depend in a systematic way upon their construction, and this can be revealed from the outside. All sentences containing *'a'* mean something about John; all sentences containing *'b'* mean something about Harry; all sentences containing *'G'* mean something about being happy; all sentences containing *'F'* mean something about being bald. All three theories are theories of truth conditions for the language *L*, that is they assign the same truth conditions to the sentences of *L*, though they differ in their internal or what

[6] Henceforth I use tacit knowledge and psychological realism interchangeably, since Davies considers that the conditions of applying psychological reality correspond to the conditions of applying tacit knowledge.

Davies calls *derivational structure*. The theories deliver whole sentences of the form:

> *s* means that *p*

Even though we shall consider theories which employ the familiar format:

> *s* is true iff *p*

The first theory, T_1, is a *listiform* theory. It has one hundred axioms, one specifying the meaning of each L-sentence. The second theory, T_2, is a *structured* theory. It has twenty-one axioms: one for each name, one for each predicate, and one for the subject-predicate mode of combination. For the name *'a'* we have:

> *'a' denotes* John

For the predicate *'F'*, we have:

> An object *satisfies 'F'* iff it is bald

And for the mode of combination, we have the compositional axiom:

> A sentence coupling a name with a predicate is true iff the object denoted by the name satisfies the predicate

T_3, on the other hand, is also a *structured* theory. It has twenty axioms, instead of twenty-one. T_3 contains an axiom for each name of L, which are the same axioms as in T_2. However, the axioms for the predicates are different. For predicate *'F'* we have:

> A sentence coupling a name with the predicate *'F'* is true iff the object denoted by the name is bald.

In other words, we have distributed the content of T_2's compositional axiom among the ten predicates in T_3. Given that Evans' proposal of tacit knowledge should count T_2 and T_3 as equivalent theories, since the same structural facts hold of both, then they are *ipso facto derivationally equivalent*. In other words, there is no distinction to be made between tacit knowledge of T_3 and tacit knowledge of T_2.

Suppose then that there is a speaker who uses the sentences of *L*, with the truth conditions that the semantic theories coincide in assigning. What would it be for that speaker to have tacit knowledge of T_2 rather than merely of T_1? The answer for Davies is that the speaker should have a causal-explanatory structure which mirrors the derivational structure of the theory, that is, the speaker should have a causal-explanatory structure for which there would be a single state of the subject for each axiom of the theory, and those states should figure in a causal explanation of why the speaker reacts in a regular way to all the sentences in which the axioms derive their meanings.

This account has the desirable property for Davies that there can be empirical evidence for or against a particular kind of causal structure in a given subject. If attributions of tacit knowledge are attributions of structure of causal-explanatory states, then such attributions make perfect empirical sense; and they can, in principle, be grounded in empirical evidence. This, according to Davies, meets Quine's challenge. What sort of evidence can we imagine that would allow us to attribute to a speaker tacit knowledge of the articulated theory T_2, rather than merely attributing the listiform theory T_1? Davies presents and supports Evans proposal who considers three types of evidence that would be of relevance to his notion of psychological reality. The first is evidence from patterns of acquisition, namely, the developmental facts about the learning of, for example, a language. Secondly, we could consider evidence from patterns of breakdown or decay, i.e. all the data that clinical neuropsychology can provide concerning the cognitive deficits observed in patients with brain lesions. Finally, we could use evidence from experiments on perceptual processing, that is, evidence taken from the psychology of, for example, object, verbal or haptic perception, to which we could add another sort of evidence that we could use as a part of a constitutive notion, namely, evidence from revision of beliefs about meanings. This will be developed below.

2.2.Problems of the first account

The account just presented has two basic problems that even Davies has recognised. One of them will lead to a revision of his account that will make his account part company, as Davies puts it, with the account of Evans. The second will require a refinement of his Mirror Constraint.

2.2.1.Lack of informational sensitivity

Davies recognises that "there is no more determinate informational description which is definitely licensed by the [first account] and which has the property that the content of an explained belief follows from the information contents of the explaining states" (1987, p.457). The first account lacks informational sensitivity, since it does not specify the contents of each operational state.[7] Specifically, the problem lies at the first account's inability to license descriptions of constituent explanatory states as states of tacit knowledge of the semantic axioms. In other words, the first account does not require that the information drawn upon by each operational state be established. In fact it provides a global attribution of contents and not an attribution specifying the contents of each state of tacit knowledge. Obviously, this is a "side effect" of trying to cover all structurally equivalent theories. Put in another way, the first account counts as equivalent any two theories that have equivalent structural configurations, even if they are not logically equivalent.[8] As we have seen above, Davies tentatively argues, following Evans, that we should count T_2 and T_3 as derivationally equivalent.

[7] Davies stands in direct opposition to the notion of psychological reality advocated by Peacocke (1986). Peacocke's notion will be presented in the next chapter. Succinctly he proposes that a theory is psychologically real if the cognitive mechanism said to instantiate the theory draws upon the information that the theory establishes. Davies himself agrees that Peacocke's notion cuts thinner than his own, which means that Davies' proposal would have to face difficult prospects if it is to stand up to comparison.

[8] Davies makes it clear that his notion does not require the states to be representational, viz. complying with the Language of Thought hypothesis: "Tacit knowledge will not essentially be a matter of explicit representation (1987, p.454), "tacit knowledge does not have to be explicitly represented; it can be realized by the presence of a processor" (1989b, p.553)

2.2.2. Use of an undifferentiated notion of causal structure

The second problem Davies recognises has to do with a criticism made by Crispin Wright (1986a, pp.204-38 and 1986b, pp.31-44). The objection could be put in the following manner. Suppose that we reveal the competence of speaker A is underpinned by T_1. Suppose we reveal some sort of causal structure underlying A's competence that is not a semantical structure but some other type of structure, (say, nutritional) apart of course from the semantical one. This special structure has the following features. For each name of L, there is a special nutritional input. Accordingly, for each disruption of inputs there will be a disruption in all meaning assignments corresponding to a particular name. Therefore, and so long as we consider *patterns of breakdown* to be genuine evidence for tacit knowledge of a theory, then for any disruption of inputs, we will have evidence favouring T_2 or T_3 instead of T_1. In short, it would seem as if mere causal structure does not warrant a certain semantic theory.

2.3.The revised account[9]

Davies finds a way to surmount these problems by introducing the notion of *systematic revision of beliefs*. Specifically, if a speaker revises his belief about the meaning of say, '*Fa*' -taking to mean, not that John is bald, but merely that he is bald*ish*- then he *likewise* takes '*Fb*' to mean that Harry is also bald*ish*. In other words, the speaker keeps meaning assignments in step with each other to preserve systematicity under revision. Then, the notion is refined in the following terms:

> (i) The resources used in derivations of the meaning specifications for s_1, s_2,...,s_n are jointly sufficient for the derivation of the meaning specification for s (for speakers who understand sentences s_1, s_2,...,s_n and are able without further training to understand sentence s).

[9] In order to make the points clearer I have avoided the presentation of an intermediate account which has no relevant interest in our discussion.

And the salient structural facts about the competence of speakers are explained as follows:

> (iii) The operative states implicated in the explanation of the speaker's actual beliefs about s_1, s_2,...,s_n together with revision to the belief about the meaning of *s* should provide an explanation of the corresponding revision in the speaker's beliefs about s_1, s_2,...,s_n.

Therefore, we can talk of tacit knowledge of a theory if:

> (iv) Within the causal explanatory structure in the speaker there is an explanatory locus of systematic revision corresponding to each proper axiom or rule of the theory. For each axiom or rule, the required notion of *systematic revision* can be spelled out in a quite determinate way.

However, this new version implies giving up of one of the aims of Evan's proposal, namely, to group together such theories of which salient structural facts hold of them. Indeed, in the revised account, T_2 and T_3 will count as different attributions of tacit knowledge so long as they have *different locus of possible systematic revision.* As we said above, Davies acknowledges the point but he claims not to be worried by it.

2.4.Analysis of Davies' basic notions

In what follows I will argue that Davies' account requires some elucidation, since some of its contentions may have different interpretations depending on the model we use to consider them. By revealing such different interpretations ofDavies' notions we will actually be led to the basic problems of the account. I will first examine the distinction Davies draws between two abstract conceptions of theories (A1 and A2), on the one hand, and two psychological conceptions of causal structure (P1 and P2), on the other. I shall be concerned with the contentious qualification of abstract, for the first two notions, and psychological, for the second ones, especially with its application to the enterprise of explaining a cognitive ability. This will be connected by comparing Davies' notions with those entertained by other theorists and cognitive scientists, such as Chomsky, to which Davies wishes his notion of tacit knowledge to apply. Specifically, the idea is how to place what Chomsky labels a competence theory in Davies' framework and then to see whether it corresponds to the notion of theory that Davies uses in his account of tacit knowledge. In my opinion there is an unsupported assumption made by Davies in applying his characterizations to, among others, Chomsky's notion of competence theories. The search for a solution will lead us to arbitrating the dispute within Grandpa's explanatory framework (see Chapter 1). Specifically, we will try to answer the following questions: Does the task level correspond to an A2 notion of theoretical structure? And what is the level in Grandpa's framework that corresponds to Davies' causal-explanatory structure?

2.4.1.Is the distinction between abstract and psychological empirically sound?

As we have seen above Davies outlined, within his Mirror Constraint, a theory, on the one hand, and a certain causal structure on the other. This account is framed after a distinction between two abstract conceptions of semantic theories (A1 and A2) and two psychological ones (P1 and P2). I recall the formulation:

(A1) The systematization of the facts about the meanings of sentences in a language.

(A2) A structure that an ideally rational creature would recognize. It is an abstract, rather than a psychological, construal because there is no assumption that the structure is realized in the actual speakers of the language.

(P1) The causal-explanatory structure of the semantic ability of actual speakers, i.e., normal or optimally functioning actual speakers.

(P2) The causal-explanatory structure of the semantic ability of an individual actual speaker, who may be abnormal and less optimally functioning.

The way in which Davies uses the distinction in his account of psychological reality is that the appropriate constraint on a semantic theory would require that the causal-explanatory structure found in normal speakers -P1- should mirror the derivational structure of the theory -A2-. The problems I see here lie in the qualification of the notion A2, which claims that it is "the structure that an ideally rational creature would recognize (...) [without assuming] *that the structure is realized in the actual speakers of the language*" (my italics). This refers to semantical theories, but as we argued above, it should be extended to other sort of cognitive theories. I think that there is a tension between the notion of an a priori theory and the theoretical structure that Davies wants to use in his criterion for psychological reality. As a matter of fact, it is at least contentious that the sort of theoretical structure to be considered a candidate in an attribution of tacit knowledge should be not only an a priori theory, but simply a theory that does not assume its realizability in a cognizer. Note the implication. We are considering that we can devise a theory without the cognizer, ideal or real. Davies seems to entertain this qualification simply because the possibility *exists*, and that the variety of a priori theoretical structures could be "recognized" by cognizers need not coincide with the actual structure accounting for the competence:

> The crucial question this time is not whether an ideally rational speaker/theorist could work out the meaning of s given the meanings of $s_1, ..., s_n$. It is, rather, whether the cognitive resources which in fact underlie the ability of a normal speaker to recognize the meaning of s are among those which underlie her ability to recognize the meaning of $s_1, ..., s_n$ It is not guaranteed *a priori* that the answers to these two questions will be the same. (1986, p.133)

It seems to me, however, that even if the possibility exists, the relevant question is if that is a general case in the domain of cognitive theorising. I believe that the basic problem here is that Davies has picked one of the few domains in which we can design theories of knowledge without the cognizer. And that is even contentious. Indeed, for a widely held hypothesis about cognition and knowledge is that we cannot even speak of knowledge or competence of a semantic theory, let alone of a cognitive one, as being detached from actual speakers or cognizers, that is, being an a priori theory. On this view, the only sense in which a semantic theory of a natural language can be considered a theory specifically about the structure that an ideally rational creature *can* recognize has to do with the very fact that a semantic theory is the structure actually realized in actual speakers. If not, how could they assign meanings to sentences? Of course, it is another thing that the speaker could recognize many other structures in, for example, artificial languages. In general, it only makes sense to speak of a theory of object perception, musical proficiency and the like precisely because these are theories about a given competence which is not independent of the causal-explanatory structure. For such a position there cannot be a priori theories of semantic, syntactic, object perception, since they are theories about cognitive abilities. Note that we only need one case in which a knowledge theory needs the cognizer to be designated to undermine Davies' characterization of "abstract" theories as one of the elements in the Mirror Constraint.

The position that requires the cognizer is sustained by some theorists, such as Chomsky, who curiously seems to be source of the sort of theories that Davies refers to. As a matter of fact, Chomsky has made it clear from the very beginning that he considers linguistics to be a psychological theory, "that branch of human psychology known as linguistics" (1972, p.88), as he puts it, denying the separation of linguistics as the abstract theory and psychology as the mechanism. Even in the case of semantics, Chomsky presses his case, as we can see in his controversy with Dummett, who claimed that his theory of meaning is not a psychological hypothesis because "it is not concerned to describe any inner psychological mechanisms" (1976, p.70). In response, Chomsky argues that "Dummett's theory of meaning is a 'psychological hypothesis", though one that abstracts away from many questions that can be raised about inner mechanisms:

> Dummett's theory of meaning is concerned with empirical facts and attributes implicit knowledge to the speaker, and "implicit grasp of certain general principles, naturally represented as axioms of the theory [of meaning], [which] has issued in a capacity to recognize, for each sentence in a large, perhaps infinite range, whether or not it is well-formed, a capacity naturally represented as the tacit derivation of certain theorems of the theory... [to each of which]...corresponds a specific practical ability," such as the ability to recognize well-formedness, "to use the language to say things," and so on.
>
> In short, Dummett's theory of meaning will incorporate statements about the speaker's capacities, practical abilities, implicit knowledge, and the like, which are taken to be true statements. But he is not proposing a psychological hypothesis concerning "inner psychological mechanisms".
>
> One might ask at this point, once again, what is the distinction between a theory held to be true of the speaker's capacities and implicit knowledge, on the one hand, and "psychological hypothesis," on

> the other. (1980a, p.110)

For Chomsky, Dummett's theory is a psychological theory because it specifies conditions that the inner psychological mechanisms must meet. For example:

> One can speak of "reference" or "co-reference" with some intelligibility if one postulates a domain of mental objects associated with formal entities of language by a relation with many of the properties of reference, but all of this is internal to the theory of mental representations; it is a form of syntax. (1986, p. 45)

Concerning his theory of Universal grammar, and acknowledging Grandpa's explanatory framework in levels, Chomsky advances the following:

> We may consider the study of grammar and UG to be at the level of the theory of the computation. I don't see any useful distinction between "linguistics" and "psychology", unless we choose to use the former term for the study of the theory of the computation in language, and the latter for the theory of the algorithm. (1980b, p. 48-9)

And he also argues:

> There is no initial plausibility to the idea that apart from the truths of grammar concerning the I-language and the truths of UG concerning S_0 there is an additional domain of fact about P-language, independent of any psychological states of individuals. (1986, p.33)

In other words, a grammar is an abstract theory so long as it *abstracts away* from particular psychological mechanisms but it is specifically a psychological hypothesis. In this sense, grammars are psychological hypotheses by specifying -intensionally- the function that these inner mechanisms are alleged to compute.[10] My view is that the position that Chomsky

[10] However, and this is important even for the semantics of that language, Chomsky remarks that the intensional specification of a language might be misleading since the language is infinite. Grammars provide a characterization of a speaker/hearer by specifying the function (pairing of sound and meaning) that they compute in the course of language use (Matthews 1991, p.19). So long as the language is infinite, it won't be possible to establish the function extensionally, since the set that defines the function is infinite.

represents here is not only restricted to him; rather, I believe that it is the view generally held by cognitive scientists. However, an appeal to Chomsky's authority obviously is not an argument, but it represents an alternative to the view presented by Davies.

The problem with the notion of A2 is, as we argued above, that even if it applies to a notion of semantic theory, that is, even if we can in fact construe a theory without taking into account the cognitive mechanisms of a cognizer, we cannot extend it to other sorts of theories in other cognitive domains. As a matter of fact, it is difficult to imagine how we can construe theories of perception, concepts, learning, or even reasoning *without assuming that the theories are about cognitive mechanisms in the mind of the cognizer.* What would an "abstract" theory of visual perception look like?

It would be therefore more reasonable is to modify Davies notion of A2 so that it could be better adapted to the task of looking for psychological reality. In that sense, A2 should be described in the following way:

> (A2') A structure that an ideally rational creature would recognize, assuming that the structure can be realized in the actual speakers of the language. It is an abstract, rather than a psychological theory, because it abstracts away from particular mechanisms of the speaker.

The problem would be then to accept two positions, one represented by Davies' defense of a qualification such as A2 and another defending the one specified by A2'. However, in my opinion the opposition is only apparent, since Davies' argumentation in his writings seem in fact to be defending an A2' position all along.

2.4.2.Is Chomsky's competence theory Davies' notion of theoretical structure?

As a matter of fact, when we pose the question in this way, that is, distinguishing two positions regarding cognitive theories, namely, between A2 and A2', then we face the problem of explaining how it is that Davies seems to defend Chomsky's position concerning the Chomskyan account of psychological reality, i.e.:

> In his influential book *Spreading the Word*, Simon Blackburn has distinguished a number of different positions one might take on the relationship between ordinary language users and syntactic and semantic rules. One, according to which 'psychological reality' is allowed to the rules is of course *Chomsky's realism*. (...) The position that I am defending -what I take to be Evans' position- is intended to be some version of Chomsky's realism. (1986, p.137)

He also acknowledges to be elucidating the notion of psychological reality that Chomsky entertains:

> In the previous section I assumed that the idea of a causal-explanatory structure is an empirically respectable one, and then drew some distinctions with a view to legitimizing the Chomskyan notion of tacit knowledge. (1986, p.135)

In sum, Davies seems to defend the Chomskyan notion of *competence*, as the theory-level notion, recognising that the grammar *is a competence theory*, which is the theory level for Chomsky. But then he places his view about cognitive theories within the A2 position, which he labels "*a priori semantic structure*". The problem we face then is how to reconcile this characterization with the above distinctions between A1, A2, A2' and P1?

I guess that there are three ways out of such a situation. One is to assume that Davies is arguing that, Chomskyan assertions about the psychological character of semantic theories notwithstanding, Chomsky's competence model is in fact an A2 theory and not an A2' one. This solution has the advantage of not requiring any special argumentation, but it is of course a speculative consideration. Another option is to contend that the condition about the realization of the theoretical structure in actual speakers is not at all essential to individuate A2, and hence that A2' is pointless. This notion could be then be described as:

> (A2") A structure that an ideally rational creature would recognize. It is abstract, rather than psychological, because it abstracts away from particular mechanisms of the cognizer.

This requires then that the individuation of the abstract theory be made fully dependent on the notion of "structure recognition" by an ideally rational cognizer without any commitment to the realization of the structure. Useful as it may be, this position only sweeps the problem under the rug, robbing Davies' distinction between A1, A2 and P1 of any force it could have.

A third possibility is that of modifying Davies account to the effect that A2' *should* be the notion to be confronted in the Mirror Constraint. On this view, what Davies reckons as abstract theory, is nevertheless a *cognitive theory*, a theory made by a theorist about a cognitive competence actually realized in cognizers, and therefore a full-blooded psychological theory. In the particular case of language, it would be a grammar that would aim at explaining the cognitive competence of a speaker/hearer.

Summing up, there are some difficulties to place Davies' proposals within the context of current cogntive-science theorising. To begin with, Davies' A2 notion of abstract theory, which lacks any commitment to actual realizability in actual cognizers, clashes with the widely held hypothesis that any cognitive theory is committed to such realizability. If we accept Davies' A2 position, the tension appears in believing, first, that he is providing an account for cognitive theorizing in actual programmes such as Chomsky's, since Chomsky himself has made it explicit that a competence theory should be committed to its realizability in actual speakers and, second, the very mission of sanctioning the attribution of a theory to a cognizer seems to imply that the theory must be assumed to be realized in a cognizer. The tension can only be eased by overlooking these objections, by adopting an A2' notion (making realizability a condition for an abstract theory) or by adhering to an A2" notion (dispensing of the need to mention realizability). Grandpa's framework can help us in deciding which of the three is more appropriate.

2.4.3.Davies' account under Grandpa's explanatory framework

As we have said above, in the case of psychology, or cognitive science, many authors have argued that an adequate psychology will comprise explanations at different levels. More precisely, explanations in cognitive science should accord with what we have called in chapter 1 the "classical cascade" (Franks 1995), an explanation that *cuts across* the different levels that are normally proposed, namely, the *task, algorithmic* and *implementation* levels. The *task level* is the top level. It identifies what sort of *function* the system is performing. It provides an abstract formulation of the information-processing task that defines a given psychological ability. As we saw in the last chapter some see it as providing only a specification of a function in extension (Peacocke 1986), that is, giving the semantics of the function (Franks 1995). Others take the task level to specify the information-processing task -calling it the 'function-composition view'-, in the sense that it adds to the functional-specification details of the informational contents operated upon by the theory. The function-in-extension view requires the introduction of an intermediate level between the task level and the algorithmic level, which is called, according to Peacocke (1986), level 1.5, the level in which the information that a system draws upon is specified. In the last chapter, we concluded that as far as the latter position required another intermediate level to account for the information -level 1.5 in Peacocke's version- the advantages of parsimony recommended the function-composition position. The *algorithmic* level is the *how* level. It establishes the form of representation and the informational transformations that a mechanism exerts when a task is performed. It takes into account the computational constraints in specifying the psychological processes, or computations, by means of which the task is actually performed, which may differ in different creatures. And finally we have the *implementation level* where the way in which computation and intention are instantiated in a physical system is specified.

It is important to make clear, at least here, that Grandpa's framework of levels is actually a heuristic proposal of how to build theories in, above all, cognitive science so that by employing such an approach the empirical projects could be much more productive. Hence, such a proposal could be seen, in the best possible scenario, as a constitutivity criterion for cognitive theories: What should a cognitive theory have or how should we devise a cognitive theory in order for it to be considered an adequate cognitive theory. What such an approach cannot be said to be is an account of reality for a cognitive theory. Therefore, even if we could approach the notions in both accounts it should be always kept in mind the different projects underlying the proposals.

2.4.3.1.Is the task-level theory Davies' abstract theory? The task-level described in the last chapter seems, prima facie, to correspond quite neatly to the abstract theory proposed by Davies. The task level establishes the function that the cognitive system computes, and Davies' abstract theory refers to the "abstract structure" recognized by a cognizer. However, as we have seen above, there are two different views on what exactly we should count as a task theory, one considering that the function should be specified in information-processing terms, and the other reckoning that the level should only specify the function-in-extension. In Davies' account, the theory level cannot nevertheless be the function-in-extension position, since for Davies the theory to be considered is one which has a relevant structure about which we can stipulate an equivalence by virtue of sharing with other theories the same salient structural facts, the same competence structures in speakers. Therefore the function-in-extension comes short of specifying the relevant structure, provided that the function-in-extension cannot be considered to establish any structure whatsoever; it is only a list of cases or a mapping between inputs and outputs (see Peacocke 1986).

Therefore, we have to consider that the correspondence must be established under the function-composition view. The only difference is that Grandpa's model requires that the theory has to be realized in actual cognizers, which is not what the notion of abstract theory -i.e., A2- requires. In such a situation we are forced to consider the possibility that either there is no correspondence between both accounts, and that A2 is actually what we have to compare with the causal explanatory structure, or that we have to choose between A2' or A2" to be able to establish the correspondence.

My opinion is that Davies should be defending A2' but the fact that he espouses A2 is due to two reasons. The first is that his proposal intends to meet Quine's challenge, a challenge based on the idea that there can be semantic theories, or other cognitive theories, devised independently of cognizer's examination. For example, the idea that we can establish the truth conditions of semantic theories without taking into account certain cognitive constraints, such as for example relevance, rational expectations, implicatures and the like. This view, as Davies himself recognises on occasion, is already mythical, very few people reckon nowadays that meaning can be established without cognition. The second reason, which is more interesting for me, is that he needs A2 to be able to apply his Mirror

Constraint, since without it then there is a conflation between the notions of A2' and P1.

2.4.3.2.What is Grandpa's corresponding level for Davies' causal-explanatory structure? Could it be possible to establish the correspondence between the causal-explanatory structure and the algorithmic level? For one thing, the algorithmic level takes into account the computational constraints in specifying the psychological processes, or computations, by means of which the task is actually performed, and this could be well considered to be the relevant psychological structure of the cognizer. However, Davies proposes the correspondence at the level 1.5 of Peacocke, which is the function-composition level:

> In fact, the simple equation of the processing level with the level of tacit knowledge description is potentially misleading. A description at the tacit knowledge level specifies the information that the system draws upon. But a full description at the processing level should surely do more that, it should specify, in addition, how the information is drawn upon. To the extent that the processing level is to be identified with Marr's level two -the level of the algorithm- the tacit knowledge level should be distinguished as a slightly higher level of description. (Peacocke [1986] labels it level 1.5). (Davies 1989b, p.547)

This incurs a paradox, since it implies that both notions, A2 and P1, correspond to the *same notion* in Grandpa's framework, that is, the causal-explanatory structure *is characterized in the same way* as the one about the theory level.

However, there is a way to understand Davies' contentions, which will reveal the nature of the problem that faces Davies in the application of his criteria for sanctioning the reality of a specific theory. The idea is that when Davies is talking about the theory level, he is referring to the *description* of the theory, the wording that specifies it, whereas when he refers to the causal-explanatory level he intends to refer to some sort of mental structure. Davies puts it this way:

> In *Knowledge of Language* [Chomsky] recommends distinguishing between internalized language -that is, I-language- and grammar. I-language is 'some element of the mind of the person who knows the language' (p.22); a grammar, in contrast, is a theory of I-language. A grammar is not a cognitive structure; it is a linguist's theory (...) If a particular grammar is a correct theory of the I-language of a speaker then the language faculty of that speaker can be characterized -at one level of description- by that

> grammar. (...) For a grammar to be a correct theory of I-language is for it to be psychologically real, or tacitly known. (1989b, p.545)

In other words, the distinction that Davies' draws is between a mental element -whatever that is- and its *mode of characterization*, a theory. It is only acknowledging that point that we can understand the full content of the proposal as well as the reason for the disagreement with Grandpa's framework.

2.5.Too weak a criterion

As we have seen all along in this chapter, Davies' account makes the following contention:

> (v) 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.

Let us accept that such an account is *metaphysically* sound. However, problems surface once we come down to transforming such a conception into a criterion for sanctioning the reality of a given theory, namely:

> (vi) A theory *T* is real if it complies with the Mirror Constraint, namely, if the derivational structure of the theory (A2 in Davies' taxonomy) matches the causal structure in the cognizer's mind (P1).

Any comparison that is supposed to entertain these two elements should comply with some requirements. To begin with each one of them must *be about different things*, and there must be independent features of the elements in question that enable them to be identified in a different way depending on which side of the comparison they belong. Among other properties, the independence of both elements should allow for the disclosure of each of them without the disclosure of the other. Then, if that is so, the structures that Davies proposes to use in the Mirror Constraint should therefore be independent structures, that is, entities that could be revealed independently of each other. But, as we have seen, this does not hold true. One is the mode of characterization of the other. If we take the above citation at face value, then the difference between A2 and P1 vanishes. Indeed, since Davies assumes Chomsky's idea that the contrast between a grammar and I-language is that the former is the linguist's theory of the latter. Furthermore, as I have argued throughout this chapter, the idea can be extended so that the difference between a cognitive theory about a given cognitive ability and the mental element that accounts for that competence is that the former is a theory of the latter. What we have then is a competence on one side of the comparison and a theory of the competence on the other. So long as one element of the comparison is the mode characterization of the other, we are thus comparing a *thing* with its description. Accordingly, the elements of the comparison cannot be said to be about different things; both A2 and P1 actually refer to the very same structure. Therefore, this comparison cannot be used to establish the reality of a theory against another one. This implies that in the case that we could compare two theories about the very same competence, what we would be entitled to say is that one is an *incorrect* theory about that competence, rather than lacking psychologically reality. And this is so because the *incorrect* theory would not correctly describe the structure of what is it about, i.e., the competence, rather than describing some other sort of structure, for instance, an abstract one in the sense of A1.

Moreover, my point is that regardless of all this, the root of the problem with the Mirror Constraint is that it is not possible to compare a theory with its competence. Davies' notion cannot be developed into an empirical criterion to sanction the reality of one particular theory over another because it is *epistemologically* impossible to separate the theoretical from the mental structure. Note that I am not implying here that there is no way to evaluate a theory, that is, that we cannot have evidence in support of one particular theory. What I am saying is that it makes no sense of talking about a competence as a putative element that we could entertain in a comparison. We cannot compare a "mental element" with a linguistic expression, because the "mental element" needs to be interpreted in order to be compared, that is, it requires a mode of characterization. Specifically, the idea is that grammar understood as it is in the above citation is the (unique) way to access the causal structure: the theory is the *only* way to know the mental element. As a matter of fact, we won't ever be able to individuate and know the causal structure of the Mirror Constraint other than by interpreting it in terms of a theory that describes it. What this amounts to is that we cannot distinguish the causal structure disclosure from that of constructing the theory, because *there is no fact of the matter that can distinguish the enterprise of revealing the structure of the theory from the causal-explanatory structure*, the theory being the mode of characterization of the causal structure. It could be said that empirical evidence is actually how we uncover the causal-explanatory structure. True, but evidence is not the causal-explanatory structure, and the way it discloses a certain causal structure is by sustaining or dismissing a given theory. If we agree, as we have concluded above, that a theory is supposed to characterize the structure in a cognizer's mental structure, then the task of revealing the cognizer's causal structure is to be characterized in the theory. They cannot be taken to constitute two separate enterprises. The facts that we obtain by examining a cognizer are facts framed in theoretical terms. So if we come up with a piece of evidence like 'there is a locus of systematic revision for word C', this can make sense within the context of a theory, by taking into consideration that the competence of grammar requires that there should be a locus of systematic revision. There is no evidence without theory. Theory construction and causal disclosure go hand in hand in cognitive science. The fact is that when Davies characterizes the notion of causal-explanatory structure:

> (..) the basic idea [is] of a psychological structure that allows language users to recognize the meaning of new sentences built from familiar constituents; of common cognitive resources at work in the comprehension of different sentences with common constituents (1986, p.135),

he is in fact describing the structure of the theory.

The idea could be viewed from another point of view. The independence of the elements stipulated in the Mirror Constraint requires, as we have said, that each one of them can be disclosed independently of the other. Let us suppose that we have no theory about the linguistic competence. Then, how could we reveal the causal structure within the cognizer so that we could use it later to compare it with some theory? What would evidence in favour of a given causal structure look like? Any answer to these questions requires a theory. As a matter of fact, there is no way in which that process can be accomplished since *causal-explanatory structure is to be specified by reference to the theory*. If we want to reveal causal structure, we have to look to some causal entities that *explain* the possession of a theory, so that if we have no theory, nor do we have causal structure. Therefore, Davies' criterion of psychological reality depicts a relation of *asymmetric dependence* between the two structures of the Mirror Constraint, the derivational and the causal. What we find here is that the Mirror Constraint provides a *reflection* rather than a *correspondence*, namely, it is in fact a *mirror* constraint not for the fact that it describes a matching, but because the theory looks itself in the mirror.

We can trace back the strains of the notion to when Davies looks for the right type of causal structure:

> The problem is: why does *causal* structure justify articulation in a *semantic* theory? The only fair answer to this query is, I believe, that it does not. That is to say, *mere* causal structure does *not* justify articulation in a semantic theory. We need to distinguish, and have no yet begun to distinguish, between causal structure that is relevant to the attribution of tacit knowledge and causal structure that is irrelevant. (...) What the philosopher is looking for is a constitutive account of tacit knowledge; and for a constitutive account of a syndrome. He needs to be able to say what it is for a causal structure to be of merely physiological significance; and what it is for a co-occurrence of symptoms to be psychologically accidental. (1987 pp.448-9, p.453)

Yet this task is not different from construing a competence theory for the ability. Moreover, when Davies sketches the different sort of causal structures that could match the derivational structure of the theory, defending the view that those justified in a comparison are those that are *psychological*. We can see the point clearer in the following passage:

> Opposed to Chomsky's realism, in Blackburn's taxonomy, are two positions on which 'psychological reality' is denied to the rules. One of these, unsurprisingly, is the hard-line Quinean position. The other is a position which 'finds the missing link between us and the rules in neurophysiology' (rather than psychology). It is this position that serves as a useful contrast to mine. It can best be understood by looking both at Blackburn's positive account and at his objections to it.
>
> On the positive side we have this: 'The idea is that our brains have a causal structure. Some "bits" are responsible for some aspects of our competence, and other "bits" are not'. Also, it might be that a particular kind of brain damage is found to produce a loss of linguistic ability structurally analogous to the derivational loss consequent upon the removal of a particular rule or axiom from a semantic theory: 'This would be empirical evidence that the rule or axiom is actually embodied in the user's neural

> processes. And, on the side of objections, we have this: 'But, it *might* be that our brains... do not encode information bit by bit. It is simply not true of them that if you destroy a particular area you lose one particular piece of information.'
>
> Now, suppose that it were nomologically impossible to produce localized brain damage in a subject in such a way as to delete precisely her ability with a particular word W, or a particular construction C. Then, according to the position under consideration, no semantic theory that had an axiom assigning a semantic property to the word W or to the construction C could be structurally adequate. But, says Blackburn, 'that seems incredible'. And he is surely right. (..).
>
> This position, described by Blackburn, has a number of notable features which differentiate it from my preferred account of tacit knowledge. First, and most important, although the position makes use of the notion of a structure in subjects that mirrors the derivational structure of a theory, the structure in question is *not psychological*. This is not merely the point that tacit knowledge is somehow unlike common-or-garden conscious knowledge and belief. It is the claim, rather, that the structure is not even psychological in the sense in which the early stages of visual processing are psychological. The remarks about bits and areas of the brain reveal that, on this view, the structure is *neurogeographical*.
>
> Second, there is not, on the position as described, a sharp distinction between what is constitutive of structure and what is evidence for structure. It appears at first that, on the neurogeographical view, what would be constitutive of a structure corresponding to the presence of an axiom for the word W would be the presence of a brain area, location, or bit, precisely responsible for mastery of W. In that case, facts about neurophysiologically possible damage would only be *evidence* for structure. (...)
>
> The contrast between the neurogeographical notion and the notion that I am attempting to characterize is rather like the contrast between two notions of a *syndrome*. The two notions that I have in mind are both distinct from the 'minimal' notion that a syndrome is nothing more than a constellation of symptoms. Both agree that only some constellations of symptoms constitute syndromes. According to the first notion, these are the constellations that have some common neurophysiological explanation, while according to the second notion, what is required is a common cognitive psychological explanation. (1986, pp.137-139)

Here Davies discusses the *neurogeographical* view, and any other non-phycological (non-P1) view for that matter, making an assumption about such positions that overlooks the root of the differences. The assumption is that those defending the neurogeographical or neurophysiological causal-explanatory structure share with Davies the same theory about linguistic competence, that is, that the only difference is in the position about the type of causal structure that underpins our linguistic ability. However, this assumption is unsupported. What the advocates of non-psychological causal-explanatory structure defend is precisely a non-psychological theory about our linguistic competence. The Churchlands (i.e., 1986, 1989) deny, for example, the *reality* of the theory at the psychological level, claiming that neurophysiology is the right level to account for cognitive abilities. Connectionists, on the other hand, reckon that cognitive theories should be worded in sub-symbolic terms. Therefore what Davies shows in the above discussion is not the inappropriateness of the causal-characterization, but that of the theory. And this is precisely because the theory *is* the causal structure of the cognizer regardless of the theoretical position considered. What Davies possibly points to is that there are certain causal facts that are not relevant. This is fair enough, but that has to do with what sort of *theory* happens to explain competence, since this point concerns *correct* accounts versus *incorrect* ones. In the example that Davies develops of the nutritional structure, the problem is not that the structure is *physiological*, but that the explanation is incorrect, or more precisely, that the connection between evidence and constitutive basis for competence is incorrect. This says less for the type of causal structure than for the sort of theory to use in cognitive explanation. What is more, the reference to *neurogeographical* causal structure is again a problem about a flawed attribution not about flawed type of causal structure. Causal structure that explains the possession of an ability there is only one for a given theory: its true causal structure.

One objection to this line of argument is that it seems to concern the long-debated issue of the underdetermination of theories by data.[11] This thesis amounts to the idea that different incompatible theories can make the same predictions. Where one is successful, so will be its empirical equivalents. There is no fact of the matter as to which one of two incompatible, but empirically equivalent, theories hods true. In other words, if one believes that there is a privileged class of observational (evidential) statements, then its possible to propose that there are many alternative theories equally well-supported by all true observational statements. However, this objection misses the point. The problem for the Mirror Constraint is that it is based on a spurious comparison between two structures, one of which is the mode of characterization of the other, whereas the underdetermination of

[11] See for example, Quine (1969), Newton-Smith (1978), Sklar (1991), Leplin and Laudan(1991), Ellis (1991), Boyd (1973) and Laudan (1990).

theories by data concerns the empirical groundings of theories. Indeed, for *even if theories were well determined, the Mirror Constraint would still be in the same position in deciding which of two theories should be licensed to attribute tacit knowledge.* The problem with the Mirror Constraint is not with evidence; rather, it is with the possibility of entertaining a comparison between one element which correspond to a theoretical structure with another element, a causal-explanatory structure.

Let us see how we could show the difference between both positions. For example, say that we try to understand an unknown computer which seems to execute a certain function (Oatley 1980). One hypothesis is that the function is a "legal chess move", but another interpretation is that the computer implements "arithmetic operation X between integers". Both hypothesis have an equivalent theoretical structure, so that they explain the behaviour and internal functioning in a robust way, and for that reason they are equivalent with regard to the Mirror Constraint. On the other hand, let us suppose that both have the same empirical consequences, and therefore are empirically equivalent.[12] Then what the thesis of underdetermination of theories by data prescribes is that *there is no fact of the matter that can support one interpretation against the other*. However, what my objection is concerned is not about evidence, but with the notion of causal structure as an element to be compared with a theory. So that even if we had evidence supporting one of the two interpretations, we would still be unable to compare them with the causal structure. Perhaps the key point is to note we should not confuse evidence with the notion of causal structure. What evidence supports, constitutively or not, is a given interpretation, that of "legal chess move" or that of "arithmetic operation X", which is the theoretical structure; data does not depict a causal structure, it merely reveals it under a theoretical interpretation. Therefore, the symmetry in the case of the computer between the underdetermination thesis and Davies' account is delusive.

Furthermore, my point is neutral on the sort of evidence that should count constitutive of a given theory. In this sense, it could well be that Davies is right in pointing that evidence that comes from revision of beliefs is constitutive, but if that is so then it is

[12] This is not only stipulative, since according to one particular proof, the Lowenheim-Skolem theorem, any computable function could be redescribed as an arithmetic computation between integers.

constitutive of a given interpretation. Of course, it is an empirical question as to what sort of evidence comes out as constitutive of the competence of speakers; it can be psychological, physiological or whatever.

Let us see other examples that show how the Mirror Constraint can lead us astray. Suppose that we want to explain a given cognitive ability. Say that theory Y accounts for this ability which has a derivational structure of the form A-B-C. Suppose that we construct such a theory by positing that there are three different operative states that account for such ability: m - n - o. The finding is robust, so that for any instance of the ability the structure holds. Hence, we construct a Mirror Constraint for that ability by assuming that those states account for the internalization of the theory by the cognizer. Then we happen to come up with another theory, which we label Z, with a derivational structure with the form D-E-F, such that the explanation of the ability appeals to three different states. And then we explain the possession of the theory by the existence of states m-n-o in the mind of the cognizer. Theory Y and theory Z have no doubt different semantics or, if you will, different informational implications. Thus, we have two theories with different semantics that are explained by the same causal structure of cognizers. In such a case we have built two different Mirror Constraints for -what seems- the very same causal structure, and therefore we cannot distinguish between the attribution of theory Y and theory Z, and of any other theory, for that matter.

This can be applied to a particular example. Say that we have explained a certain ability. Suppose that such an ability is the derivation of *modus tollens*. We then explain the ability by the possession of a logical theory that interprets the operative states and revision of beliefs about the axioms of the theory. That would be the end of the story for Davies. But then, suppose that we happen to interpret such operative states and revision of beliefs as *probabilistic derivations* which can account for *modus tollens* derivations. In other words, we assume that probabilistic transformations *underlie* logical derivations. The cognizer's internal states are the causally implicated facts that explain the competence of ability X, and the states and inferential steps of both theories happen to coincide in number and interrelations. In other words, we stipulate that the resources used to explain each one of the logical rules of the *modus tollens* are the same in both theories, with the difference that in one they are logical principles and in the other they are probabilistic principles. For example, we have resource A which is used in the production of inference a_1 within the *modus tollens* derivation. If theory Y and theory Z have a certain resource m that is used in the production of inference a_1, then both theories are equivalent with regard to their causal structure, *no matter the semantics of resource A*. Moreover, concerning the notion of systematic revision, the locus of systematic revision will be the same for both theories. If underlying a logical inference there is a probabilistic inference, then *the systematic revision of the logical belief will always go hand in hand with the probabilistic one*. Accordingly, we have two genuine Mirror Constraints for the very same type of evidence, which is not what we are looking for.

What has gone wrong? The problem is, of course, that the putative causal-structure stipulated in these examples is a myth; it is just "the mirror image" of the theoretical structure. If it were not for the fact that the theory has a derivational structure of three states A-B-C, we would never had stipulated three operational resources. Additionally, when we refer to a certain resource A which is used in the production of inference a_1, the fact that we attribute it logical or probabilistic properties depends on the theory and of evidence used to sustain any of the two interpretations. As Cummins puts it, "[i]n actual practice, of course, we don't identify the [internal] structure first, then provide an interpretation, but the reverse: our best shot at a grammar is posited as our best shot at an indirect specification of the structure it is supposed to interpret." (1983, p.44)

Is there a way out? I think there is. First, we can save the Mirror Constraint if we consider it to be a criterion of what constitutes an explanation in cognitive science, that is, a criterion of how to construct an explanation based on (v) above. The idea is that an explanation in cognitive science should amount to building a Mirror Constraint. Davies has himself noted the way to such an interpretation:

> ...Peacocke makes use of a notion of *differential* explanation (*Holistic Explanation*, pp. 63-89). For an explanation to be differential the invoked generalization (which need not be a fundamental law) has to specify a function (in the mathematical rather than the teleological sense) linking the explaining condition and the explained condition. In a way expressed by the function, the explained condition is sensitive to the explaining condition. To be a little more precise, the invoked generalization distinguishes, within the sufficient explaining condition, some constituent condition that is functionally related to the explained condition. It is then said to be the occurrence of the constituent condition which -given the background

> provided by the remainder of the explaining condition- *differentially* explains the explained condition.
>
> In the case of the speaker C, we shall want to say that the total explaining condition for his belief about the meaning of the sentence '*Fa*' has two constituents. One is the presence in a certain storage unit of the information about '*Fa*' and the other is the flow of nutrients in channels X_a and Y_F. It is the first constituent that differentially explains C's belief about '*Fa*'.
>
> With just this much grasp on the notion of differential explanation, we can see how it could be used in an account of tacit knowledge. For, in describing the causal explanatory structure of speakers, we could focus on facts of the following form: the states which differentially explain the speaker's beliefs about the meanings of s_1, s_2,...,s_n are sufficient for a differential explanation of the speakers belief about *s*. (1987, p.456).

Therefore, we need to take only a small step to see that Davies' notion is in fact a criterion to establish what counts as a cognitive explanation. The idea is then that the task of the theorist is, first, to stipulate those causal entities, states, or processes implicated in the explanation of the possession of the ability; secondly, isolate such states, entities or processes that are responsible for the possession of the ability from the whole lot which are not; and finally include them in the causal-explanatory structure of an ability as the "common cognitive resources at work in the comprehension of different sentence with common constituents" (1986, p.135). In other words, the construction of a competence theory is a psychological enterprise, even if it is an "armchair" task, so to speak, since the purpose is to reveal psychological facts that pertain to the causal-structure, facts that are then presented in form of sentences.

On the other hand, the objections to Davies account should not be interpreted as implying that there is no sense in asking the question "which one of two possible theories should be credited as being real?". This question is justified, and Davies' characterization of the answer is conceptually sound: The theory which corresponds to the causal-explanatory structure of the normal cognizer. And it is also justified to say that the choice of one theory over another can be supported empirically.

However, one suggestion to modify the account would be that what should count as support for the theory is not the evidence about "there is a locus of systematic revision for word C", but that the conditions of the elucidation of such piece of evidence. Specifically, what should be constitutive of structure is the robustness of the conditions of empirical elucidation. Hence, if we say that there is an operative state that draws upon that information, it is the robustness of the experimental set up, among other things, what supports the hypothesis. For example, if subject A responds to question q, the support should be given with a array of particular evidence of the sort "the subject understands the question", "the experimental set up examines grammatical competence and not any other function". Additionally, we must support the evidence with, first, theories that specify what counts to be a revision of belief, and, second, with empirical support that grants that when one subject of an experiment under conditions *C* changes his belief about a word W, he actually changes his belief about word W. Of course, this would imply a pragmatical approach. We could propose the following condition to sanction the reality of one particular theory:

(vii) A theory *T* is psychologically real if there is constitutive evidence for it.

Here constitutive evidence could be specified as such evidence for which there is robust support, and robust support being, in turn, sufficient surrounding theory, methodology and experience. The idea is obviously not new:

> (...) nevertheless, though there (may) be no principled way of deciding which natural number recursive function best capture the activity of some system, scientists still assign a unique computational interpretation to the measured relations. They do so based on their own understanding of how the physical system *relates to other accepted theories*. (Hardcastle 1996, p.67, my italics)

These theories can be cognitive, biological, ecological, developmental (the capacity of the physiological and psychological development and implementation of the system). Hardcastle even goes further and proposes to grant a cognitive function by using counterfactual scenarios and normal operation constraints:

> I submit that there is no principled answer [to the question of which function best capture the activity of some system]. We take counterfactual considerations as fundamental to our computational hypotheses; that is we rely on more than the actual data gathered -or even the possible data that could be gathered. Moreover, any possible world we entertain in order to expand our intuitions about what the system is

> doing is necessarily going to deny one or more principles of physics. (ibid, p.67)

However, this is only one of many other developments that we can undertake to complete Davies' account.

2.6.Too strong an account

I have argued that Davies' basic account, namely,

> (v) A theory is psychologically real if it describes the causal structure in the mind of the cognizer that accounts for the competence it specifies

can be considered to be a robust account. However, it is my view that it could be too strong to accommodate certain attributions that might be considered to count as psychologically real theories. This is so because the account leaves out some conceptually possible and empirically plausible alternatives that violate (v).

The attributions of cognitive theories are a central part of psychological practice. However, it is not clear that the underlying hypothesis about causal attributions in contemporary cognitive science would accord with Davies' account (cf. McClarmock 1995, p.21). I present three conceptual possibilities which are taken from actual psychological theorising. The presentation will only be a sketch of their conceptual possibilities, though I shall develop the last one in later chapters. As we shall see, the basic idea is that such proposals imply a sort of attribution which is not restricted to a completely internalized structure.

2.6.1.The extended mind

The first possible attribution of a cognitive theory that would not be licensed by Davies' notion, and yet considered to be genuine attribution of knowledge, is proposed by Andy Clark and David Chalmers (1995) in a paper with the title "The Extended Mind", a possibility that later was labelled as "The Scaffolded Mind" by Clark (1996). The idea is that when satisfying some tasks, a part of the world/environment functions as a process which complements those performed in the brain and which are necessary to fulfill a given cognitive task.

Clark acknowledges the origin of the idea in Vygotsky who "stressed the way in which experience with external structures (including linguistic ones, such as words and sentences [...]) might alter and inform an individual's intrinsic modes of processing and understanding." (ibid, p.45). They even acknowledge an observation of Simon: "Search in memory is not very different from search of the external environment" (Simon 1981). The notion of scaffolding includes all kinds of external aid and support, whether it be provided by other individuals or by the inanimate environment. The rationale is that in many cases the human organism is linked with an external entity in a two-way interaction, creating a 'coupled' system that can be seen as a cognitive system in its own right. All the components in such a system play an active causal role, and together they govern behaviour in the same sort of way that cognition usually does. If we remove the external component, the system's behavioural competence will decline, just as it would if we removed part of its brain. Clark and Chalmer's hypothesis is that this sort of coupled process counts equally well as a cognitive process, whether or not it is wholly in the head.

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. The authors use for clarification aims the game of Scrabble. In this case we make use of physical re-arrangements of letter tiles to prompt word recall. Clark and Chalmers contend that it is natural to explain a choice of words on the Scrabble board as the outcome of an extended cognitive process that involves the rearrangement of tiles on the tray. It could always be possible to try to explain the action in terms of internal processes, but in a sense the rearrangement of tiles on the tray could be considered to be not part of the action, but part of the very same "thought".

Kirsh and Maglio (1994) have investigated the issue from another point of view. They have introduced a new notion, *epistemic action*, to account for certain situations similar to those of Clark and Chalmers, contrasting it with *pragmatic action*. 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. Pragmatic actions are actions whose primary function is to bring the agent closer to his or her physical goal. The distinction between both is not a simple task, though, and Kirsh and Maglio devote a whole article to clarifying it.

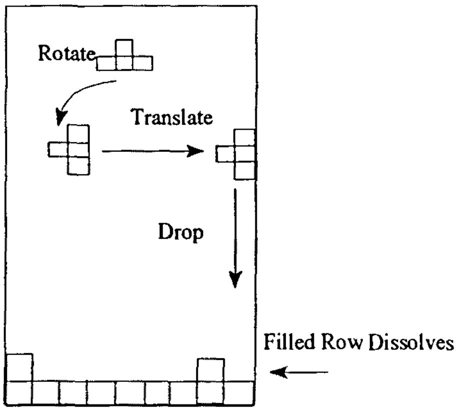

Figure 2.1. In Tetris, shapes, which Kirsh and Maglio call *zoids*, fall one at a time from the top of the screen, eventually landing on the bottom or top of shapes that have already landed. As shape falls, the player can rotate it, translate it to the right or left, or immediately drop it to the bottom. When a row of squares is filled all the way across the screen, it disappears and all rows above it drop down. (From Kirsh and Maglio 1994)

The notion of epistemic action has its origins in some examples in cognitive science, especially those 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. According to Kirsh and Maglio, there are many such actions in everyday activities. These include familiar memory-saving actions such as reminding, for example, placing a key in a shoe, or tying a string around a finger; time-saving actions such a preparing the workplace, for example, partially sorting nuts and bolts before beginning an assembly task in order to reduce later search. Kirsh and Maglio have shown that these sort of actions occur without the need of symbol manipulation. The example that they develop concerns a video game named Tetris (see figure 2.1). They have found that when subjects play Tetris, the actions of players are often best understood as serving an epistemic function. According to Kirsh and Maglio, the best way to interpret the actions is not as moves intended to improve board position, but rather as moves that simplify the player's problem-solving task. A player who moves a piece to the left of the screen and then reverses it back to its original position performs a series of actions that leave the physical state of the game unchanged. However, those moves allow the player to learn something, or succeed in computing something that seems to be worth more than the time lost by the reversal. Then, the idea of Kirsh and Maglio is that epistemic action captures the change between states arrived at after actions such as moving a piece. Specifically, they believe that when a player extra rotates a *zoid* (a Tetris shape), for example, such a move may serve certain cognitive functions such as:

(1) Unearth new information very early in the game (such as identify the form of a zoid).
(2) Save mental rotation effort.
(3) Facilitate retrieval of *zoids* from memory.
(4) Simplify the process of matching zoid and contour.

On the other hand, translation in Tetris implies to shift a zoid either right or left to permit placement in a column. According to Kirsh and Maglio sometimes payers translate a zoid right to left and back to its original position in order to judge the possible fit in one column. As long as the accuracy of judging spatial relationships between visually presented stimuli varies with the distance between the stimuli, a zoid dropped from a height of 15 squares has a greater chance of landing in a mistaken column than a zoid dropped from a height of 3 squares. Therefore, the obvious function of this translate-to-wall routine seems to verify the column of the zoid. By quickly moving the zoid to the wall and counting out the number of squares to the intended column, a player can reduce the probability of a mishap.

Let us consider another example, that of the calculation by means of an abacus. Suppose we have an individual who shows an arithmetic ability, multiplication, *only with the manipulation of an abacus*. The calculation is made by the individual, but he needs to rely on the arrangement of pebbles or beads. Could we deny that individual the attribution of an arithmetic theory that follows the steps of arithmetic operations which are normally executed following the arrangements of an abacus? In other words, suppose we have a theory that describes the whole operations of multiplication via an abacus, irrespective of what might be done by the brain and what by the abacus. Could we attribute such a theory to the individual? If we attribute it, then the causal structural facts in the cognizer will fail to satisfy Davies' notion, since some of the sub-operations and states of the whole operation are instantiated in the abacus. If we don't attribute it, then we will have to solve the difficult question of how such an individual is capable of "complying" with the operation without knowledge of the theory.[13] We could try to look for a theory that takes into account this sort of "delegation" of tasks, but then we would still need to attribute it. Furthermore, how could we attribute partial knowledge of the theory, if without the "aid" of the abacus the remains of the theory would not have any use?

In sum, if this idea happens to be plausible, at least in some domains, and even if it is true for only *one* type of competence attribution, then Davies' account is not suited to account for an attribution of such theories. Another thing is whether, for the theory, such an attribution merits the label of psychological reality, that is, if a theory that describes a causal structure is partially provided by the brain and partially by the world, it should be considered to be psychologically real. In my opinion this question could be seen only as a terminological dispute, and the choice between accepting or not the qualification of psychological reality of such theories will not solve the real problem, which is to assimilate those kinds of theories into normal cognitive theorising. Davies' account is proposed specifically to support the realist principles that underlie the cognitive science's quest for an explanatory account of cognition. If we restrict it to the theories that internalize all the processing steps completely, the account will lose its scope and therefore its usefulness, since it would reject out of hand an attribution for which a theory could be considered to be adequate. Moreover, I have argued there that the attribution of the theory to a cognizer required the inclusion of such "extended" processes, provided that without such an inclusion we would not be able to attribute any theory whatsoever.

2.6.2.Bounded knowledge

Another possibility that we should take into account is that a system might comply with a

[13] This question is trickier that seems at first sight, since the attribution of arithmetic abilities to humans do not usually take into account the fact that we verbalize (or write) much of our arithmetical operations, and there is no clear view of whether verbalizing is, as it seems, an indispensable element of our operations, or the role they actually play. It could well be that verbalization plays a similar role that the pebbles and beads of the abacus (see chapter 4).

theory, not by *thoroughly* executing its "derivations" but by being constrained, in some or all parts of the derivation, by the design of the system. If such is the case of the brain, it may happen that we could comply with some cognitive theory while not reproducing the complete theory necessary to explain the ability, and lacking a part, or even all, of the locus of systematic revision that Davies' notion is asking for. The idea can be traced to the concept of *bounded rationality* (Simon, 1981), which has been developed by various authors. A possibility of such "bounded knowledge" is described by Peacocke:

> The Berwick-Weinberg explanation of the holding of the Subjacency principle in the modified Marcus parser itself suggests a way in which this possibility might obtain. (...) Subjacency restricts the ways in which transformations may move a phrase. In Berwick and Weinberg's theory, Subjacency is explained as a consequence of certain properties of the finite control table of the parser. There are limitations on the kind of conditions that will not be violated. When Subjacency is explained entirely derivatively in this way, it need not be a *principle specifying information drawn upon* when the child learns new grammatical rules. (...) If the Berwick-Weinberg account is correct even as an account of a possible language-acquisition device, then we have a model of how it might be possible that human languages *fit* a certain constraint though *no rule stating that constraint, nor any state with the same content as such a rule,* enters the causal explanation of why this fit obtains. (1989, p.127)

Accordingly, the parser might constrain, or be constrained by, the possible procedures in such a way that makes certain derivations of the theory partially redundant or unnecessary, though they should figure in the causal-explanatory characterization of the competence. In the case just presented, Peacocke does not consider the matter of just which sort of constraints could account for that. McClarmock (1995) develops one idea about the type of constraints that might be considered. He distinguishes between procedure used *by* a system in the course of its operation and those used *in the design* of such system. In this sense, a system may be built to work *in accordance with* some general principle or strategy (say, avoiding additional inferences in favour of searching the environment for more information) because the designer knows that a system with particular capabilities available in a given case (e.g., better at perception than at more abstract inference) will, in general, be better off favouring its strengths The procedures for weighing which options might just be set up so as to accord with the appropriate task account that specifies the general logic of the capacity.

Peacocke also presents another case that could fall under the category of "bounded knowledge". This concerns the developmental nature of cognitive abilities. There exists the possibility that a certain competence should be accorded a specific theoretical element, but that due to the nature of the development of the ability, the element would need only be in its initial state of development and not in its acquired state, namely:

> [W]hen a principle *is* psychologically real in the initial state, it does not follow that in the acquired state that principle (or some parametrization of it) is psychologically real. Again, it is on the present conception a substantive, empirical claim that it is so. Consider Chomsky's Projection Principle, which states that lexical structure is represented categorically at every syntactic level; and let us suppose that it is, for a given individual, psychologically real in the initial state in the sense of the Criterion of Informationally-Determined Acquisition. We would then expect the psychologically real grammar of that individual's acquired state to respect the Projection Principle. But this could well fall short of some (possibly parametrized) Projection Principle being psychologically real in the acquired state. Indeed it is redundant for the device to possess the information that in the acquired language, lexical structure is represented categorically at every syntactic level. All that is required is that the lexical structure *be* so represented, and not necessarily that there exist a state with the parametrized informational content that it is so. (1989, p.127-28)

However, as the quotation implies, for Peacocke this possibility would make such an attribution not psychologically real for such a particular element. But then, would the whole theory where this element is included lose its psychologically real status? Peacocke makes it clear that the theory would need to consider it because it is used to explain the competence of the cognizer. His choice here is to draw the distinction between a truth being universal and it being psychologically real in one subject. But then, as we saw above, the problem will be to explain how a theoretical element can be universally true and at the same time enter in a causal explanation of a particular cognizer.

2.6.3.Cascade blockage

As we have seen in the last chapter Grandpa's explanatory model stipulates what we have called the "classical cascade", where a functional explanation runs through different levels of analysis of a given cognitive capacity. A successful cascade is a map in which the same function is computed through three levels: the task level, the algorithmic level and the implementation level. A basic requirement place on 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. Then, there are ways in which a correct functional attribution might violate the inheritance of the cascade. In some cases the cognitive system seems to satisfy a functional requirement by employing a number of strategies that cannot be interpreted as part of its functional analysis. In other words, a system may accord with a certain functional description that is not implemented as a form of internal structure. We interpret this 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. When a task level fails to be inherited, then the algorithm specified at the next level down is computing a completely different function, and hence is drawing upon different information. If this actually happens, as I will try to show in later chapters, the situation will have consequences for Davies' account of psychological reality. Indeed, if we have a theory of a specific cognitive capacity, which is correct for a given system, but for whose satisfaction the system employs a different ability, then we face an inconsistency. This possibility will be central in my dissertation and will be developed below.

Chapter 3

Peacocke's account of psychological reality

In this chapter I examine Peacocke's (1986, 1989, 1993) proposal regarding the psychological reality of cognitive theories. Succinctly, for Peacocke a theory is psychologically real if it complies with an explanation of what he calls the level 1.5. Peacocke distinguishes a level of description of a system that lies between what I have been calling task level and algorithm level. 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 it. In this chapter, I will argue that Peacocke's account is in the right direction for providing a robust account of psychological reality by giving necessary conditions for it, but which falls short of offering sufficient conditions. I will first present Peacocke's account with the aim of subsequently sketching some cognitive scenarios where the account falls short of providing sufficient conditions for psychological reality. Finally, I will develop what I consider to be the additional conditions needed to complete the account.

3.1.Peacocke's account

According to Peacocke, cognitive theorising omits a certain level of explanation within Grandpa's framework. Some theories offered by philosophers and cognitive scientists have to be viewed as an attempt to answer questions at this level. As we saw in the first chapter, the task level can be seen to specify only the function-in-extension, and this is the way in which Peacocke identifies the task level. The next level was seen to be that of the algorithm which describes the mechanism by which the task level is implemented. The level with which Peacocke is concerned lies between the task level considered as a function-in-extension and the algorithm level. For that reason he dubs it 'level 1.5'. This level states the information on which the algorithm draws. The notion of "draw upon" is offered as an intuitive conception supported by a number of examples. Succinctly, it can be spelled out in the following way:

> A state draws upon some information whenever such state carries the information which is causally influential in the operation of the algorithm or mechanism. (1989, p.102)

3.1.1.Drawing upon information

To illustrate his notion, Peacocke offers some examples. One is the *Kind Perception* (Peacocke 1986). This example is taken from Marr's experimental work. We can perceive a person as a man in almost all of the different orientations the person may have relative to us in three dimensions; within a certain range, it also does not matter where he is in relation to the subject. The perceived man himself may also have different shapes and still be perceived as a man. Marr and Nishihara sketched a theory of the structures and processes which make this possible, a theory which uses the idea of a 3D model description (Marr and Nishihara 1978). A 3D model description is a description, possibly hierarchical, of a type of shape; it uses various volumetric primitives, and uses a coordinate system in giving the relative position of the parts of the object to one another which are centred on the object itself. The 3D model description also specifies the permissible range of variation in the internal spatial relations of the various parts.

Marr and Nishihara hypothesized that the 3D model description is used as follows. The perceptual system has, by a certain stage perceptual information given in viewer-centred (egocentric) coordinates about the objects and their environment. The system defines a coordinate system on an object whose shape and kind has not yet been characterized in purely non-egocentric terms; for example, an axis might be placed on it by considerations of symmetry, or elongation. The object's parts, their shapes and relative positions are then characterized relative to this object-centred coordinate system by a process of searching through descriptions with those in the catalogue. Thus, the system is able to identify the perceived thing as a man, which is the label on a 3D model description in the catalogue. For Peacocke, the explanations that mention 3D model descriptions go beyond the task level (again, understood in terms of function-in-extension). In that sense, it is possible to conceive of a device which, inefficiently, simply lists all possible egocentric spatial descriptions of an object which will result in its being seen as a man. At the task level, this device may be characterized by the same function in extension from egocentric to non-egocentric descriptions, but it does not employ 3D model descriptions. A mechanism that involves a 3D-model description for the concept *man* may, for instance, draw on information that there is only a certain range of angles the legs of a man bear to his torso. No such information is drawn, for Peacocke, by the inefficient listing device. But it also seems that, while some of the explanations that involve 3D-model descriptions operate at the algorithmic level, not all of them will do so. We can have many algorithms for searching through the catalogue that are consistent with the role of the 3D model just outlined. For example, an algorithm may proceed either by checking first for a match of global structure and then seeing whether the membership of the sets at each node match, or alteratively it may check the sets at the top node and work progressively through the tree, testing at each stage both nodes and tree structure. The fact that one rather than another of these searching and matching procedures is used need not affect the fact that the same catalogue of 3D-model descriptions is searched in both cases. When search and matching procedures vary, the time taken to non-egocentric perceptual identification of a particular shape may also vary. The algorithms at the algorithmic level therefore differ, but some of what is explained by the 3D-model descriptions remains the same, though that model description is used by different algorithms. For instance, restrictions on the range of shapes that can be seen as a man may be explained by information drawn upon in the 3D model. Thus, in giving a 3D-model description, we are specifying the level-1.5 explanation.

Another example is the perception of *size and depth* in monocular perception. We can conceive of a system that explicitly stores something with the general informational content that all values of *r* (retinal size), *p* (physical size) and *d* (depth) are related by the equation $p = d x r$. In this system there is a common element in the explanation of instances of the regularity of efficiently calculating size and depth: the general informational content given in this equation is drawn upon in computations which determine perceived values. Storing the information that $p = d \times r$ is consistent with the use of several different algorithms in different organisms that compute those values.

The final case is that of *reading*, such as the ability to pronounce words one sees written on a sheet of paper. According to Peacocke (1989) we can fix on a given function described at the task level, a function from words to sequences of phonemes. We can also conceive three different bodies of information that might be drawn upon by the algorithm that computes the function:

a) "Reading lexical route", which is a list that specifies for each complete word the corresponding sequence of phonemes.

b) An algorithm which draws upon information about the pronunciation of sublexical syllables, and computes the sequence of phonemes for the whole word from information about phonemes corresponding to its parts.

c) An algorithm which draws on information about the pronunciation of a proper subset of the given class of whole words and extrapolates from their pronunciation via a similarity relation to the pronunciation of the remainder (no information about syllable/phoneme need be used).

For Peacocke we can have different algorithms which can account for each of the possibilities. For example in case *b*, different algorithms may compute the sequence of phonemes by proceeding through the word from left to right, another may proceed from right to left, a third may make the assignments by proceeding through some deeper tree-structure assigned to the word by a syntactic or semantic theory, and then reordering the phonetic representation to obtain the pronunciation of the surface word. The identity of the particular algorithm which draws upon the information that *ad* has a certain pronunciation is not then crucial for the explanation of the general fact about this subject that he pronounces any word of the form -ad- as something of the form [æd]. The explanation that cites this piece of information as drawn upon is correct whatever the detailed algorithm that draws upon it. The notion of information being drawn upon is a fully causal notion. Peacocke asserts that

> (...) explanations by facts stated at level 1.5 is a form of causal explanation. Facts about the meaning, syntactic structure, and phonetic form of expressions are causally explained by facts about the information drawn upon by algorithms or mechanisms in the language-user. (1989, p.113)

The notion of "causally influential" is not fully constrained. Peacocke recognises this, and wants it to be constrained so that it can accommodate both a "strong" and a "weak" claim (in terms of Higginbotham 1986, p.358). Specifically, Peacocke does not commit his proposal to the condition that the information be represented in some sort of way in some state, that is, he does not commit it to a language of thought hypothesis.[14] In fact, he accepts that the account could also allow a "weak" claim so that the state would not have such causal power unless the information "drawn upon" were true. In this regard, consider two algorithms for computing x^2-1 on input x. The first squares x and then subtracts 1 from the result. The second subtracts 1 from x, stores the result, then adds 1 to x and multiplies by what is stored. Then, for Peacocke, the second algorithm draws upon the information that $x^2-1=(x-1)(x+1)$, even if it is not mentally represented, since the algorithm relies on the information for its success. However, Peacocke sometimes seems to defend an intermediate position where the information should be "possessed" in some way by the system:

> In these remarks, I have committed myself to the possibility of complex states which realize a subject's subpersonal possession of a piece of information, even though the state does not consist in the possession of a representation of that information. (1986, p.391)

The problem is that he does not fully specify which sort of mental representation this

[14] The Language of Thought hypothesis is an hypothesis normally attributed to Fodor (1975, 1983) about the form of mental representation that postulates that mental representations are an unarticulated internal language (sometimes called Mentalese) in which the computations supposedly definitive of cognition occur and in which it is possible to express every distinction that may ever be drawn in any natural language.

intermediate position represents.

3.1.2. Transition types

Up to now, the criterion applies to informational *states* that instantiate, for example, rules of grammar of linguistic facts, such as the rule for the pronunciation of *ad*. However, Peacocke extends his theory to include the notion of psychological reality for *rules of inference*, that is for transitions between informational states. To accommodate such a move, he introduces the notion of *transition-type*. For a certain rule of inference T, we can say that some mechanism or algorithm in a subject uses the transition-type expressed by the rule T. When some mechanism or algorithm uses the transition-type expressed by T, it is in a state with the causal power of causing the organism, if it is in a suitable pair of states with informational contents of the form *a is a NP* and *b is a VP* to move into a state with an informational content of the form *a^b is an S*.

3.1.3. Informational Criterion for Psychological Reality

According to Peacocke level 1.5 is where a theory can be described as psychologically real. He proposes that for a theory to be psychologically real is to be in agreement with his *Informational Criterion for Psychological Reality*, specifically in a case for grammar of a particular language. This language has rules $R_1 ... R_n$ of grammar G, rules that state that $p_1 ... p_n$ respectively. Then, for the rules to be psychologically real for a subject is for the following to be true:

> Take any statement *q* of grammar which is derivable in G from rules $R_1 ... R_n$: then not merely is it true that *q*, but the fact that *q* holds for the subject's language has a corresponding explanation at level 1.5 by some algorithm or mechanism in the subject, an algorithm or mechanism which draws upon the information that p_1 upon information that p_2... and upon the information that p_n. (1989, p.115)

More generally we could adapt it to be able to accommodate any sort of cognitive theory:

> **Informational Criterion of Psychological Reality for a rule of axiom**. For any statement q of a theory X which is derivable in X from axioms $A_1...A_n$, which state that $p_1...p_n$ respectively. Then there is an explanation for the fact that q holds for a cognizer such that there is an algorithm or mechanism in the cognizer that draws upon the information that p_1, upon the information that p_2, ..., and upon the information that p_n.

The account has then to adapt the notion *transitions types*. For grammar G with axioms $A_1...A_n$ that state $p_1...p_n$ and proper rules of inference $T_1...T_n$, these axioms and inference rules are psychologically real in case:

> Take any statement q of grammar which is derivable in G form $A_1...A_n$ by means of $T_1...T_j$: then the fact that q holds for the subject's language has an explanation at level 1.5 by some mechanism or algorithm which draws upon the information that p_1..., and the information that p_n, and does so by using the transition-types expressed by $T_1...T_j$.(1989, p.116)

Again, more generally:

> **Informational Criterion of Psychological Reality for a transition-type.** For any statement q of a theory X that is derivable in X from axioms $A_1...A_n$ by means of $T_1...T_j$, axioms which state that $p_1...p_n$ respectively. Then there is an explanation for the fact that q holds for a cognizer such that there is an algorithm or mechanism in the cognizer which draws upon the information that p_1, upon the information that p_2,..., and upon the information that p_n, and does so by using transition-types expressed by $T_1...T_j$.

3.1.4.Equivalence Criteria

Following Peacocke we can establish a certain idealized theory, which is called *content correlate* of a level-1.5 explanation. Peacocke acknowledges the fact that there are *two* ways of construing cognitive theories, as content-using or non-content-using. For him, level

1.5 appeals to content-using theories. Content-using theories are theories that ascribe states with contents, that is, states that refer to objects, properties, relations or magnitudes. For Peacocke, the contents of states that have a common element in their explanation at level 1.5 are all in a certain sense derivable consequences of the information in the common explaining state. For example, a content-using theory of perceptual phenomena can be construed in the following way. Marr's theory of visual perception use the notion of a $2^{1/2}$ D sketch.[15] Part of the theory will be that the $2^{1/2}$D sketch concerns depth and orientation, that is, what, according to the theory, elements this mental representation refer to. The theory may then say that further states with content are computed from the $2^{1/2}$ D sketch. In other words, in understanding the theory one needs to assume that the representations it mentions have a meaning. Conversely, the theory of visual perception based on the $2^{1/2}$D sketch could be construed as a non-content-using theory. In such a theory, there may be mention of what is in fact the $2^{1/2}$ D sketch, and it may in the theory be associated with representations. *From outside the theory, one can say that one of these representations refers to a particular depth, and the other to magnitudes determining orientation. But the theory itself asserts no such thing, and any computational explanations it offers explain states under descriptions not involving content by prior states that in turn cannot be described as involving content.* Therefore, the distinction has to do with the metaphysical choice of the theorist. Some theorists might not want to use the notion of content by relying on the realization of Fodor's (1981b, pp.226-7) formality condition: the computational processes apply to representations in virtue of the formal properties (the syntax) of representations.

For Peacocke, a content-correlate theory is thus a theory that prescinds from the particular types of psychological states that have informational content. In this theory we focus on contents at level 1.5 in which we derive by proofs or computations the contents of the states which have explanations at level 1.5 from the contents of the explaining states at that level. This theory serves for any type of correlate of theories. Content-correlate

[15] A $2^{1/2}$D sketch is a stage in visual perception after the stage of establishing the two-dimensional properties in the retina. The $2^{1/2}$D sketch codes the orientation and relative depth of the visible surfaces and the contours of surface discontinuities. It is labelled the $2^{1/2}$D sketch because the co-ordinate frame is centred on the viewer The depth that is coded at this stage is not a position in three-dimensional space, but relative distance from the viewer's retina.

theories may vary in complexity. The content-correlate of the level-1.5 explanation which computes the depth of visual perception would be of the following form: it would consist of no more than a multiplication axiom relating three magnitudes, and an arithmetical apparatus for deriving its consequences. At the other extreme, Peacocke concedes that no one is currently in possession of a clearly correct semantic theory that would be the content correlate of a level-1.5 explanation of a subject's understanding of the English language.

Peacocke is eager to point out that his criterion of psychological reality has no problem in accommodating revisions of the current theories that are considered to be psychologically real, since the notion is intended to be a *general* notion of psychological reality. He even accepts that there might be another criterion stricter than his:

> The Informational Criterion is, then, not in conflict with the more specific, and we hope constantly improving, criteria for psychological reality used by the practitioner of psycholinguistics. (1989, p.117)

It is therefore not clear how many psychological realities we can have, though it seems that Peacocke intends to cover all notions under his:[16]

> The Informational Criterion rather aims to make explicit what state of affairs all [the psycholinguist's] evidence is evidence *for*. (ibid., pp.117-18)

3.1.5.Comparison with Davies' notion

Both Davies and Peacocke recognise that in some sort of reading both their notions can be considered to be equivalent:

> A description at the tacit knowledge level specifies the information that the system draws upon. But a full description at the processing level should surely do more than that; it should specify, in addition, how the

[16] Peacocke proposes for example a criterion for the psychological reality for Universal Grammar, in what he calls *Criterion of Informationally-Determined Acquisition*: "(...) for a proposed principle of universal grammar to be psychologically real is for it to give a specifically linguistic content drawn upon, or a specifically linguistic transition-type used, in the individual's acquisition of a psychologically real grammar for a particular language (1989, p.125)"

> information is drawn upon. To the extent that the processing level is to be identified with Marr's level two -the level of the algorithm- the tacit knowledge level should be distinguished as a slightly higher level of description. (Davies 1989b, p.547)

And Peacocke asserts:

> Here I confine myself to recording this belief: there are plausible ways of elucidating the informational conception and plausible ways of elaborating and extending Davies's notion to other domains, under which the two approaches can be shown to be equivalent. (1989, p.128)

However, the issue of whether both accounts are equivalent is a tricky question. First, Davies' account seems to be an *epistemic criterion* (a theory is real for a subject if we have identified the beliefs that explain its possession by that subject) whereas Peacocke's account (a theory is real of a subject if the cognitive system is shown to draw upon the content of the theory) is epistemically neutral: it is simply not necessary to have an explanation of the theory in epistemic terms to give an account of its reality, nor is it necessary to identify any specific structure in the subject. It is true that both proposals should not be considered to be orthogonal only for that, since they could give different perspectives of the very same conception. That, however, remains to be clarified.

Second, my opinion is that Davies' and Peacocke's proposals cannot be taken to be equivalent. For one thing, Peacocke's notion is ontologically less strong than Davies'. Peacocke's account is only based on the criterion of *information*. Conversely Davies' notion needs a *causal specification*, and the scope of each could concern a consideration made by Peacocke (1989, p.120) on possible objections towards his proposal. He argues that when we give a criterion for reality of some theory we may be aiming at three different things: a) what is to be realized; b) what is the realization *relation*, and c) the realization itself. He commits his proposal to *a* and *b*. By contrast, Davies is in fact aiming at *c*, since it specifies the way in which the theory has to be realized, a locus of revision for each axiom of the theory. As a matter of fact, the ontological implications of Peacocke's account only require that the information drawn upon by theory *must be causally significant* in the competence transactions of the cognizer, whereas Davies' notion prescribes a very strong condition: the isomorphism of the theory's derivations and the causal structure in the mind of the cognizer.

In other words, the only condition that a theory has to comply with to ensure its *reality* is for its informational implications to be taken up by the subject's cognitive system. In this sense, Peacocke does not prescribe a structure of any kind in the mind of the cognizer; he does not require any states, or particular form or structure in the mind of the cognizer to possess any piece of information in any particular way; it does not impose any condition on *revision of beliefs* or any other epistemic considerations. For example, the information could be carried by a state or by two or by a thousand distributed states, the important point is that the system *as a whole* draws upon the information specified by the theory. In short, what Peacocke asserts is that the information drawn upon the theory must be drawn upon by the system. Yet, the fact that they might not be equivalent does not rule out that they are mutually exclusive. If Davies' account turns out to be correct, then Peacocke's account will be correct. Peacocke's notion is in fact embedded in Davies's notion.

But finally, provided my analysis in the previous chapter, it is my view that Davies' account does not work as an *ontological* criterion of psychological reality, so long as it needs to be modified in order to accommodate different conceptual possibilities of knowledge attribution that could be considered to be genuine attributions but which are not accepted within Davies' account. Conversely, regardless of the objections I present below, my view is that Peacocke's account -or some variant of it- should be included in any notion of psychological reality. Summing up, I believe that both accounts cannot be compared on the same basis. This last point is developed below.

3.2.Objections to Peacocke's account

As I have already stated, Peacocke's notion gives a *necessary* condition for any notion of psychological reality, though it falls short of giving *sufficient* conditions for it. True, Peacocke offers a way to specify what a *correct* theory amounts to, in terms of information, but it is not a criteria to differentiate two extensionally equivalent theories. On the one hand, it is far from clear how to determine the piece of information which is causally relevant for a system; on the other, an informational specification of a theory is not sufficient to account for all possible cognitive theories.

3.2.1.Informational specification

How do we determine what information a certain cognitive system draws upon? As we have seen, for Peacocke psychologically relevant notions of information obviously need much more investigation. I think there are reasons to back this judgment. As Davies pointed out in the last chapter, it is widely accepted that no information-processing description is ever uniquely correct. As a matter of fact, there might be many correct informational specifications of the same event, object or property in all possible worlds whereas a system might only be sensitive towards one of these pieces of information of that very same event, object or property: Furthermore, Peacocke gives no criteria to choose which one of two pieces of relevant information is causally relevant for a system. Consequently, we might find that a system might agree with the information of a theory, but be drawing upon some other piece of information.

There is, for example, the example of the sight-strike-feed mechanism of the frog. Frogs catch flies by way of a strike with their tongue. It is assumed that mediating between the environmental presence of a fly and the motor response of the tongue strike there is some sort of mechanism that registers the fly's presence in the vicinity of the frog, causing its tongue to strike. 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 assumed story goes on to consider that the information drawn upon by state *S* is that of "fly" or "fly, there", deriving this information from the fact that the function of its underlying mechanism is to detect the presence of flies. However, there is a notorious problem with this sort of story. The present account assumes that the function of the mechanism grants the attribution of information to the mechanism, that is, if the function is to register the presence of flies in the vicinity of the frog, then the information drawn upon the mechanism must be that piece of information specified by "fly". Yet, there is an alternative construal of the information drawn upon by the internal mechanism, namely, that the information that the mechanism in question draws upon is in fact about "little ambient black things". In this case, the function of the mechanism is to mediate between little ambient black things and tokenings of a state with a causally relevant piece of information 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 information that the mechanism draws upon is different in each case and, hence, the frog's mechanism is functioning correctly even when the frog strikes at a little ambient black thing that is not a fly but a lead pellet (usually referred to as a "BB") that happens to be in the vicinity. The selective support for such a possibility, that is, the guarantee that a frog with this mechanism drawing upon that piece of information could have survived and reproduced, would be provided by the contingent fact that a sufficient number of little ambient black things in the frog's environment during natural selection were frogs (or edible bugs). Then the problem is how we are going to determine what the information is that a frog draws upon when it sees a fly pass its visual field? Is it a "fly", or is it a "black edible spot"?

3.2.2.Informational sufficiency

The central question in evaluating Peacocke's proposal is whether a complete informational account *exhausts* the description of a specific cognitive theory. In other words, can all explanations of cognitive functions or competence theories be exhausted by determining the information that their states draw upon? The following discussion will lead us to the consideration that the problem may lie in the question of whether the informational account is the optimal, reasonable or empirically sound level at which to pose the notion of a psychological level.

As we saw above, Peacocke asserts that explanations by facts stated at level 1.5 are a form of causal explanation. Facts about the meaning, syntactic structure, and phonetic form of expressions are causally explained by facts about the information drawn upon by algorithms or mechanisms in the language-user. My proposal is that *not only* these facts are required. As Peacocke himself recognises simply realizing an informational state is not sufficient for his account. For an informational state in his sense it is necessary to specify certain relevant connections that underlie the ability to be explained, namely:

> No state is the relevant informational state that 'ad' is pronounced æd unless it has certain relations to the ground of the ability to perceive an inscription as an 'ad'. (1989, p.113).

In other words, an account of psychological reality based on the notion of information must be supported by a sum of complementary *theoretical* conditions:

> (...) the contents of states which have a common element in their explanation at level 1.5 are all in a certain sense derivable *consequences* of the information in the common explaining state. They are, for instance, consequences -*in the context of a suitable surrounding theory*-" (1986, p.108, second italics mine).

He also gives an example of more complex contents, such as the content that 'man' is true of an object in the case it is a man. In such a case the required relations

> (...) will concern the state's connections with the subject's possession of the concept *man*, with the ground of his ability to recognize the word 'man', and with his general grasp of predication. (1989, p.113).

I agree with Peacocke in the need for a surrounding theoretical support, but disagree on the sufficiency of such appeal. As a matter of fact, I do not believe that these conditions are possible objections to a pure informational account, nor do I object to the following qualification:

> If a grammar is psychologically real by the test of the Informational Criterion, then there will indeed be the possibility for us, the theorists of making certain deductions from the information drawn upon by the mechanism in question. We can thereby move to conclusions about the output of the mechanism. To exploit this possibility *we have to ensure that in drawing conclusions from the informational content of states in the organism, we use only inferential principles which suitably correspond to the means of drawing on that content which are available to the organism itself.* (1989, p.124, my italics)

The specifications I am thinking about *are independent* of identifying the information that the system is drawing upon. These are conditions that constrain the attribution of a competence theory in a way that is relevant in cognitive terms but which do not fall under the specification of the information-drawn-upon. These could be placed under the set of possible answers to the question *how does this system handle information?* The nature of the conditions specified by the possible answers make a *cognitively relevant* difference in considering the equivalence of theories that draw upon the same information, and hence they should be considered constitutive of an account of psychological reality.

The logical counter-objection to this constraint is to appeal to the notion of algorithm. However, as I will try to argue below, the *task* or *function* that the theory of possible knowledge attribution specifies is precisely the task of how to handle the information in a given way. I will present examples taken from actual cognitive science theorising, though idealised in a way to make the points clearer. We will see that we have theories that compute the same function or have the same function, draw upon the same information, which are not theories at the level of algorithm and which are incompatible with each other, in addition to theories which draw upon the same information but compute or have different functions.

3.2.2.1.Theories of visual perception. The way in which the brain analyzes the visual stimulus has provided a corpus of theories that describe many different functions and mechanisms. I will refer here to one particular property of the visual system that seems to have an established consideration but with which we can stipulate an alternative that will help us see how there can be two theories drawing upon the same information but which underlie two very different theories with very different consequences for each of them.

T_1: Channels of visual perception. It seems to be well established (Livingstone 1988; Livingstone, and Hubel 1988; Zeki 1993) that the pathways that project the visual information to the visual cortex, and some areas of the cortex form a number of separate channels, each concentrating on a different property or dimension of the visual stimulus. There is one for processing form, one for colour, and one for movement and stereo depth. These three channels are intermingled initially in the pathway from the eye to the brain, but in secondary visual areas, the areas of the cortex to which the primary visual cortex projects, there seem to be discrete areas responsible for handling the different dimensions. Each of these areas may operate as an independent module having a distinct, specific visual processing function: colour, movement, form. In secondary visual cortex areas there are more than these three dimensions. Yet, we will stipulate that in this theory the dimensions fall under these three locus. In this theory then, perception is actually a set of distinct, heterogeneous processes, operating in parallel, which are somehow linked together to give

an illusion of homogeneity. If we are looking at a scene each of the three channels will be concentrating on extracting information about only one of these dimensions, and these three channels operate largely independently and in parallel, with some particular properties, for example:

a) Information in the different dimensions is loosely synchronized. Therefore, this will have many implications. As the history of painting has shown us, a small local patch of colour can be used to define the colour of a much larger bounded area and when the information from the form channel is put together with that from the colour channel, the missing information is filled in to give an impression of a uniformly coloured area. Likewise, it has been shown that when colour and form are arbitrary linked, that link is very easily broken. Treisman (1986) has shown that if three letters drawn in arbitrary colours, varying from trial to trial, for example a green X, a blue S, and a red T, are presented for 200 milliseconds, on about the third of the trials illusory conjunctions occur and subjects report seeing, for example, a red S. Treisman has argued that dimensional characteristics such as colour are not initially associated with a particular stimulus, but are in some sense 'free floating'. The linkage occurs later, either through the sharing of a common spatial location, particularly one being attended to, or through a top-down process. In sum, information in the different dimensions is initially rather loosely linked together allowing, say, an artist to create effective images with a very loose coordination between different dimensions, particularly colour/form and outline/texture.

b) Information in the different channels are processed independently. This will have the implication that the different channels will have different acuity properties. The form channel will have high acuity, it is good at discriminating fine details of the image. But the colour channel has poor acuity. Therefore, when defining shape with colours it is much more effective to have large areas of colour, whereas areas of fine detail will be best handled by non-colour systems. One phenomenon related with this property is called "bleeding". With sufficiently large areas, brightness or colour contrasts are exaggerated across boundaries, but with small areas of brightness and colour exactly the opposite effect occurs, effect receiving the label of bleeding. This effect is used by artists in two ways. In drawing and engraving, it produces the appearance of areas of even shading using lines. And in the pointillist paintings such as those of Seurat it produces subtractive colour mixing between spatially discriminable spots of light. If the dots of paint in a pointillist painting are seen from different visual angles, sometimes the low acuity of the colour channel will blur neighbouring dots together, as if the two patches of colour were mixed (blue and yellow dots will appear greenish), whereas at a different angle the dots are analyzed by the form channel as different patches of colour.

T_2: Unified channel. The incompatible theory is then easily seen. If we suppose that the scene is analyzed by a unified channel where all the different sources of information are mixed together, then the results will be strikingly different. Even if the information of movement, form and colour are taken into consideration, maybe not separately but as intermingled dimensions, the way in which this theory establishes how the information is handled implies very different consequences. Information is rigidly synchronized and not independently processed. Indeed, for none of the 'pictorial' or 'qualia' properties described above will apply here. The information about a patch of colour that does not cover all the area delimited by an outline is not seen as covering that area, contrary to what the previous theory predicts for the same information. If this were the case, the history of art would have been completely different.

T_3: Channels with synchronization. What is more, we can devise a theory that has the same channels that T_1 establishes but for which there is a condition, maybe simply due to the physics of the system, where synchronization among channels takes place. Accordingly, we can have the same structural analysing properties as T_1, but with the same consequences as such of T_2 where none of the effects of the visual scene occur.

What picture can be derived from these divisions? At Grandpa's task level we could say that T_2 and T_3 can be considered to compute the same function, which could comprise certain constraints such as "blurr colour in delimited areas". At level 1.5 of Peacocke, that is, at the level of the "information drawn upon", the three theories can satisfy the *informational criterion*, since they extract and process the same informational bits from the visual scene. Indeed, suppose that one rule of our visual perceptual system is R_1: "line segments that move together belong to the same object". Then, the information that this rule draws upon is true for the three theories presented.

It could be said that the mechanism implementing T_1, and not T_2 or T_3 could be drawing upon the information that "for any area if there is a patch of colour in it, consider the whole area to be covered by that colour regardless of the actual covered area". That, however, is unnecessary. The mechanism need not draw upon such information to yield the consequence of that assertion. This is simply a consequence of the way in which channels are realized. Note that you need only the channels to be synchronized for the effect to go away, and that is counterfactual supporting. Therefore, if some time for some unknown reasons the channels become synchronized, then the mechanism of T_1 will just yield the same consequences as the other two. The important point is that that piece of information is *not causally influencing* the mechanism. Finally, at the level of implementation we find an inversion for T_1 and T_3 is more similar than any of the two with T_2.

3.2.2.2.Theories of Attention. Attention is one of the areas in cognitive science to which a great deal of research has been devoted since it began as a discipline with an agenda. There have been many approaches, some of which general while others are focused on fields such as automatic versus controlled processing, visual attention and others (see Neumann 1995 for a review). It is difficult sometimes to subsume such approaches into general views on attention, but probably some idealization could be achieved and be plausible for the present discussion. The fundamental point is that theories of attention refer to certain mechanisms of the cognitive system that are directed at *handling* information to serve particular cognitive purposes. Therefore, from certain point of view we can consider that *what* information is drawn upon by such mechanisms is the least important thing for them. They handle information in a certain way and that way is what individuates such mechanisms, and what makes them either efficient or inefficient. Each state of such mechanisms is causally influenced by information, but by *whatever* information the mechanism is handling. It is influenced in a sense that "activates" the connections and tasks that are hardwired in the mechanism. To suggest an analogy, attentional mechanisms can be seen as motorways designed in a given way to favour the goals of the whole system and what theories do is to stipulate "road maps", the cars being the information which can be seen as equal in each one of the theories. In the case of theories of attention we can draw the opposition between two actual theories that draw upon the same information but which have different functions (computational level specifications)

T_4: Limited capacity-early selection. Some theorists have conceptualized attention as essentially the consequence of some kind of *system limitation* (Broadbent 1971 is the main reference): the result of limited or insufficient processing resources or processing capacity in the brain. In this sense the rationale behind all the operations of attention, imposing their essentially selective character is the limited information-processing capacity of a system. This has some consequences of subjunctive force, namely, if the brain had infinite capacity for information processing there would be little need for attentional mechanisms. Hence, at Marr's task-level the theory can be said to describe the basic function of attentional mechanisms so as to *protect the brain's limited capacity system from informational overload.* The idea is that the limited capacity restricts processing beyond some level to a selected subset of available information. This can be realized in a system where there is a central processor that possesses strictly limited resources.

There are a number of possible divisions within theories of attention that can be used to subserve my aim. One very important issue concerns, for example, the stage or level of analysis at which the supposed bottleneck of limited capacity is located, and accordingly the level at, or before, which selection must take place. There are two proponents. One (T_{4a}) is known as *early selection* (Francolini and Egeth 1980; Hoffman 1986, Kahneman and Treisman 1984, among others), where selective filtering of sensory information must occur at a relatively early stage of analysis, that is, before the stage of perceptual recognition or categorization. The opposite theory (T_{4b}) stipulates the contrary view, the *late selection view* (Coltheart 1984, Deutsch and Deutsch 1963; Posner 1978; Schneider and Schiffrin 1977, among others) in which attentional mechanisms operate when information has been already partially categorized. This opposition has many implications for the whole cognitive economy. For example, if we postulate that it is a late selection that is performed then this will imply that we might have information semantically encoded that will never be susceptible of becoming conscious, thereby opening a completely different picture of a cognitive system in comparison with the early selection view. However, each of the two proposals treats information in the same way because their functions at the task level do not change, they answer equivalently to questions about *what* or *how* concerning attentional mechanisms. Selection is synonymous here with selective processing, that is, the shutting out of non selected information from further analysis.

A second opposition within the framework of attention as overload prevention has been proposed in the past. It has to do with the number of loci of selection (see Neumann 1995). In the previous theories it was implied that we had only one locus of selection which could either be an early or a late locus. This could be one side of the alternative (**T_{4m}**). However, there have been theories (**T_{4n}**) that have proposed more than one locus, or even a continuum selectional process from the very beginning in, for example, the retina up to the higher association areas. In this opposition we again have two different systems that draw upon the same information but which apply to very different cognitive architectures.

There is still another opposition within theories of attention that could be relevant. In all the theories reviewed we have assumed that attention serves to regulate the flow of information in the sense of determining which portion of the input information passes through all the processing stages. But, how is the selective transition performed? There are two possibilities (Neumann 1995). It is possible that the non-desired information is somehow *prevented* (**T_{4x}**) from further processing, or it could be that the desired information is somehow given an advantage that *enables* (**T_{4y}**) its further processing. In other words, we can draw an opposition between inhibition or facilitation of information. Both theories can be seen to draw upon the same information though at the same time describing very different mechanisms, since within the latter opposition we should stipulate *how* information can be enhanced or inhibited, and different submechanisms to account for them.

T_5: Attention as selection-for-action. Some authors (especially Allport 1989) have regarded attention or the variety of attentional systems as a positive mechanism that uses information

to satisfy the need for action in the most efficient way, rather than being solely a processing *constraint* of the system. In that sense the task-level specification concerns an attentional system that must ensure the coherence of behavior under different conditions which the system faces such as:

a) Time. The environment can change at any moment and very quickly, and in ways that may be critical for the organism that require proper and quick action by the system.

b) Goals. A cognitive system may have a wide variety of goals at a specific moment. If the system is not to fall apart it needs some mechanism of priority assignment that adjusts the possibly incompatible goals or the most urgent actions

Then *what* the attentional system does, in terms of Marr, is to assume the responsibility of selective priority assignment among competing available information sources.

We can imagine two systems that perform each of the two "functions", T_4 and T_5, and both systems can be considered to draw upon the same information. In one the information is selected *because* the system has to prevent overload. In the other, the information is selected *because* the system has to use information in an efficient way. In other words, the system might use the mechanism for different cognitive objectives. If we implement algorithms for each of them, they will draw upon the same information (say that "there is a predator coming my way" is selected or priorized to the piece of information "what a nice female"), but for completely different reasons.

3.2.2.3.Theories of intelligence. The best example of different theories that draw upon the same information but which have radically different consequences is that of the theories of intelligence. It is true that there are no concrete theories of intelligence, but the widely held belief is that intelligence is the product of an "intelligent handling of information, whatever intelligence is". That is, we could have two mechanisms that draw upon the same information but one would be intelligent and the other not. This is about as radical as a difference as you can get.

Because there are no well-established, accepted theories of intelligence, I shall resort to theories of deduction. Much problem solving involves reasoning a way to a solution. Given this piece of information *x* and that piece of information *y* what can be concluded? Many of such inferences require the manipulation of information in special ways. Consider questions such as: which job should I apply for? Where am I going to go on holidays? In reaching a solution we have to infer what we will enjoy most or which may yield the best prospects. Take the following two sentences:

(1) a. There is some salad in the fridge and there is some chicken in the oven
b. There is some salad in the fridge

In reading these, we seem to feel that the sentence **a** fully entails sentence **b**. It seems to be a fact about human cognition that we can sometimes employ *some cognitive resources* to apprehend obvious (and non-obvious) entailments such as (1). Which mechanism is responsible for such an appreciation?

Inference rules. The simplest approach to explaining elementary entailments is to suppose that people possess mental rules that carry out the corresponding inferences. In our case we could activate a rule like *P and Q, then P*. When this rule is applied to a particular case then we can accept or reject some entailment. Mental rules of this sort are quite similar to rules in formal natural-deduction systems in logic. The most important feature of these rules is that they obey the logical form of the sentences that trigger them.

Deduction rules are not, however, a model of the way in which a cognitive system reasons. A psychological model of reasoning differs from rules in logic in incorporating strategic information. In their usual formulations, logical rules will generate an infinite number of new sentences from a fixed set of premises, whether or not these sentences are relevant to the task at hand. A realistic psychological system cannot afford irrelevant sentences. If we learn that everyone has 23 pairs of chromosomes, we do not want to conclude automatically that the Pope has 23 pairs of chromosomes. A solution to this problem is to revise the rule in such a way that the conclusion is drawn only when it is needed as a goal for the system.

Mental Models. Another way to understand how we recognize elementary entailments relies on an analogy to perceptual recognition. In this sense our ability to classify (1) as correct rests on an ability to construct "mental models". The term mental models has been used in many different ways but several common properties in these conceptions have emerged:

a) Mental models constitute a person's causal understanding of a problem and are used to understand and solve it.

b) They bear a similar relation-structure to the situation they represent.

c) Each entity in the world is represented by a corresponding token in the model. The properties of entities are represented by properties of tokens in the model and relations among entities are represented by relations among tokens.

d) Although the model can be constructed on the basis of language, its structure is not represented in language form.

The mental-model theory assumes that individuals understand the premises of the entailment and construct a model of the situation. When such a model is constructed, the individual can appreciate the different structural relations and extract or simply "see" the solution of the problem. Let us consider how one method deals with arguments such as (1). According to this theory, people represent a simple sentence such as (1), *there is a salad in the fridge and there is a chicken in the oven*, as a single symbol token, say *p*. The negation of a simple sentence such as *there is not a salad in the fridge* appears as an explicit negation sign (i.e. "¬") preceding the token: ¬p. This is similar to the representation in rule systems and in formal logic, and we can borrow the term *literal* from these systems to refer to simple tokens and their negations. If two or more literals are true in a particular state of affairs, the literals are aligned in a single row. Thus the sentences *There is a salad in the fridge and there is a chicken in the oven* each appears as in **a** in the table below and the sentences *there is a salad in the fridge and there is not a chicken in the oven* appear as in **b**:

a.	p	q
b.	p	¬q

Each row here is considered to be a mental model. Then we recognize entailments by noticing that the entailed sentence is true in all mental models of the entailing sentence. If we consider (1), then the representation of the premise is **a** above, and the representation of the conclusion is the single token *p*. Since there is only one model (i.e., row) in the premise representation and since the conclusion is part of this model, the premise entails the conclusion. (All this can be found in Johnson-Laird and Byrne 1991)

What is essential in a mental-model mechanism is that the rules are "implicit" in the constructed model. Additionally, the mental model construction makes the additional constraints that have to be incorporated in the inferential model to allow it to be a "psychological plausible" model useless. This is because the construction of mental models are just the construction of relevant situations. Therefore, if we encounter the premise "everyone has 23 chromosomes. John has 23 chromosomes" the mental model exhausts all there is to solve the problem and does not generate further models such as that of "modelizing" every single person we know.

Given this scenario, it should not be difficult to accept that both theories might draw upon the same information of the problem, identify the same relevant structural and logical relations between elements by different strategies and yield the same conclusions. Again we cannot say that the difference is at the level of the algorithm, because we could say that the "function" that the theory describes is, in fact, the "handling of information in a certain way", and that for each theory we could find different algorithms to compute it.

3.3.Information-under-a-cognitive-architecture

The aim of the preceding revision was to present some cognitive functional attributions in which the Informational Criterion falls short of distinguishing between possible scenarios in which there are cognitively relevant differences. We have seen that there are theories that describe cognitive structures whose function is to *handle information in a specific way* so that the way to specify them is independent of the notion of the *sort* of information which the mechanism is drawing upon. For example, attentional mechanisms are specified by the way in which they control the flow of information. The point is that an equivalence based on the information-drawn-upon groups together theories that describe cognitive systems with quite distinct, and sometimes opposite, behavioural outcomes. This seems to suggest that an account of psychological reality would need some complementary, though constitutive, conditions that might fall outside the notion of information.

It should be clear just what my contention is. I claim that there are cognitive mechanisms that are *informationally-neutral*, or at least not wholly specifiable in informational terms. The contention could become clear recalling the distinction that Peacocke draws between content-using theories with those that do not. As we saw, content-using theories are theories that ascribe states with contents, that is, states that refer to objects, properties, relations or magnitudes. We could use this notion to clarify my claim by asserting that the mechanisms I am aiming at do not possess states with *content*, but rather states with *instructions*: The states do not refer to objects, properties, relations or magnitudes, rather, they instantiate actions to be accomplished when suitable. Suppose an attentional state whose function could be described by the rule "move the piece of information *a* to location *z*". This description of the rule is in itself a piece of information, but it is not a *content-using* state. It is not the piece of information of the type that Peacocke is interested in. It is a piece of information that describes the rule. In other words, my claim is that it is obvious that we could describe the states of such mechanisms as the description of the mechanism, the rules to which the mechanism might be shown to accord with. Another thing is the informational content upon which the mechanism is depending. The point is then that the mechanisms we have reviewed do not rely on any sort of informational content; they are content-neutral. We could say that they are mechanisms that could be labelled as control mechanisms; they control the flow of information in a specific way.

There could be the temptation to consider that the objections fall under Peacocke's classification of non-content-using theories. This would be a misunderstanding of my point. Peacocke characterizes theories which are not content-using as theories that do not assert in the theory the contents of the states that it ascribes. As we have seen, the distinction has to do with the metaphysical choice of the theorist. However, my point does not concern the metaphysical choice of the theorist. I refer to theories that are *necessarily* not content-using, because they do not draw upon any information.

I think that the point could be easily accepted by Peacocke. There are many theories that do not appeal to content, and *could not* appeal to content. The contentious part of my claim is whether these content-neutral theories are cognitively important, and point to some relevant constraints that should be considered in any account of psychological reality. My analysis is that theories that describe flow-control of information are important so long as they may determine how efficient a cognitive system is in situations in which the difference implies such different outcomes as death and survival. It is precisely for this reason that researchers have devoted time and effort to devise such theories, and the level of constraint of such theories would fall outside the notion of information-drawn-upon.

However, two different objections could be posed against my claim. The first is that the reviewed examples should be seen as theories that fall under an explanation at the level of the algorithm. There are two answers to this objection. One answer could be succinctly put in the following terms. There are theories in cognitive science that describe mechanisms *whose function at the task level is that of handling information in a given way*, that is, that contents need not enter in the specification of the function that the system computes, and as such, they hold a one-to-many relationship with the variety of algorithms that account for them. All these theories can be described as being the level of the task, since a cognitive function of a mechanism can be to *control the flow of information in a specific way*, leaving for the level of the algorithm to describe *the specific way in which information is controlled*. In other words, if we were to specify the explanation of a mechanism within Grandpa's framework, we would have to specify the top-level, the task, as "handling information in such and such way". The *task* of a cognitive system need not be circumscribed to content-using functions, since there are many cognitively relevant operations that do not trade with contents. Additionally, there may be cognitive systems that are not solely constrained by their contents, but by other operational conditions. As we have seen, a theory can stipulate of a mechanism of attention that its aim is to "limit the load of information and move it to the categorization module".

The second answer grants Peacocke the attribution of the constraints of the reviewed theories at the level of the algorithm. But if we accept that then the Informational Criterion of psychological reality faces the conundrum of counting equivalent theories that describe such different cognitive scenarios as one implying survival and the other extinction, one yielding intelligence and the other stupidity. Therefore, assuming that an account of psychological reality is provided at least to have some empirical value, the account requires to be constrained by some complementary criterion to discard the attribution -within the set of equivalent psychological real theories- of those theories that yield an incompatible behavioural and cognitive outcome. However, in my opinion as long as parsimony is normally called for in science, it would be worthier to look for a unified criterion rather than to offer two complementary accounts.

The second objection that could be posed against the claim that a pure informational criterion falls short of a robust account of psychological reality would be to consider that the reviewed examples are not to be considered as competence theories, since they describe the cognitive architecture of a system. Therefore, we could not say that they are properly knowledge attributions. As Stich and Ravenscroft (1994) have put it for other purposes:

> If one thinks of theories as the sorts of things that make claims, and thus as the sorts of things that can be true or false, then one might be inclined to say that only the" knowledge stated in the form of propositions can be accounted as a theory. All that is a rule-based or program attribution cannot be accounted , for it can do a good or a bad job.

Again, there are two answers to this objection. The first one considers that competence theories are theories of cognitive abilities. Some cognitive abilities can be effectively described as organized pieces of information about the world. But there are other abilities that can be described as an efficient handling of such pieces of information, regardless of their contents. Hence it would be unwarranted to leave out theories that describe cognitive abilities solely because they are not "about the world". The second answer grants the characterization of non-competence theories to theories which are not constrained by contents. But then, we would again face the challenge of accounting for their attributions. Hence, we would need to devise another sort of account of psychological reality to be able to distinguish between candidate theories. Once more, parsimony recommends not to split what you can subsume; therefore, even in this case it would be better to look for a more general account.

All this should not undermine the suitability of the notion of information-drawn-upon in accounting for the psychological reality of cognitive theories. Furthermore, it is my view that Peacocke's proposal is a solid step towards an account of psychological reality. As I will try to develop below the notion of a "content-correlate" theory will be very useful. My view is that the problem in Peacocke's account stems from taking the heuristic strategies that have been used to specify Grandpa's explanatory framework too seriously. As we saw in chapter 1, Grandpa's explanatory framework is divided into three levels. The first level seems to state *which* function is computed and *why*, which is subsumed under the question of *what is the function computed?* The second level states *which* algorithm computes the function, which is subsumed under the question *how is the function computed?* The third level states how the algorithm is realized in hardware. The fact is that Peacocke has undertaken a genuine development of such a framework and that in doing so he has identified the right level of constraint for cognitive theories. However, Peacocke has also pressed Grandpa's framework to its limits, and these limits are imposed by the difficulties in incorporating the totality of the cognitively relevant constraints at the level of the task, taken as a function-composition view, or, what is the same, at level 1.5. We have seen that these constraints cannot be put under the consideration of being at the level of the algorithm, because they are constraints that specify the sort of function, at the task level, that the system computes. This means that the way in which a theory has to be constrained is not easily accommodated within Grandpa's framework. As I will try to show below there is a distinction in cognitive functional attributions, orthogonal to Grandpa's framework, that can account for such inconsistencies. This distinction focuses on two different explanatory projects. One corresponds to the explanation of the way in which cognitive agents seem to comply with a class of functional efficiencies that might be selectively relevant. It is an explanation that has to couple an agent with its environment, and it is to be characterized according to a given 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 hypothesis is that Peacocke's proposal concerns the first one of such projects, and that the problems appear because the Informational Criterion inadvertently subsumes the first and the second sort of attributions.

I now sketch what sort of constraints are needed to complement the notion of information in an account of psychological reality for a given theory. The full development is presented in Chapter 6. Essentially, the suitable constraint for Peacocke's proposal of psychological reality is an account of an information-under-a-cognitive-architecture proposal. The role of such a complementary condition is not only directed at specifying how the mind handles information, but to constrain the psychological reality by subsuming cognitive relevant characteristics that fall under different levels of the Marr's framework. For me the complementary condition is what is known as *cognitive architecture*. Note for the moment that this notion of cognitive architecture generally refers to a related notion that has been discussed by many authors, especially Pylyshyn (1984). For them, there is a level of cognitive specification which represents the right level at which to view mental processes. This structure is the sort of functional resources the brain makes available- what operations are primitive, how memory is organized and accessed, what sequences are allowed, what limitations exist on the passing of arguments and on the capacities of various buffers, and so on. My proposal is that a full account of psychological reality must be constrained by the cognitive architecture of a system. Therefore, it will be unnecessary to specify *where* in such framework we regard psychological reality to be located. This proposal will allow the integration at the same level of equivalence or incompatibility theories that are individuated by the following constraints:

> **Knowledge Representation.** What is the language in which our knowledge is represented? This is a central question of all disciplines interested in cognitive science, and it could even be said to be the "conundrum" of cognitive science. Knowledge has been described in many ways. One kind of knowledge can be verbalized, visualized, declared in some manner, and for these reasons has been called *declarative knowledge*. A second type of knowledge consists of skills, cognitive operations, knowledge of how to do things, and has been called *procedural knowledge*. This distinction is not very precise, but it points to specific considerations which have been widely discussed in the past forty years. Declarative

knowledge can be, in turn, represented in two ways. *Analogical representations* preserve properties of objects and events in an manner in which the representational system has the same inherent constraints as the system being represented. The second type of representation is *symbolic*. Such representations are structured, semantically interpretable, objects which hold an arbitrary or conventional relation with the objects they represent.

Modularity. If we extend the notion of cognitive architecture to the general mental structure of mind, in terms of Fodor (1983), then we could have another constraint to cognitive theories. Such constraints might have very important implications, as Fodor has repeatedly stressed and is at pains to remind the community of cognitive scientists:

> The limits of modularity are also likely to be the limits of what we are going to be able to understand about the mind... Specifically, if central processes have the sort of properties that I have ascribed to them [namely, non-modular], then they are bad candidates for scientific study. One relatively minor reason is this. We have seen that isotropic systems are unlikely to exhibit articulated neuroarchitecture. (...) The moral is that, to the extent that the existence of form/function correspondence is a precondition for successful neuropsychological research, there is not much to be expected in the way of neuropsychology of thought. (...) There are, however, much deeper grounds for gloom. (...) The condition for successful science (in physics as well as psychology) is that nature should have joints to carve it in: relatively simple subsystems which can be artificially isolated and which behave, in isolation, in some way similar to how they behave *in situ*. Modules satisfy this condition; Quineian/isotropic-wholistic-systems by definition do not. If, as I have supposed, the central cognitive processes are nonmodular, that is very bad news for cognitive science. (Fodor 1983, pp.126-128)

Note that using Fodor's argument here is independent of whether the mind is modular or not. The point is that the two models of mind, modular and nonmodular, underpin very different cognitive theories.

Computational models. Any cognitive theory is based on a computational model or paradigm. In general, cognitive science is based on the information-processing model, where mental representations are taken as symbols and the computational processes get such symbols transformed to account for our cognitive abilities. However, the informational implication of a theory is independent of the computational model within which it is modelled, and therefore it could, and it fact has happened, that we could have two informationally equivalent theories that are based on different computational models. The example of connectionism, which stands in opposition to the symbol paradigm, is pertinent here. Here we could have two theories that make the same informational implications but which describe very different cognitive systems. Should they be counted as equivalent? Peacocke argues that his notion can adapt to a connectionist theory, but the question is: should we take two informationally equivalent, but paradigmatically different, theories to be equivalent?

Teleology. A state can be said to instantiate some sort of information. However, some theorists have proposed that a state should have a teleological requirement on content instantiation. A physical state of an organism will count as realizing such-and-such a functional description only if the organism has genuine organic integrity and the state plays its functional role property for the organism, in the teleological sense of 'for' and in the teleological sense of 'function'. The state must do what it does as a matter of its biological purpose, so to speak. We then need a very important addition, a functional attribution has to be made taking teleological considerations into account. And the fact is that we could have two equivalent informational theories that subserve different teleological functions. Should we count them as equivalent? As Cummins puts it "if some future interstellar archaeologist were to discover a computing device of an advanced but extinct population, questions about what the device could do would have to be distinguished from questions about what it was intended to do. Having discovered something it could do, it would be perfectly in order to construct an interpretive functional analysis to explain *how* it could do it, with no regard to whether it was *designed* to do it. Were we to discover that the device could fly, we would be obliged to explain *how*, whether or not anyone ever intended it to fly. " (1983, p.41)