

## The Philosophical Import of Connectionism: A Critical Notice of Andy Clark's *Associative Engines*\*

---

MANUEL GARCÍA-CARPINTERO

Andy Clark's book (Clark, 1993; henceforth abbreviated as 'AE') is a fresh examination of the conceptual issues raised by connectionism. The book develops, from new perspectives, the theme of connectionist systems' advantages over Classical systems in recreating aspects of advanced cognition—like 'generalization', 'graceful degradation' and so on. It also defends connectionism against the claim that it cannot properly represent structured thought, both by suggesting how connectionist algorithms could do so and by arguing that, when closely examined, Classical systems do not appear to be on a much better footing in that regard either. The arguments are informed by some recent developments in the relevant fields of Cognitive Science, and they are also clearly and engagingly presented. The book makes informative, stimulating and enjoyable reading.

This appraisal does not imply that the present writer finds Clark's work successful in achieving its most ambitious philosophical goal. Contemporary philosophical interest in connectionism adds one more chapter to an old dispute between opposing philosophical views on the mind—views which, for the sake of convenience, I shall label henceforth *Cartesian* and *Skeptic*. (The labels are intended to be suggestive more than accurately descriptive.)<sup>1</sup> Cartesians are united in the connected beliefs that (i) the categories of folk

---

\* Andy Clark, *Associative Engines: Connectionism, Concepts, and Representational Change*. Cambridge, MA.: MIT Press, 1993. Pp. xiii + 252.

I would like to express my gratitude to Andy Clark, Begoña Navarrete, David Pineda, Ignacio Vicario and an anonymous referee for *Mind and Language* for many valuable comments that led to improvements. Research for this paper has been funded by the Spanish Government's DGICYT, Ministry of Education, as part of the research project PB93-1049-C03-01.

**Address for correspondence:** Departamento de Lògica, Història i Filosofia de la Ciència, Universitat de Barcelona, Baldri Reixach, s/n 08028 Barcelona, Spain.

**Email:** garcia@cerber.mat.ub.es.

<sup>1</sup> They also involve an implicit homage to the views of the researcher whose work has done most to advance what I call Cartesianism in recent times, namely, Noam Chomsky.

psychology—those we use to make sense of ourselves and our fellow humans as rational beings, subjects of rational criticism and appraisal—stand or fall just to the extent that those categories' causal-explanatory role might be vindicated by applying to them the same standards of evaluation which are invoked when dealing with other concepts intended to do causal-explanatory duty—given the scientific-realist attitude dominating ordinary scientific practice—and (ii) that so far there are no reasons, empirical or conceptual, to think that they will fall. Otherwise, the label 'Cartesian' encompasses severely disparate views on the mind: those of philosophers who think that the requirements in (i) involve a commitment to the *identity* of folk-psychological properties with physical properties—like David Lewis contemporarily—together with those of philosophers who think that only something weaker, like 'strong supervenience', or maybe just some sort of nomic covariation, is actually required—like Descartes himself or John Searle nowadays; those of philosophers who think that (ii) can only be true to the extent that we can provide an 'individualistic' interpretation of the contents mentioned in folk-psychological concepts, compatible with 'methodological solipsism'—like Descartes himself and a former stage of Jerry Fodor—together with those of philosophers with 'externalistic' viewpoints, like Tyler Burge.

The *Skeptics*, on the other hand, reject at least one of the claims that identify the Cartesians. Aside from that, this label also extends over a ragbag of theorists, including stubborn believers in folk-psychological categories who, unwilling to give any hostages to empirical fortunes, take a 'soft', behaviouristic understanding of them—like the Wittgenstein of the *Philosophical Investigations* or, more contemporarily, Davidson and Dennett (at least in their classical writings in the seventies as they are commonly understood)—together with eliminativists of a more severe scientific-realistic persuasion like the Churchlands. Now, the big philosophical issue surrounding the debates on connectionism concerns the claim that Skepticism might be vindicated by the findings of recent research on connectionist modelling of cognitive processes—which stands in direct opposition to the contention made formerly by Cartesians, notably Jerry Fodor, that empirical research in Cognitive Science vindicates their own views. Clark's book is no exception in being overwhelmingly concerned with this issue.

Cartesianism received a big boost mainly through the empirical research stimulated by Noam Chomsky's proposals and the philosophical interpretation of it associated with the work of Jerry Fodor. Clark calls that cluster of ideas *the Syntactic Image*. I shall review its main tenets, together with the link between the Syntactic Image and Cartesianism, in Section 1. I shall attempt to be very careful in setting up the issues, for two reasons: first, because, as will become apparent in Section 3, much of the evaluation of Clark's line of criticism depends on being clear about them; second, for the sake of readers not particularly familiar with the specifics of the philosophical debate—to whom this review is also addressed. Clark's grand design is to reach the conclusion that connectionist findings ruin any hold on the true nature of the mind that the Syntactic Image might be thought as having;

what emerges as his alternative view is something closer to the proposals of the Skeptics with behaviouristic proclivities. I shall try to examine the difficulties of this view in Section 2, and I shall indicate also some grounds for complaining that details crucial to an evaluation of Clark's own alternative are missing. Nonetheless, it is relative to its grand design, I think, that the book fails. Very important conceptual distinctions are overlooked at the crucial turning points—a shortcoming that mars the arguments and spoils their plausibility. This is, at least, what I intend to argue in the final section of this review.

### 1. *The 'Syntactic Image'*

It is very important to be absolutely clear about what the commitments of what Clark labels the 'Syntactic Image' really are, and what they are not. I shall review the issue in this section. I merely take myself as presenting the views that can be found in such works as Fodor (1987) or Fodor and Pylyshyn (1988); but no attempt will be made to substantiate this exegetical remark with detailed references or hermeneutical considerations.

We start with the facts of *systematicity* and *productivity*. Or better, given that (i) in the presence of a plausible assumption (namely, that the productivity we are speaking about is one ultimately to be explained by cognitive facts about 'finite' beings like us), productivity implies systematicity without being implied by it, and also that (ii) some might think that to take productivity as a fact would beg the question against the friends of connectionism, we start with the weaker fact of systematicity. The fact is usually not given by definition, but just by gesturing towards paradigmatic alleged instances.<sup>2</sup> Let me therefore try to state more explicitly what I take it to be.<sup>3</sup> I shall first motivate my account of systematicity and only then provide it.

<sup>2</sup> Clark's definition is in this respect typical: 'More precisely, a thinker is systematic if her potential thoughts form a kind of closed set—i.e. if, being capable of (say) the thoughts "A has property F" and "B has property G", she is also capable of having the thoughts "A has property G" and "B has property F".' (AE, p. 147) Given that a 'set' might be 'closed' by any set of closure principles one may wish to devise, including the null one, and no constraint is placed on the nature of the intended closure principles—facts which, incidentally, make redundant the additional vagueness-introducing hedge 'a kind of'—we are left with the example to make do for the announced surplus 'precision'.

<sup>3</sup> An alternative definition may be found in Davies, 1991. He takes systematicity to characterize processes, and makes it relative to a way of describing them. Because of my dissatisfaction with the second aspect, my approach is slightly different—although the phenomenon we both try to grasp is the same. I am aware that the proliferation of notions of systematicity may be deplored. If I am going to add to it, it is because in my reckoning the confusion this could provoke weighs less than possible misunderstandings that could arise from using notions already proposed. The notion developed here is much more detailed than the one assumed in Fodor and Pylyshyn, 1988; moreover, it tries not to beg any question against friends of connectionism who, like Clark, have behaviouristic leanings. I try to carefully present what everybody involved in the present discussion should acknowledge as a fact. This distinguishes it from the expla-

First, the motivation. There are types of behaviour whose instances are assumed to have—in both folk- and scientific psychology—*intentional* explanations *logically independent* of each other; and there are behaviours which, also relative to each other, are assumed to have *logically related* intentional explanations. Understanding and producing sentences of natural languages is the typical example of the second variety, but it is far from being the only one: perceiving the grammaticality in a given language of certain strings of expressions, and producing grammatical strings; perceiving and producing words—i.e. expressions commonly written between spaces—as words of a given natural language, and perceiving and producing strings of sounds as belonging to a given language are further examples of the second variety.<sup>4</sup> Let us refer to types of behaviour in the first class as *unsystematic*, and to the types in the second class as *systematic* behaviours.

Systematic behaviours presuppose unsystematic ones. To understand, and to sensibly produce, sentences of a natural language presupposes that one understands the meaning of certain lexical units (*not* words, in the formerly indicated sense, but typically things smaller than words: the root of a verb, the endings of verbs, etc.). The latter type of behaviours—to understand and sensibly produce lexical units—are *intentional*. This just means that behaviours of that type have their explanatory origin in an intentional state or a combination of such states—beliefs, intentions, desires, in general states with *representational* or *informational content*. To understand and sensibly produce lexical units is explained, at the very least, by such intentional states as, say, *knowing that such-and-such a type of expression is associated with such-and-such a concept*. Of course, the intentional states which figure in explanations of those behaviours might turn out to be complex, even *intentionally* so. But the intentional origin of each one of the behaviours in the type is not assumed to be *logically related* to any of the others.

The explanatory source of the types of behaviour we are considering is

---

nation in McLaughlin, 1993, Sec. 2, to which I am otherwise very sympathetic. The assumption of 'constitutive bases' for intentional capacities could be found questionable by those philosophers—for reasons explored in Section 2.

<sup>4</sup> For the sake of presenting in a careful way what everybody involved in the present discussion should acknowledge as a fact, I am going to focus on public-language linguistic abilities. It is my view, however, that these are neither the only, nor even the core cases. The core cases involve abilities which are required to have public-language linguistic abilities, and which can be possessed and are actually possessed without having public-language linguistic abilities. For instance, if an organism is able to perceive a red sphere one meter in front of him, he will typically be able to perceive also a green sphere of the same size one meter in front of him. If an organism is able to perceive a small red sphere one meter in front of him, and also a big red sphere one meter in front of him, he will typically be able to perceive also a red sphere one meter in front of him intermediate in size between the preceding two. There is also systematicity (in the sense developed in the main text) in these cases. See also McLaughlin, 1993, Sec. 2. Finally I should stress that, as these examples also suggest, in my view the 'symbols' mentioned in Classical accounts of systematicity might well be 'picture-like' instead of 'word-like'. The features of symbols which are crucial for classicists are indicated below.

an *ability, capacity or disposition*, and an intentional one. That the intentional abilities explaining a given user's understanding of a lexical unit are thought not to be logically related to those explaining any other is another way of saying that the ability to understand a given lexical unit is not thought to give its possessor the ability to understand any other lexical unit. In contrast, the ability explaining the understanding and production of sentences is intentionally complex, because the ability to understand any sentence is assumed to be related—by sharing simpler intentional abilities with it—to the ability to understand other sentences.

The precise nature of the complexity of the intentional abilities explaining systematic behaviours is only surmised in folk-psychological theories. It is slightly more carefully articulated in some philosophical theories of semantic abilities. In some easier cases, it has been specified in a scientifically more precise way. As I mentioned previously, there are many more examples of systematic behaviours than the paradigmatic case constituted by our ability to understand sentences: perceiving the grammaticality in a given language of certain strings of expressions, and producing grammatical strings; perceiving and producing words as belonging to a given natural language, and perceiving and producing strings of sounds as belonging to a given language are further such examples. Phonology, for instance, provides precise empirical theories regarding the latter case, giving detailed accounts of how the ability to perceive and produce sounds as belonging to a given language is articulated out of simpler intentional abilities, like, say, the ability to identify sounds as having certain articulatory traits, or producing sounds with those articulatory traits. To give the exact details of the complexity involved in any interesting case is therefore a matter for difficult theorizing; it cannot be said that it is a *fact* that a given type of behaviour is intentionally complex in any fully specific way. However, I take it that it is a fact that the examples I have given are examples of systematic behaviours—when the statement of this fact is understood as not committing us to any particular fully detailed general account of the intentional ability in question, but only to a vaguely specified theory of the assumed systematicity's nature. This is the contention that even the arch-behaviourist Quine grants, quoting approvingly a passage by Postal: 'The claim that there are linguistic rules is simply the claim that individuals know their language and have not learned each of its sentences separately.' (Quine, 1972, p. 443)

Because assumptions regarding simple and complex intentional explanations for behaviour concern *abilities*, i.e. states with an *explanatory* role, they entail counterfactuals and other subjunctive statements. The claim that understanding the root of verbs is intentionally simple implies that if a given speaker were given a verbal root which he has not encountered before, his abilities to understand lexical units would not be sufficient for him to understand the new one. The claim that understanding words is intentionally complex implies that if a given speaker were given a word he had never encountered before and the word were made out of units appearing in words he had encountered before, he would be able to understand it. To say that a

given class of behaviours are or are not systematically related is to say something (positive or negative) of an explanatory nature about the cognitive abilities that produce them, and thus it is to commit oneself to something (positive or negative) about how the subject would behave with respect to new cases. To say that the abilities a given subject has to classify the sentences in a given class as grammatical are not systematically related is to say that his knowledge is of a list-like nature, and thus it implies something about how he would generalize to sentences not in the class (he would not). By the same token, such a claim also implies subjunctive statements about how he might have *acquired* the abilities: it implies that he could not have acquired them so that, after learning that some of the sentences in the class were grammatical, he would judge that some others also in the class were grammatical too. And it also implies something about how the abilities would be lost: it entails that losing one would not necessarily involve losing any other. On the other hand, to say that the abilities are systematically related implies subjunctive statements about definite patterns of acquisition and loss. Moreover, different accounts of the systematicity involved in a given class of systematic behaviours make different predictions about patterns of acquisition and loss. Even when two different accounts of the systematicity involved in a class of systematic behaviours (say, recognizing as grammatical the sentences in a given class) are 'extensionally equivalent' (i.e. the class is the same in both cases), they may well make different predictions about patterns of acquisition and loss (one may predict that, given that the subject had already acquired the ability on the basis of a certain proper subset *S* of the class of grammatical sentences, he would be able to recognize a sentence *s* also in the class as grammatical, while the other may make the opposite prediction).<sup>5</sup>

The subjunctive statements implied by a theory of the alleged systematicity, furnish also the observational predictions with respect to which the theory is confirmed or disconfirmed. It is very important to remember that those subjunctive statements are to be understood as hedged by conditions

---

<sup>5</sup> See Evans, 1981, where all of this is argued for and clarified by means of useful mock-up examples. We find these contentions in Hadley, 1994: 'In their discussion of systematicity, F[odor] & P[ylyshyn] assume the grammatical competence of cognitive agents, but they are not concerned with the ontogeny of systematicity. Nor do they discuss *degrees* of systematicity. Rather, their focus is upon a particular causal precondition of systematicity, viz., *structure-sensitive* operations. [...] unlike F&P's conception, the concept of systematicity defined here fundamentally involves issues of syntax learning.' (Hadley, 1994, p. 249) Then he goes on to define different degrees of systematicity. But there is a fundamental misconception here. For, as we saw, claims 'upon a particular causal precondition of systematicity' entail (just by being *causal, explanatory* claims) claims about the 'ontogeny' of systematicity. To say that a certain behaviour is *caused* by determinate structure-sensitive operations entails claims about how that system of operations might be acquired (and about how it might be lost, and also about many other aspects which people better than I am at designing experiments to test specific causal claims are undoubtedly able to imagine). Hadley's 'degrees of systematicity' result just from differences in the background theories of the systematicity at stake.

(*ceteris paribus* clauses) which we are not fully able to make precise at the moment. For instance, *cetera* could be not *paria* in that, when confronted with the new word, the speaker suffers a certain impairing brain condition. That would provide for the consequent of the subjunctive conditionals to be falsified, even though the antecedent is true; but because we take the circumstance to have been tacitly excluded by the *ceteris paribus* clause, we understand that it does not falsify the subjunctive statement or the theory entailing it. The more precise the background theory regarding the complexity in question, the less indeterminate the hedges need be. All of this is also the case regarding similar statements in any other contemporary explanatory undertaking. Thus, barring a general skepticism about science, there is nothing to stop us from thinking that counterfactuals such as those implied by considering a type of behaviour systematic or unsystematic can be empirically confirmed or disconfirmed. Indeed, what makes it the case that the examples given before *are* examples of systematic behaviours is precisely the fact that the relevant subjunctive statements (the ones entailed by relatively vague theories of the intentional complexity to be assumed) are confirmed by experience.

With this as background motivation, let us turn now to the task of defining systematicity. Systematicity is, I shall assume, a property of intentionally determined properties (like *being grammatical in Spanish*—or maybe *being grammatical in the Spanish idiolect of S*—, *being a Spanish word*, and *being a Spanish sequence of sounds*) or relations (like *sentence  $\sigma$  expressing in Spanish proposition  $\pi$*  and *word  $\alpha$  expressing in Spanish concept  $\chi$* ). An *intentionally determined* property or, more generally, *n*-adic relation (an *ID*-relation, for short) is one such that there is an intentional explanation for the fact that it holds among the items in an arbitrary *n*-adic sequence.<sup>6</sup> I shall assume that there is at least one such true explanation. (Perhaps there is a point in having a more general notion of systematicity, but for our purposes it is adequate and convenient to constrain it to intentionally determined properties.) This is then the definition: A *n*-adic *ID*-relation *R* is *systematic* if and only if an intentional explanation for any *n*-adic sequence *r* in the extension of *R*

<sup>6</sup> Given the assumption that behaviours are events with intentional explanations, being a disposition to behave so-and-so is an *ID*-property. It is perhaps more customary to define systematicity directly as a feature of internal mechanisms; this is so in Fodor & Pylyshyn, 1988. My definition obtains indirectly the same result, for it implies that, whenever a property is systematic, there are systematic relations (in F&P's sense) holding among intentional states that explain instances of the property. The motivation for my more roundabout route was given before: I try not to beg any question which people with behaviouristic proclivities might not grant. (I rely on intuitive understandings of the crucial notions invoked to introduce intentional explanations—particularly that of 'representational' or 'informational' content—but I think that they can be explicitly explained.)

involves other ID-relations holding among parts of  $r$ 's relata in such a way that  $R$  holds of at least some other such sequence.<sup>7</sup>

There is systematicity, then, only when a particular kind of explanation exists for membership in the extension of a given property; i.e. only when there exists a theory accounting for the systematic property at stake. I shall refer henceforth to this theory as *the system-theory*. A claim of systematicity can therefore be known with certainty only when such an explanation is fully known; and the hypothesis that such a claim is correct is in order only when the explanation is at least surmised to exist. (I shall also refer to the theory of the surmised system-theory as 'the system-theory'; no confusion should arise if the context is fully taken into account.) As I said before, we have good reasons for sustaining the hypothesis in the different instances I have mentioned previously, even if we cannot claim certainty. What is controversial is the further, anti-behaviouristic claim that *at most one* of two system-theories making the same predictions about the systematic behaviours under investigation can be true; and not just on the basis of 'simplicity' considerations. (It is precisely this that Quine denies in Quine, 1972.) That is to say, it is a realist, anti-instrumentalist interpretation of the system-theory what behaviourists find unacceptable.

The argument for the 'syntactic image' (or for the *hypothesis of the language of thought*—abbreviated LOT henceforth—as it is more generally known) starts from the assumption that there are systematic properties. Because there are systematic properties, there are properties membership in whose extension is explained by positing complex intentional abilities. Which complex abilities in particular are posited depends on the specific theory of the systematic behaviour involved: as we just saw, claiming systematicity entails the existence of such a system-theory. But, whatever the theory, the complex intentional abilities will have two essential properties: first, because the abilities are *complex*, several (two or more) of the abilities explaining membership in the systematic property's extension for *different* items will involve *the same* simpler abilities *in the same way*; second, because the abilities are *intentional* and their complexity is *intentional* complexity, they will involve, as their simpler parts, *representational states*.<sup>8</sup> For instance, the explanation for the

<sup>7</sup> A  $n$ -adic ID-relation  $R$  is *productive* if and only if an intentional explanations for some  $n$ -adic sequences in its extension involves other ID-relations holding among parts of the sequences' elements in such a way that  $R$  holds for an infinite number of  $n$ -sequences.

<sup>8</sup> Evans, 1981, sec. III claims that the intentional states involved in the explanations of the systematic behaviours I am considering here (the states constituting our 'tacit knowledge' of the phonological, morphological, syntactic and semantic properties of our mother tongue's expressions) differ in important respects from paradigm cases of intentional states—beliefs or desires which we are aware of having, or could be easily aware of having. The latter, but not the former, can be put 'at the service of many different projects'. The paper also suggests that the states constituting linguistic competence are not 'really' intentional states—although it does not explicitly say so. The claim is clearly correct; the suggestion is at odds with the practice of linguists and psychologists, who describe as a matter of course the relevant states in intentional terms. In my view, the psychologists are right: a correct account of the nature of intentional states applying to



pairing of different sentences with their meanings will involve the pairing of the same lexical unit with its meaning (this exemplifies the first essential property); and this simpler part common to different complex abilities will itself be a representational state (this exemplifies the second). The abilities explaining that several different sound patterns are Spanish—to mention a second case—will involve, say, the recognition of the same articulatory traits (first essential property); and, these simpler abilities being *recognitions*, they are themselves to be representational states (second). Representational states are postulated (at least, according to the present reconstruction of this area of our conceptual practices) both in the rough folk theories of systematic behaviours and in the more detailed philosophical or scientific theories.<sup>9</sup>

Relative to the details of the system-theory, the whole process constituting a complex ability might be modelled as a rule ('simpler sounds with such-and-such articulatory traits can be followed/are to be followed in Spanish by sounds with such-and-such traits', etc.). What is explained by one of these complex abilities (any instance of a systematic type of behaviour), thus, can be traced back to something that might be described as *knowledge of a rule*. However, as has frequently been pointed out, nothing follows from this for our purposes—because such a 'knowledge' could well be 'hardwired' into the system, without involving any 'psychologically real' representation of the applied rule.<sup>10</sup> On the other hand, as we just saw, in the most plausible accounts which we can give of systematic properties—both scientific and folk, I must insist—the parts of the complex intentional abilities are always constituted by intentional relations to *representations*—items with informational content which reappear, playing the same explanatory role, in different complex abilities explaining different instances of the systematic property at stake: representations of the articulatory traits of sounds, in the phonological case, or of the lexical units and their modes of combination, in the morphological case, or of the meanings of the individual words and subsentential expressions in the semantic case.

So, to summarize what has been argued so far: there are systematic properties; and their existence entails the existence of complex intentional abilities, which share simpler intentional abilities involving relations to representations. Again, these simpler intentional abilities (involving relations to representations) are represented (to insist one more time) both by the simpler concepts that, intuitively (folk-psychologically), are explanatorily required

---

paradigm states would show that there are enough analogies between them and the states involved in tacit knowledge to validate their practice. But I cannot do more here than state the view.

<sup>9</sup> See Sebastián (forthcoming) for such an example of how scientific psychology is committed to systematic properties. According to the research in speech-perception summarized in this work, there are rules articulating basic sound-units into (possible) Spanish syllables and syllables into (possible) Spanish word-size sounds. The ability to perceive a sound as a Spanish word involves a representation whose informational content is the syllabic structure of the sound.

<sup>10</sup> By Fodor, among others; see Fodor, 1987, pp. 21–6.

for having complex thoughts, and also by the theoretical posits of sophisticated scientific theories in fields like perception of language. Now, although, as we just mentioned, no causal-explanatory role need be given to the representation of the 'rule' into which, given the system-theory, any complex ability could be translated, an essential causal-explanatory role is certainly given to the representations themselves. For, as we saw, the difference between systematic and unsystematic behaviour-types lies precisely in the counterfactuals and other subjunctive statements which are true in the former case and false in the latter. But the entailment of subjunctive statements indicates the presence of properties with a causal-explanatory role; and the differences in the entailed subjunctive statements just mentioned are predicated, in the end, on the fact that the abilities explaining systematic behaviours share intentional states while the abilities explaining unsystematic behaviours do not. In fact, according to the functionalistic picture of mental properties, the intentional states themselves and the involved representations are *individuated* by their causal-explanatory roles: a representation of a certain articulatory trait is identified precisely by its causal role, among other things, in the recognition of certain sound-patterns.

Cartesians interpret this causal-explanatory role given to representational states as involving *exactly the same commitments* that causal-explanatory states are generally thought to have outside psychology. (This tenet of Cartesianism is usually referred to as 'Intentional Realism', and marks its crucial difference from the instrumentalistic attitude towards intentional states taken by behaviourists like Quine.) In addition, Cartesians think that a crucial such commitment, to be found generally in science, is the one philosophers have been trying to state precisely under the label of 'Physicalism'. This is the requirement (which I shall present in a manner as neutral as possible among different metaphysical stances) that a certain *explanatory* relation must exist between the causal-explanatory posits of high-level theories of the macroscopic world, and the ones at lower-levels—ultimately, at the utopian most basic microphysical level—for the former to be acceptable, or taken as 'real'. (The differences in the metaphysical pictures, on which subject I shall try to keep neutral, concern the nature of the relation: identity, or some variety of supervenience.) So, for instance, clouds are macroscopic entities, and the property of *being a cloud* is related to certain causal-explanatory patterns (provoking rain and so on). Now, this property is acceptable to the extent that *being a cloud* is correlated (again, trying to be neutral on the metaphysical issues) with that of being a certain kind of physical entity, in such a way that having these physical traits, plus the physical laws, *explains* that causal profile associated with *being a cloud*. (This is what, in the end, distinguishes *being a cloud* from *being a witch* or *being born under Leo*, according to scientific realism.) Exactly the same should be true, therefore, of the representations that play an essential causal-explanatory role in accounting for systematicity. There should be correlated properties described independently, in lower-level terms, such that possession of the lower-level character explained the systematic behaviour that the representation is pos-

ited to account for. In particular, the lower-level correlates of the representations involved in the explanations of systematic properties should account for the properties essential to them which we highlighted earlier: namely, that they constitute complex abilities which explain instances of a systematic property (they are 'parts'), and that they appear in the explanation of different instances of the systematic property playing the same causal-explanatory role in all of them (they are 'movable'). Under the further assumption of functionalism, this will also account for their having the content they have, for, as we indicated previously, *having informational content* is, under the assumption, a property functionally individuated by having a certain causal-explanatory role.

This requirement might seem a heavy burden. However, the defenders of the syntactic image understand that the conceptual idea of a von Neumann-Turing machine or digital computer, together with its physical implementation, already indicate how it can be satisfied—and is in fact satisfied in practice. For in these machines we can certainly find the lower-level correlates of the representations that satisfy the requirements of the previous paragraph.<sup>11</sup> On the basis of an obvious analogy with expressions of natural languages, the properties of such correlates allow us to call those correlates *symbols*: they are entities with nonsemantic (physical) descriptions, which have semantic interpretations, and contribute, in a semantically systematic way (given by the system-theory), to the semantic interpretations of more complex entities of which they are non-semantic (physical) parts. They have therefore the essential traits of symbols composing natural languages; it is for this reason that it makes sense to think of the system of representation-correlates accounting for an intentional complex ability as a *language*, indeed, as the *language of thought*.

As is well known, Fodor postulates LOT as an empirical chain, the result of an inference to the best explanation: it is just that the von Neumann-Turing picture is 'the only game in town', the only account we have at the moment which promises to succeed in explaining systematic properties. I myself see the issue in a slightly more aprioristic (although still empirical) way—as I shall try to explain in the following section.

## 2. *Inter-level Explanatory Relations and LOT*

The skepticism motivated by connectionism, as I said in the introduction, comes in two varieties. There is that of eliminativists—those who share with

---

<sup>11</sup> I would like to emphasize that the point made here is rather weak: it is just that digital computers indicate how the Classical account of systematicity could actually be correct. In particular, it should not be taken as implying that Classicism is committed to the view that the mind's functional architecture is specifically that of a von Neumann machine, or that of a Turing machine. (Fodor & Pylyshyn, 1988 make the same point; see also McLaughlin, 1993, pp. 182–3, and McLaughlin & Warfield, 1994, pp. 378–9.) I develop further the view that the constitutive claim of Classicism is rather more abstract

Cartesians the view that there is no principled difference between the categories of folk-psychology and those of science in general, but think that connectionism shows them to be at fault in that respect;<sup>12</sup> and there is the skepticism of those who think that folk-psychological categories are protected, almost a priori, from any such failure—for whom connectionism shows why it is misguided to think of them in the same way we think of other scientific categories. For skeptics in the second class, connectionism pinpoints the conceptual confusion of both Cartesians and eliminativists in fully assimilating intentional properties to natural properties. Clark's skepticism belongs in the second group—although, as I shall try to show in this section, there are reasons for not being satisfied with the clarity of those of his views in this regard put forth in the book.

Skepticism of the second variety gets some support from the outrage we tend to feel in the face of the possibility contemplated (and in fact defended) by eliminativists. Intentional properties play too important a role in our lives for them to fall victim to the fates of *phlogiston* and *impetus*. (That there are intentional properties seems to be presupposed by the very possibility of giving arguments for eliminativism!) However, as critics of eliminativism have frequently pointed out, we should not be misled by this comparison: *phlogiston* and *impetus* are very bad analogies indeed to illustrate the fate that could await folk-psychology, given the Cartesian view. The real commitments of Cartesians (the ones whose attainability by folk psychology eliminativists should really attempt to expose) are better put in the following way: folk psychology, and its categories, should stand to scientific psychology in more or less the same relation as naive physics (the commonplace knowledge of the physical world that allows us to navigate our way through it) stands to scientific physics. Here the conceptual relations are much more difficult to state in any simple way. On the one hand, the properties and laws of naive physics (say, that it takes more effort to raise heavier bodies than lighter ones, that it takes more force to stop the faster of two moving bodies that weigh the same, and so on) are explained by (and therefore, are *ontologically* dependent on) those of scientific physics—which opens up the possibility that discoveries regarding the latter might force corrections, maybe even deep corrections, on the former. On the other hand, scientific physics is *epistemically* dependent on naive physics, because the properties and laws of the former are known to us only to the extent that they are *empirically* confirmed; and they are empirically confirmed partially in virtue of their explanatory connections with the properties and laws of the latter—which seems to forbid any full-scale correction of naive physics by scientific physics, so as not to cut the branch on which we sit. This is not the place to examine whether or not there is an insoluble form of paradox here (it is

---

than it is taken to be by some friends of connectionism in a related paper, García-Carpintero, in press.

<sup>12</sup> Notoriously, the Churchlands. See, for instance, Churchland, 1989.

my view that there is not); it has been mentioned only to point to a more accurate measure with respect to which to determine the true strength of the threat posed by Cartesian views to folk psychology.

Clark's own alternative to Cartesianism is unclear at the crucial juncture. He says that he has abandoned the (*AE*, p. 223):

overriding commitment to finding scientific analogues (albeit at some high level of description) to the folk solids so as to allow their reductive identification with straightforwardly causally potent inner states. Explanatory goodness need not be tied to such straightforward causal potency. Instead we may adopt a macrostrategy [...] in which explanatory goodness is defeasibly [...] established by gross behavioral patterns (actual and counterfactual. [...]) This position (which owes more than a little to Gilbert Ryle) falls short of out-and-out behaviorism in its being coupled with a desire to tell rich and illuminating stories involving a variety of inner representational states and in its explicit acknowledgement of the falsifiability of folk accounts by general discoveries concerning underlying cognitive architectures.

The idea of accounting for explanatory goodness by adopting a 'macrostrategy' is explained earlier (*AE*, pp. 208–14), by contrasting it with the adoption of a 'microstrategy' (the distinction is attributed to unpublished work by David Ruben). In the second case, the correctness of a macroscopic explanation is tied to linking the macro-properties and laws to lower-level properties and laws; the requirements we tied to Cartesianism in the preceding section are, I suppose, a case in point. According to the first view, on the other hand, it just depends on the truth or falsity of counterfactuals and other subjunctive statements stated at the macroscopic level itself. Thus, to exemplify the picture (without committing Clark to the specifics of the example, which is mine), the 'gross behavioural patterns' in what I called earlier *systematic behaviours* motivate us to postulate complex intentional abilities, involving representational states. Now, for the abilities to be deemed as real and explanatory, the truth of the subjunctive statements that positing them, as we saw, entails is enough. It is not additionally necessary that they be 'mirrored' at the lower level of the language of thought, as Cartesians require (a requirement that they feel justified in taking as confirmed by the explanatory success of the von Neumann-Turing picture, which according to Clark has been brought into disrepute by connectionism). This is essentially Quine's view on the status of grammars in the paper we referred to in the preceding section.

Clark's is, it must be readily acknowledged, an appealing picture. It pleases our sense that eliminativism is not even a possibility (or is at least one with a very low degree of probability), without forcing us to attempt a clarification regarding the precise nature of the links between macroscopic

and microscopic properties and laws—the difficulty of which might be gathered from what we indicated previously. However, there are deep-seated difficulties with this picture—difficulties which Cartesians have mentioned all along. I would have liked Clark to give some indication of his views regarding the most obvious difficulty, which I shall point out now.

The problem, bluntly put, concerns the truth-conditions of the same-level counterfactuals on which Clark relies for distinguishing correct from incorrect explanatory claims. Clark's patron saint Ryle held the view that they encode 'inference-tickets': licences to infer observational circumstances from observational circumstances. The relevant counterfactuals merely codify regularities in behaviour, relative to observable circumstances. This view (constitutive, of course, of a version of 'logical behaviourism') expresses a deep anti-realist, in fact instrumentalist, interpretation of such mind-innards as the posits of the theories of systematic behaviours—folk or scientific—which I mentioned previously. It is generally thought to have been refuted; but of course philosophical views generally thought to have been refuted are resuscitated every now and then, sometimes in all fairness. In any case, it would be good to know if this is indeed Clark's view of the matter, and if so, some of his answers to the questions incorporating the standard objections: is this intended as a general view of non-basic scientific theories? The positive answer (irrealism regarding the theoretical entities posited by non-basic scientific theories, that is to say, *every* scientific theory with which we are nowadays familiar) seems highly implausible; but, if the answer is negative, why should an exception be made in the psychological case? And so on.

However, Clark does not seem quite to take Ryle's view (although their close affinity, which he himself acknowledges, will be made transparent as we proceed with some of the criticism to be advanced in the final section). For he says, in the quotation just provided, that he is prepared to contemplate the 'falsifiability of folk accounts by general discoveries concerning underlying cognitive architectures'. Thus, it does not seem that mere regularities in behaviour are enough, according to him, to validate the explanatory nature of folk-psychological claims. (This, he claims, differentiates his ideas from those of Dennett, generally thought to be the closest of contemporary philosophers to a behaviouristic stance; see *AE*, pp. 214–6.) It seems therefore that folk-psychology involves, according to him, some commitment about the non-intentional nature of the causes of behaviour. It must be pointed out, alas, that what Clark has to say about such commitments is not altogether clear—to this reviewer at least. There is a 'Requirement of Normative Depth', according to which (*AE*, p. 216) '([t]he inner workings of an intentional system must be of a kind compatible with the description of that system as capable of making mistakes which involve the failure to respect those commitments in episodes of on-line processing.' As far as I can see, the workings of *anything at all* are compatible with some 'description' such that there are 'episodes of on-line processing'—this simply means, I take it, *particular occurrences in the item's lifetime*—that can be counted as failures 'to respect' the 'commitments' specified in the description. Some precise con-

straints should be placed on 'description' and 'processing' for this not to be the case; but Clark's text says nothing of their nature. The second requirement is that the innards of a true believer should give rise to conscious states (AE, p. 217). There is, of course, no need to belabour the darkness in which we are at present regarding the intended meaning of 'consciousness'. Is it the physical basis for qualitative states that is required? Or enough complexity for having 'thoughts about thoughts'? Clark does not say. But in any case, Clark's commitment to the falsifiability of folk-psychological claims suggests a more 'full-blooded' view of the abilities or dispositions posited by those claims than true behaviourists are prepared to grant. This only makes more urgent the need to be clearer about the content of those same-level counterfactuals to which he gestures to state his anti-Cartesian views on the metaphysics of the mind.

Gareth Evans (from whom I have taken the term 'full-blooded') gave some time ago an a priori argument in favour of the structured character of thought (Evans, 1982, pp. 100–105). He contended, however, that the argument, by itself, did not lead to LOT (1982, pp. 100–101):

I certainly do not wish to be committed to the idea that having thoughts involves the subject's using, manipulating, or apprehending *symbols*—which would be entities with non-semantic as well as semantic properties, so that the idea I am trying to explain would amount to the idea that different episodes of thinking can involve the same symbols, identified by their semantic and non-semantic properties. I should prefer to explain the sense in which thoughts are structured, not in terms of their being composed of several distinct *elements*, but in terms of their being a complex of the exercise of several distinct conceptual *abilities*. [...] we can shed some light on what it means to see a thought as the result of a complex of abilities by appealing to what is meant when we say that the understanding of a sentence is the result of a complex of abilities. When we say that a subject's understanding of a sentence, 'Fa', is the result of two abilities [...] we commit ourselves to certain predictions as to which other sentences the subject will be able to understand; furthermore, we commit ourselves to there being a common, though partial, explanation of his understanding of several different sentences.

In this passage, Evans seems to be contemplating a view like the one put forward by Clark (who in fact refers approvingly to it: see AE, pp. 198–206). Thoughts are *essentially* structured; thoughts are essentially involved in the production of systematic behaviour. Such structure as they have, however, is to be constructed as a structure of (intentional) abilities or dispositions, without any commitment to its being replicated at other, lower levels. This notwithstanding, the system of complex abilities can be taken as fully-fledged explanatory.

In a previous paper, however, Evans appears to commit himself to a more Cartesian view. There he explains in what having *tacit knowledge* of a *structured* semantic theory ( $T_2$ ) consists—as opposed to having tacit knowledge of an unstructured, list-like semantic theory ( $T_1$ ). He also resorts for that purpose to complex abilities or dispositions; but he says (Evans, 1981, 329–31):

Now, it is essential that the notion of a disposition used in these formulations be understood in a full-blooded sense. These statements of a tacit knowledge must not be regarded as simple statements of regularity, for if they were, anyone who correctly judged the meanings of complete sentences would have a tacit knowledge of  $T_2$ . When we ascribe to something the disposition to V in circumstances C, we are claiming that there is a state S which, when taken together with C, provides a causal explanation of all the episodes of the subject's V-ing (in C). So we make the claim that there is a common explanation to all those episodes of V-ing. Understood in this way, the ascription of tacit knowledge of  $T_2$  does not merely report upon the regularity in the way in which the subject reacts to sentences containing a given expression (for this regularity can be observed in the linguistic behaviour of someone for whom the sentence is unstructured). It involves the claim that there is a single state of the subject which figures in a causal explanation of why he reacts in this regular way to all the sentences containing the expression. [. . .] The decisive way to decide which model [the structured or the unstructured, M. G-C] is correct is by providing a causal, presumably neurophysiologically based, explanation of comprehension. With such an explanation in hand, we can simply see whether or not there is an appeal to a common state or structure in the explanation of the subject's comprehension of each of the sentences containing the proper name *a*.

In this passage, therefore, Evans does seem committed to the idea that the 'full-bloodedness' of the explanatory complex abilities amounts 'to the idea that different episodes of thinking can involve the same symbols, identified by their semantic and non-semantic properties'—as he put the idea of a language of thought in the text quoted first. For the 'decisive way to decide' regarding the correctness of the explanatory attribution of a system of complex dispositions involves finding a *non-semantically individuated* 'common state or structure' corresponding to the semantically individuated common elements in complex abilities: in other words, finding a system of symbols.

However, I do not think that there is any real inconsistency between the two passages. In the first, Evans is contending that the structured character of thoughts can be sustained by a priori argument—that it follows, from the very concept of *having thoughts*, that thoughts are structured. Understandably, he does not want such a structured character, allegedly known a



priori, to entail the existence of a non-semantically individuated mirror structure. For the existence of such a mirror structure is certainly not known a priori (assuming that it is indeed known to exist); but if it were entailed merely by the structured character of thought, the latter could not possibly be an essential property of thoughts known a priori. In the second passage, on the other hand, Evans is providing an analysis of the notion of *tacit knowledge*. It is legitimate, for that purpose, to have recourse to the *methodological hypothesis* that any truly explanatory macro-property (and the dispositions or functions with respect to which it is individuated) is to be explained by lower-level characters and the laws in which they play a role. This is, as I said previously, the core of the idea of *Physicalism*. It would be extremely implausible to claim that this methodological hypothesis is an a priori truth; we do not even know if it is a truth at all, and whatever partial confirmation we have for it is empirical (the success of the hypothesis in guiding the research leading to the discovery of chemical correlates for genes, and so on). In any case, Cartesians claim, granting its status as a methodological guide, the notion should still constrain psychology with the very same strength that it constrains any other 'special science'. And Cartesians of Fodorian persuasion feel reassured in believing that it does apply to psychology as much as it applies to geology or biology by the undoubtedly meagre, but no less real, empirical successes of the von Neumann-Turing model of cognitive processes.<sup>13</sup>

Clark thinks that connectionism renders such confidence baseless. As I

---

<sup>13</sup> Davies, 1991 contains an a priori argument for LOT, based on Evans' defence of the necessarily structured character of thought. I think Davies fails to appreciate that any argument for LOT requires as a premise the thesis of Physicalism, which, in whatever form, is a posteriori. (The specific argument for LOT only requires that the components of complex representations posited by what I called the system-theory have a non-semantic, lower-level description—say, a computational or neurological one. However, this requirement is more illuminatingly understood as deriving from a general constraint applying to every explanatory category posited by special sciences. It should therefore apply also to whatever non-semantic properties with respect to which the symbols are individuated; in this way, they should ultimately have physical specifications. This interpretation of the requirement is more illuminating because, in my view, the truly characteristic mark of Cartesians is the view that psychology is not relevantly different from other special sciences). The general principle of Physicalism (the precise formulation of which I am purposefully avoiding) constitutes in my view a justified belief, but justified only a posteriori (on the basis of previous successes in biology, etc.); and the contention that it applies also to psychology is, if any, only a sensible conjecture, justified equally a posteriori (on the basis of the Fodorian considerations). A crucial passage in Davies' argument (pp. 238–9) assumes in my view Physicalism, without explicitly acknowledging it. Moreover, I think that Evans commits himself to LOT, and to some version of Physicalism, only in the second passage I quoted, but self-consciously puts it in abeyance in the first—while Davies refers indistinctly to both texts, and goes on to offer an interpretation of the first passage which renders it consistent with his own view (Davies, 1991, pp. 244–45). To me it is clear that Evans is contending in that passage that the a priori argument for his 'Generality Constraint' is not an argument (which would then be a priori) for LOT; Davies' interpretation is in my view too strained.

shall make clear in the next section, his arguments do not look to me so convincing. In any case, that is, I suppose, the reason for his abandonment of a commitment to the 'full-blooded' character of complex intentional abilities, and for his search for a middle ground between traditional behaviourism and the more stringent requirements of Cartesianism. In this section I have given reasons to think that the conceptual foundations for such a middle terrain are left almost unexplored in his book.

### 3. A Distinct Connectionist Account of Systematicity?

The Cartesian picture is committed to LOT, as I have explained. Moreover, it gets some support from the empirical research currently undertaken by cognitive scientists: successfully modelling cognitive processes as Classical programs gives some credence to the view that the Cartesian commitments might be factually met. However, the Cartesian picture is not committed to the specifics of von Neumann-Turing programs. There could be models of systematic cognitive processes differing in important details from von Neumann-Turing programs while still agreeing with the fundamentals of the Cartesian view, i.e. while still positing a 'language of thought'. Moreover, those possible alternative models might be such that, even if in some very abstract sense they could be seen as 'implementations' of Classical programs, the differences with von Neumann-Turing programs could be relevant to account for facts recognizably belonging to the cognitive domain. (Clark makes this point forcibly; *AE*, p. 61.)

For all we know, connectionist programs might well be such models. There seem to be good reasons to hold this 'compatibilist' attitude. On the one hand, connectionist models can undoubtedly claim some successes, in fields where, even if close to very basic perceptual abilities, some systematicity seems to be at stake.<sup>14</sup> On the other hand, connectionist researchers have great difficulties in successfully modelling the sort of systematicity that Classical programs are paradigmatically successful at reproducing and explaining.<sup>15</sup> By this I mean systematic behaviours produced by someone who has consciously learned explicit rules constituting a system-theory for a systematic property; for instance, someone who has learned a fragment of the grammar of a natural language, or—better still—for a wholly artificial

---

<sup>14</sup> This may be exemplified by the much-discussed example of NETtalk; the sort of systematicity at stake in that case is one akin to the one in the examples given in footnote 4 before.

<sup>15</sup> I have consciously said 'in successfully *modelling*' instead of 'in successfully *explaining*' because connectionist researchers seem at times to be satisfied, by way of explanation, with developing networks which, after the training period, behave in a systematic way close to the one exhibited by the targets they had selected for explanation. This is, no doubt, a reflection of the strong behaviouristic tendencies by which researchers in the field are swayed. I shall comment below on some remarks that Clark makes in this regard, which in my view confuse heuristics with explanation.

language, by being explicitly given syntactic and/or semantic rules. In between lie the truly fascinating cases, namely, those—like our knowledge of the grammar and semantics of our mother tongue, or our ability to think rationally—(i) that give rise to properties which are, beyond reasonable doubt (as explained in the first section), systematic, (ii) whose systematicity can be represented, to a certain extent, as if they were the result of having acquired and internally represented a system of explicit rules, but (iii) that can be accurately represented in this way only *to a certain extent*, for there are both conceptual and empirical reasons why these abilities cannot literally consist in an internal replica of a system of explicit rules.<sup>16</sup> In my view, no paradigm (Classical or connectionist) can claim, at the moment, to account for these intermediate and decisive cases. They both seem to do well (better than the opponent at least) with some of the properties of the interesting cases, while faring poorly with others. The conceptually most interesting aspects of research in connectionism have to do—to my mind at least—with the clarification of these issues. In other words, they have to do with the clarification of a concept crucial for any philosophical research, namely, that of our tacit knowledge of the grammar and semantics of our natural languages and of the resources for rational inference. (A compatibilist approach like the one suggested here is forcefully defended in McLaughlin, 1993, sec. 5, as ‘implementation connectionism’.)

This is not what concerns Clark in this book. Against the ‘compatibilist’ open-minded view suggested in the previous paragraph, he develops an argument, based on what is currently known regarding connectionist networks, against the ‘Syntactic Image’. To advance this anti-Cartesian view, Clark’s argumentative strategy has two parts. He argues that connectionist networks, in those cases where they do seem to produce systematic output, definitely falsify LOT. Then he suggests, in a more speculative vein, that the anti-Cartesian lesson allegedly learnt from these cases can be extended to those others with which, at this stage, networks do not seem able to cope. (He in effect suggests here that it could be only an illusion to think that LOT explains systematicity, even in those cases I deemed paradigmatic earlier.) In this section, I shall review his arguments in reverse order. I shall begin by indicating why his considerations about the cases that are relatively well accounted for by the Classical paradigm and with which connectionism fares poorly are deeply unsatisfactory. Then I shall examine the reasons why, according to him, the lesson from the cases where connectionism can be

---

<sup>16</sup> The conceptual reasons have to do with a famous argument which can be gleaned from Lewis Carroll’s ‘What the Tortoise Said to Achilles’ (*Mind*, 4, 1895, pp. 268–70). The empirical reasons have to do with the excessive stress that should otherwise be put on the competence/performance distinction—to explain the mismatch between the real abilities of speakers to understand and recognize as grammatical sentences of their natural language, or to perform rational inferences, and what should be expected from them were their implicit knowledge correctly represented as the internal representation of the relevant explicit rules.

reasonably considered successful is that, no matter how reluctantly, LOT should be abandoned.

Clark's proposal for an alternative connectionist account of manifestly systematic properties (and his criticisms of Cartesianism in this regard) arises in my view from a basic confusion between *heuristics* and *explanation*. Let me expand on this point. David Marr (1982, ch. 1) gave a popular philosophical model for cognitive research—which fits well the Cartesian expectations—by identifying three explanatory 'levels'. Let us review it with respect to some systematic property, such as *being a Spanish word* or *being a Spanish sentence*. There are cognitive facts in virtue of which these are systematic properties, facts regarding the nature of the competence of competent speakers of Spanish. Marr's model concerns this domain of cognitive facts. There is *level 1* (we can ignore Marr's other designation, which may be misleading), at which intentional characterizations are given. That is to say, level 1 is constituted by a system-theory accounting for the observed systematicity ('accounting for' in the sense that the system-theory entails subjunctive statements confirmed by the observable behaviour of the subjects); i.e. for how competent Spanish speakers are able to project their knowledge of the morphology or the syntax of Spanish onto words or sentences they had not encountered previously. There is *level 2*, at which algorithms implementing the system-theory are given. And there is *level 3*, at which neurological realizations are provided for the algorithms at level 2. Because the 'implementing' relation between the algorithms and the system-theory, and the 'realizing' relation between the neurological description and the algorithmic one are *explanatory relations*, a system of *symbols* is contemplated both at levels 2 and 3—in the sense of 'symbol' previously explained: entities with non-semantic descriptions (that is why they are at levels 2 and 3), which contribute to the formation of different complex structures (whose *movable parts* they are, then), in mirroring correspondence with the articulation indicated by a semantic description. Finally, a criterion for the distinction between levels is the possibility of 'multiple realization': the same system-theory of level 1 might be implemented by different algorithms, each one of which might in turn be realized in different neurological structures (although, of course, in any given particular case only one algorithm will implement the system-theory, and only one neurological structure will realize it). Needless to say, many more layers of levels could be contemplated, but Marr saw no conceptually interesting point in doing so.<sup>17</sup>

<sup>17</sup> Under the assumption that the intentional theory at level 1 is taken 'extensionally', Peacocke, 1986 justifies the need to introduce a 'level 1.5'—still semantical, but *intermediate* between Marr's semantic level 1 and his algorithmic level 2 on account of 'multiple realizability' considerations. The 'extensionality' assumption is the assumption that the sole business of theories at level 1 is to determine the extension of the property under discussion: which words are to be identified as Spanish, or which strings of words as grammatical Spanish sentences; *how* the theory gets that result (i.e. which structural description, if any, it assigns to the sentences or the words) is immaterial to constrain the relations between levels. Under this assumption, Peacocke is undoubtedly

Now, you can fruitfully take this model of levels as a heuristic guide: 'begin by constructing a level 1-theory of the domain (or by refining a preexisting folk one), look for algorithms that can implement it, and pay attention all the time to the fact that the system must have a neurological realization'. Undoubtedly, most classical research on cognition works that way; and some of it can claim astounding successes.<sup>18</sup> But providing a heuristic guide is not the point of the model. The model is a model of *acceptable explanation* in the cognitive domain. Marr's point can be summarized in this way: a cognitive theory (restricting ourselves to theories of systematic properties, the truly interesting ones after all) provides explanatory information only to the extent that it gives an algorithm which implements a system-theory of the domain which has a neurological realization. (The point has been made regarding systematic properties, but, of course, the same can be said regarding less interesting theories of unsystematic properties—which are always presupposed by system-theories, as we already said.) It is clear that for this model to be correct, the heuristics leading to its successful application is wholly immaterial. One can start, for instance, by blindly training a network, changing the starting weights and the 'learning' algorithm, and so on (or even by reproducing some neurological properties in a network), until one gets something that accurately reproduces the behaviour of humans in such a way that similar subjunctive statements are confirmed by its behaviour, but regarding the structure of whose innards one is as much in the dark as one was regarding humans; and only after analysis of the network (maybe beginning even at electronic level 3, selectively damaging it to see what happens) be able to come up with an accurate level 1 theory. Again, the point is that *only then* can we claim to have an explanatory understanding of the field. Seen in this light, of course, the model is in perfect agreement with the main contentions of Cartesianism: indeed, it is another way of putting them.

Clark's failure to notice this distinction in his discussion of these issues (ch. 3) is manifest in the following quote (*AE*, p. 60):

Unlike the classical, Marr-inspired theorists, the connectionist does not begin with a well-worked-out (sentential, symbolic) competence theory and then give it algorithmic flesh. Instead, she begins at level 0.5 [this has been described previously (pp. 51–2) as a specification

---

right to require that level 1.5—that at which the 'information on which the algorithms draw' is also specified. In the case of cognitive theories of systematic properties (and they are of course the cognitive theories we are really interested in), Marr's philosophical aims are only to be fulfilled if there is a semantic level putting more constraints than merely 'extensional' ones; in effect, if there is a level constituted by what we have been calling a system-theory. Thus, Peacocke's point is well taken. It is only that, under the explicit assumptions of this paper, level 1 already fulfils that task: it is *not* given 'extensionally'.

<sup>18</sup> See the research summarized in Konishi, 1988 for a good example. (I owe my knowledge of this research to Nuria Sebastián.)

of 'the set of weights capable of mediating the desired state transitions', 'the basis upon which . . . the system comes to be able (after much training) to negotiate the targeted cognitive terrain', M. G-C.], trains a network, and then seeks to grasp the high-level principles it has come to embody. This is a great boon for cognitive science, which has been dogged by the related evils of ad-hocery and sentimentalism. . . . Connectionist methodology, by contrast, allows the task demands to trace themselves and thus suggest the shape of a space in a way uncontaminated by the demands of standard symbolic formulation. We thus avoid imposing the form of our conscious, sentential thought on our model of cognitive processing—an imposition which was generally as unsuccessful in practice as it was evolutionarily bizarre.

Clark may be right here about the advantages of connectionist *heuristics*; but that does not affect the point at stake, for it concerns *explanation* and not *heuristics*. Maybe it is true that if we let our folk-psychological intuitions guide us in designing the system-theory of the domain we will probably be misled, or will trick ourselves into inventing *ad hoc* pseudo-explanations. Be that as it may, the Cartesian point is that we will not have *acceptable explanations* of systematic properties until we weld algorithms which can be plausibly claimed to have neurological realizations to semantical system-theories—no matter how we come up with them. And this point is not refuted by the heuristic point.

This same failure also affects, in my view, Clark's considerations regarding systematicity (pp. 147–9). He summarizes later what he has done as 'noting that systematicity *need not* be traced directly to the classical structure of an underlying architecture. It *might* be the product of some kind of acquired knowledge' (p. 224), although he admits that '[n]othing in my treatment is sufficient to fully exorcise the ghost of full Fodorian systematicity. . . . a "might" is not a proof, and the full puzzle remains undischarged.' (ibid.). But in fact he has not even given a plausible 'might', for the same confusion of heuristics and explanation mars his proposal (*AE*, pp. 148–9):

What I am recommending is, in short, a kind of gestalt flip in our thinking about systematicity. Instead of treating it as a property to be directly induced by a canny choice of basic architecture, it may be fruitful to try treating it as intrinsic to the knowledge we want a system to acquire. For example, we want the system to learn that an open-ended set of individuals and animals, and not just Fred, can fall under the public-language concept 'happy', and that the concept of loving is not the concept of an exclusively one-way relation. We thus treat the space of public-language concepts as just another complex space and ask what we must do to enable a learning system to negotiate it. We end up treating the space of systematically interanimated concepts as just another theoretical space—a

space which may one day be negotiated by a (no doubt highly scaffolded) connectionist learning device. The mature knowledge of such a system will be expressible in terms of a (largely) systematic interwoven set of concepts. But the systematicity will be learned as a feature of the *meanings* of the concepts involved. It will flow not from the shallow closure of a logical system under recombinative rules, but from hard-won knowledge of the nature of the domain. Why settle for anything less?

This discussion concerns how a cognitive state accounting for a systematic property can be *acquired* by artificial devices. The points which are made in that regard may be interesting; in fact, I think they *are* interesting—for reasons which have to do with what I regarded previously as the philosophically most appealing hope related to research in connectionism, namely, that it might deepen our understanding of the problematic cases of *tacit knowledge*. But they do not even touch on the arguments Cartesians offer on the basis of systematicity; for they directly concern not *acquisition*, but only *explanation*. And although explanatory claims entail, as we indicated in the first section, consequences about acquisition (and loss), the consequences about acquisition which Cartesian contentions about systematicity imply might well be in agreement with the 'scaffolded' nature that Clark himself attributes to learning patterns. (Clark thinks otherwise because he seems to take the wildly implausible claim that the learning-process of grammar by children is to be literally identified to the process by means of which linguists construct explicit theories of it as necessarily linked to Cartesianism; see *AE*, pp. 37 and 39. This necessary link, however, should be argued for. The only connection I can see comes from the fact that Cartesians such as Chomsky and Fodor have relied on that implausible view on learning for some contentions about innatism which are very important for them; but those views, as far as I can see, are not a necessary ingredient of Cartesianism.) In any event, no matter how 'scaffolded' the learning path, and the 'meaning' or informational content of the cognitive states constituting the mature competence accounting for systematic properties, the Cartesian point is that the *competence itself* must be 'scaffolded' to *explain* systematicity, and the 'scaffolding' must be reflected at lower levels.

On the basis of his considerations regarding *learning*, *acquisition*, in one word, *heuristics*, Clark strives in fact to make a different point—the one he smuggles in with the suggestion that systematicity might be a feature 'of the *meanings* of the concepts involved'.<sup>19</sup> For the point here is once again a

<sup>19</sup> The same behaviouristic leanings are involved in Clark's recourse to the notion of 'non-conceptual content'—an idea suggested in Evans, 1981, which Clark takes in the way developed in Cussins, 1990. As explained by Cussins, the notion of 'nonconceptual content' is defined relative to the commitments of the theorist attributing them: nonconceptual contents are contents theoretically specified so that there is no commitment by the theorist to the thesis that the subject to whom they are attributed possesses the concepts used in the content-specification (*AE*, p. 73). According to this, a content is

behaviouristic reflection of his own view on explanation—i.e. his inclination towards the strategy he refers to as 'macroexplanation', which we discussed in the preceding section. In other words, on the basis of considerations regarding acquisition he is in fact making a point about explanation: the point that systematicity is perfectly well explained at level 1 (or, as he says, at level 0.5); that it is enough that the subjunctive statements arising from theories at this level are confirmed by empirical evidence, without any need for algorithmic or physical implementations. These subjunctive statements can therefore be taken as 'barely true'; it is not necessary for the macro-level laws to be sustained by facts about lower levels. I already indicated the difficulties of this view in the preceding section. The reader who is not well acquainted with the terms of the philosophical discussion I reviewed there can now see the issues at stake in a more familiar light. The issue is this: should the explanations offered by cognitive scientists satisfy the inter-level links required by Marr's model, or are Clark's much weaker constraints enough? No matter what the answer to this question is, we have seen that the 'alternative' Clark has to offer to Classical accounts of systematicity—in those cases where connectionism has little to say at the moment, while Classical models appear to deal well with the facts—depending as it does on a confusion between heuristics and explanation, is not in the last analysis so much an alternative as a bare dismissal of Cartesian rigours of explanation.

Clark's suggestions for dealing with clear instances of systematic properties are thus shown to stand on a flimsy basis. But perhaps there is no reason for rejoicing in the Classical camp, if, as Clark claims in the less speculative part of the book, those connectionist models that do seem to model cognitive processes where some systematicity appears to be involved can be proven to be incompatible with the 'Syntactic Image'. Let us finally turn therefore to that other part of his argument.

Consider a connectionist network which is able 'to negotiate a cognitive domain' where some systematicity seems to be involved. Let us keep NET-talk in mind as such a network, for the sake of precise intuitions. After training, the network is able to generalize, confirming the predictions of some system-theory: it is able to handle new cases in the way indicated by the system-theory. Now, stripped of all philosophical trappings (obviously necessary to make a nuanced philosophical case, but perhaps distracting at times), the considerations we gave in the second section against the weak-

---

nonconceptual or not only relative to the way it is described. However, Clark soon forgets this and starts speaking of contents which apparently are nonconceptual *in an absolute sense*: '[...] states which are contentful, albeit nonconceptually so.' (AE, p. 75) Relative to what description? The behaviouristic leanings are not too far behind: 'The idea is that some contents properly *consist in* being able to negotiate a certain domain.' If this is the idea, the notion of nonconceptual content should be clearly defined to bear it. For it is uncontroversial that there are interesting cases of nonconceptual content, in the sense defined by Cussins; but it is not so clear that there are according to this clearly behaviouristic notion.



ness of anti-Cartesian views on explanation boils down to rhetorical questions such as these: is it credible that there is no explanation, couched in terms of a lower-level mirroring structure, of why this particular system-theory accurately predicts the behaviour of the network? If there must be such an explanation, should not the truth of the system-theory depend on its existence? Is it credible that, for any possible pair of different system-theories which made the same predictions about observable behaviour, there could be no fact of the matter to determine whether one is truer than the other?<sup>20</sup> What fuels these rhetorical questions, once philosophically conceptualized, is the attitude I labelled 'Physicalism' earlier: the notion that we have found out—certainly, not a priori, but on the basis of the successes of empirical research conducted under its methodological guidance—that questions of these kinds have the expected answers, for any macroscopic domain of investigation. The network is, after all, a physical device; if it is able to generalize on the basis of what looks like a systematic informational theory of the structure of its 'cognitive terrain', this informational theory must have a mirror physical implementation.

Clark's considered view is that, even if something corresponding to the empirically confirmed system-theory can be found at lower levels in connectionist networks, it still cannot amount to a system of symbols. He gives several reasons for this, none of which I find convincing. Two reasons recur: one appeals to the *non-concatenative* or *superpositional* character of the networks' 'symbols'; the second, to their *context-dependence*. Let us consider them in turn; as we will see, when carefully examined the first collapses into the second.

Van Gelder (1990) characterizes two ways of accounting for systematicity, which he labels respectively *concatenative* and *functional compositionality*. Only the first is compatible with LOT, he claims; the second, on the other hand, is exemplified by the sort of superposed distributed representations described in Smolensky (1991). Clark finds the distinction appealing, particularly the alleged incompatibility of complex symbols obtained by superposition with LOT; for he keeps contrasting the superposed structures found in connectionist networks which account for some systematic properties with the discrete symbols of Classical systems. However, van Gelder's distinction is badly drawn.

The intuition behind 'concatenative compositionality' is in fact that the symbol-units which are *parts* of *several* complex symbols are distinguishable spatial (as in written natural language) or temporal (as in spoken natural language) parts of the wholes they conform. But of course, defining 'conca-

---

<sup>20</sup> These questions embody the traditional attack on behaviourism. Clark's partial demarcation from it (his requirements of some 'normativity' and some basis for 'consciousness'), upon whose unclarity I commented in the preceding section, would require for them to be put in a more subtle way. I believe that if Clark's demarcation were stated in a more precise way, corresponding questions could be framed with as much rhetorical force.

tentative compositionality' in this way (and 'functional compositionality' as any other alternative way of accounting for systematic properties) makes it impossible to argue later that 'functional compositionality' is incompatible with LOT, which is what van Gelder wants to argue. For there is *nothing at all* in the notion of LOT requiring that the 'words' composing the 'sentences' of the language of thought be their spatial or temporal parts (as they indeed are in natural languages). LOT arises from a more abstract commitment. Let  $\rho$  and  $\sigma$  be two sentences of the language of thought, each one representing the grammaticality of a sentence of Spanish for a given subject; and suppose that, according to the true system-theory for that subject's ability to discern the grammaticality of Spanish sentences,  $\rho$  and  $\sigma$  share a common unit  $\pi$ . It is clear that the only commitment arising from Cartesianism is to the following:  $\rho$  and  $\sigma$ 's causal-explanatory role in the subject's ability to recognize the grammaticality of Spanish sentences causally depends on their respective composition; and this composition, which plays a definite explanatory role in the recognition of grammaticality by the subject, is such that they share at least a common ingredient. It follows from this (as Fodor and McLaughlin (1990) stressed) that  $r$  and  $s$  must have a determinate and non-arbitrary physical composition (so that a complex representation could not be tokened unless its constituents be tokened); but in no way that their parts must be spatial or temporal. If, say, tensor-product resolution of the state of a network in vectors of activation per unit (constituting distributed representations) satisfied the constraint just indicated (which it does not appear to), a complex obtained in such a way might well constitute a sentence of the language of thought. Because van Gelder, I suppose, realizes this, he gives an obscure characterization of the distinction he is after. The result is that it is not clear at all that his distinction distinguishes anything.<sup>21</sup>

After making so much of the distinct 'superposed' nature of connectionist composition throughout the book, there is a passage in which Clark almost concedes the point just made (*AE*, p. 121): 'Could it be, then, that the notion of a *concatenative* encoding is actually the shallower notion of one which *looks concatenative to us*—that the distinction between functional and concatenative compositionality turns not on intrinsic properties of the representation but on how easily we human theorists can discern the structure of component parts within it?' The passage is surprising because, unless I have entirely missed the point, the concession that is being contemplated here would have momentous consequences for the book's argument. We had been told that successful connectionist models refute the 'Syntactic Image', among other things, precisely because they involve, if anything, 'superposed' representations. So what is Clark's answer? The answer is indeed consistent with the book's tenor; but it collapses superposition with context-dependence as the true and only reason why connectionist networks cannot satisfy the requirements of the 'Syntactic Image': even if the big difference between structured-

<sup>21</sup> This argument is developed in García-Carpintero (in press).

ness-by-superposition and structuredness-by-concatenation lies only in the eye of the beholder (even if, in other words, it can be ignored for any theoretically interesting purpose), there remains a crucial distinction between the 'symbols' that can be discerned in 'systematic' networks and those contemplated by LOT. The former are 'context-dependent', while the latter are context-independent (ibid.): 'the classical style of (concatenative) encoding is 100% symbol preserving. That is to say, the symbols are completely unaffected by their composition with other symbols. Connectionist modes of composition, by contrast, are symbol altering. What gets stored as part of a larger structure is not a straight copy of an original syntactic part.'

Thus, it all seems to depend on this 'context-sensitivity'. Indeed, the feature is explicitly introduced so that it excludes the presence of a language of thought (AE, p. 25):

The most radical description of this rampant context sensitivity would be that (these) connectionist systems *do not involve computations defined over symbols*. Instead, any accurate (i.e. fully predictive) picture of the system's processing will have to be given at the numerical level of units, weights, and activation-evolution equations, while more familiar symbol-manipulating computational descriptions will at most provide a rough guide to the main trends in the global behavior of the system. The proposal, then, is just that there are no syntactically identifiable elements which both have a symbolic interpretation and can figure in a full explanation of the totality of the system's semantic good behavior.

The point made here (an elucidation of some observations by Smolensky) is, I take it, the following. Consider again  $\rho$  and  $\sigma$ , two sentences of a given subject's language of thought representing each one the grammaticality of a sentence of Spanish which, according to the true system-theory for that subject's ability to discern the grammaticality of Spanish sentences, share a common unit  $\pi$ . Let us assume further that  $\pi$  is to be, if anything, a distributed representation, say, a vector of unit-activations; and also that  $\rho$  and  $\sigma$  arise by superposition in the way, compatible with LOT, we contemplated previously: that is to say, there are determinate ways of decomposing  $\rho$  and  $\sigma$  in their ingredient elements, and their relevant causal-explanatory behaviour is to be ultimately explained in terms of that definite composition. The problem Clark points to is that, if the decomposition is to provide a precise account of the computational behaviour of  $\rho$  and  $\sigma$ , no such common element carrying the common information will be found. For the fact that the parts-by-superposition of  $\rho$  and  $\sigma$  which in a connectionist setting can plausibly count as the computational implementations of  $\pi$  are in computationally different environments ('contexts') has the result that they will be somehow computationally different. And this difference will manifest itself in several computationally relevant ways, so that a full understanding of the behaviour

of  $\pi$ 's computational implementations in  $\rho$  and  $\sigma$ , respectively, will require us to take the difference into consideration.<sup>22</sup>

This is an important point, but once again I do not think it makes it at all mandatory to take an anti-Cartesian stance. In a passage I quoted previously, where Clark compares his present position to one he formerly held, he says (AE, p. 223) 'gone is the overriding commitment to finding scientific analogues (albeit at some high level of description) to the folk solids so as to allow their reductive identification with straightforwardly causally potent inner states.' The parenthetical concession here implies awareness of a point which has been overlooked in passing from context-dependence to anti-Cartesianism. Contemporary philosophical discussion of Physicalism has made familiar the fact of *multiple realizability*. Let us assume that we identify the characteristic properties of a given explanatory undertaking with those semantically tied to its proprietary theoretical terms—those appearing in the most precise basic laws that could ever be formulated to satisfy those explanatory demands on the basis of all relevant empirical data. Then the point of multiple realizability is this: the scientific requirement that higher-level properties should be explanatorily related to physical properties is not sensibly put as the requirement that higher level properties be *identified with* physical properties. For there are good reasons (in fact, it is all too obvious) to think that, put this way, the requirement is not to be satisfied: higher-level properties (say, *being an inflationary process*) are 'multiply realized' at the physical level. The issue of how the requirement should be put (in such a way that it sensibly accounts for the scientific intuition of mandatory inter-level relations), is the object of a much disputed metaphysical debate which I have tried to steer away from in this paper. If anything, higher-level properties are to be identified with 'disjunctions' of lower-level properties. In any case, it is to this fact that a gesture is made in the parenthetical concession by Clark in the previously quoted text: the 'scientific correlates' of the 'folk solids' are only to be found at 'higher levels' of description.

We defined previously the Cartesian point as the contention that we will not have acceptable explanations of systematic properties until we weld algorithms which can be plausibly claimed to have neurological realizations to semantical system-theories. In view of the fact of multiple realizability across science, the following possibility is obviously compatible with Cartesianism: the same explanatorily acceptable system-theory (say, of grammaticality in Spanish) holds for two subjects (or for the same subject in two different times), namely, one such that  $\pi$  is a common part of the sentences of the language of thought  $\rho$  and  $\sigma$ ; but  $\pi$  has different computational implementations in the two subject, or in the two times (and different neuro-

---

<sup>22</sup> This is also, I take it, the main point made in Sopena, 1994—although there it is put in a potentially misleading way, as the notion that the absence of a language of thought in connectionist networks concerns Marr's first level. The idea in both cases is, I think, that the level at which no precise counterparts can be found for entities with semantic properties is the level at which proprietary cognitive explanation is provided.

logical realizations). What they have in common is that they both computationally explain the grammaticality judgments constituting the informational content of  $\rho$  and  $\sigma$ , and involve a common element.

Now, context-dependence (coupled together with the possibility that connectionism ends up giving the correct computational theory of the cases under consideration) introduces an interesting wrinkle. What is being contemplated is the possibility that the algorithmic implementation (and, of course, also their neurological realizations) of  $\pi$  *in the same subject and time* be somehow different, when  $\pi$  intervenes in different possible grammaticality judgements. This, I think, is something to take into consideration when formulating the correct metaphysical account of inter-level relations. But does it require the retraction of some essential claim of Cartesianism? I do not see why. To the extent that every computational implementation of  $\pi$  takes part in grammaticality judgements, in the way indicated by the system-theory, they all still count as implementations of the relevant system-theory. There are still scientific analogues 'albeit at some high level of description' to the 'folk solid'  $\pi$ . (I presume  $\pi$  will not be so much folk as scientific, belonging to a syntactic theory for Spanish.) Thus, what Clark—following Smolensky—calls 'context dependence' (if I am understanding him correctly) is still compatible with 'implementation connectionism' (as presented in McLaughlin, 1993). Knowing the computational differences between different implementations of the same symbol can be theoretically important; it can even be important to understand properly some cognitive facts. But this is still compatible with the correctness of the system-theory, and the existence of symbols. Moreover, it should be noted that 'context-dependence' can be a misleading term. If implementation connectionism is correct and the Smolensky-Clark point well-taken, there might be 'context-dependence' at the lower levels; but from the higher-level viewpoint of the system theory, the same symbol  $\pi$  is always tokened, independently of the surrounding 'context'.

Thus we see that, although superposition and context-dependence constitute departures from the models with respect to which we tend to represent to ourselves by analogy the language of thought (namely, the artificial languages of mathematical logic),<sup>23</sup> the possibility that systematicity be explained by a system with superposed context-dependent representations does not force us in the least to abandon the tenets of Cartesianism. It only forces us to refine our thought about the language of thought in particular, and about

<sup>23</sup> Nor even natural language fits the bill as such a model. It is important to see that, in the sense of 'the same' in which the connectionist context-dependent syntactical implementations of the same semantic type are not the same, the syntactical realization in natural language of the same semantic content need not be 'the same' either. For the phonological and graphical sub-types of a given word-type are not the same, in this sense; neither are many of the different graphic sub-types of the same word. A word-type in natural language is also a 'disjunctive property'. (Superposition and 'context-dependence' are also aspects of the phonological sub-types of natural language types, by the way; phonemes are a case in point.)

inter-level relations in general. I conclude therefore that Clark's arguments do not help to make the philosophical point he wants to make (a point we had many reasons to doubt independently, as I pointed out in the preceding sections).

I want to make clear that the preceding discussion does not entail the acceptance of a positive answer to the empirically open question of whether connectionist systems can really exhibit systematicity—particularly of the interesting varieties. It must be remembered that superposition and context-dependence have to do with the reasons why present-day connectionist networks have difficulties in reproducing the systematicity involved in knowledge of grammar; for instance, they have to do with the phenomenon of 'catastrophic forgetting'—*AE*, pp. 145–6. And also that, as Clark himself acknowledges, the friends of connectionism tend to be somewhat gullible regarding the capacities of connectionist systems. ('Don't be too quick to assume that a network, even an apparently successful one, has actually fixed on the features on which you wanted it to fix', *AE*, p. 41, Clark advises.) Indeed, I have been convinced by people who know more of connectionist systems than I do that Clark himself might be exhibiting this gullibility in describing the ability 'to categorize words according to lexical category' of a certain network designed by Elman.<sup>24</sup> The point of the preceding discussion is merely conditional: there would be nothing of principle necessarily at odds with Cartesianism if systematic properties were explained by a system of superposed, context-dependent representations. Explaining—compatibly with the tenets of Cartesianism on cognitive explanation—systematicity on the basis of algorithms involving superposed, context-dependent *symbols* does not constitute a contradiction in terms.

On the other hand, from the fact that connectionism cannot be plausibly invoked to make the sort of conceptual contention Clark wants to make, it does not follow that its philosophical import is null. On the contrary, I suggested previously where this interest might lie; and, from that point of view, there are many intriguing suggestions to be gleaned from Clark's book. It is one thing to say that the systematicity of our linguistic competence and our ability for reasoning might be compared to what should be expected, had we acquired them in explicitly acquiring the sort of syntactic and semantic theory designed lately by linguists and philosophers. It is a very different one (explicitly advanced at times by Chomsky and Fodor, among others) to contend that we have acquired them so. This is a very implausible view—even if it were also 'the only game in town'—which should not necessarily be associated with Cartesianism.<sup>25</sup> The conceptual challenge here is to develop a better conceptual account of implicit knowledge, compatible with the empiri-

<sup>24</sup> *AE*, p. 26. I owe this point to Begoña Navarrete. See also McLaughlin, 1993, pp. 175–6.

<sup>25</sup> See, for instance, the helpful discussion in Block, 1986, pp. 646–8 of the paradoxical conclusion Fodor extracts from it regarding the extent to which concepts are innate. Clark, as I indicated previously, seems to think of this view as necessarily linked to Cartesianism.

cal facts about acquisition, and an alternative theory of its acquisition. The hope is that connectionism—implementation connectionism—might be of help here. As I said, there are fruitful suggestions in this regard in Clark's book. The exploration of this theme, however, must be left for another occasion.

*Departamento de Lògica, Història i Filosofia de la Ciència  
Universitat de Barcelona*

## References

- Block, N. 1986: Advertisement for a Semantics for Psychology. In P. French, T. Uehling and H. Wettstein (eds), *Midwest Studies in Philosophy*. Minneapolis: University of Minnesota Press.
- Churchland, P.M. 1989: *The Neurocomputational Perspective*. Cambridge, MA.: MIT Press.
- Clark, A. 1994: *Associative Engines*. Cambridge, MA.: MIT Press.
- Cussins, A. 1990: The Connectionist Construction of Concepts, in M. Boden (ed.), *The Philosophy of Artificial Intelligence*. Oxford University Press.
- Davies, M. 1991: Concepts, Connectionism, and the Language of Thought. In W. Ramsey, S. Stich and D.E. Rumelhart (eds), *Philosophy and Connectionist Theory*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Evans, G. 1981: Semantic Theory and Tacit Knowledge. In his *Collected Papers*. Oxford: Clarendon Press.
- Evans, G. 1982: *The Varieties of Reference*. Oxford: Clarendon Press.
- Fodor, J. 1987: *Psychosemantics*, appendix: Why There Still Has to Be a Language of Thought. Cambridge, MA.: MIT Press.
- Fodor, J. and Pylyshyn, Z. 1988: Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition*, 28, 3–71.
- Fodor, J. and McLaughlin, B. 1990: Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work. *Cognition*, 35, 183–204.
- García-Carpintero, M. In press: Two Spurious Varieties of Compositionality. *Minds and Machines*.
- Hadley, R. 1994: Systematicity in Connectionist Language Learning. *Mind and Language*, 9, 247–72.
- Konishi, M. et al. 1988: Neurophysiological and Anatomical Substrates of Sound Localization in the Owl. In Edelman, Gall and Cowan (eds), *Auditory Function, Neurological Bases of Hearing*. New York: J. Wiley & Sons.
- Marr, D. 1982: *Vision*. San Francisco: Freeman.
- McLaughlin, B. 1993: The Connectionism/Classicism Battle to Win Souls. *Philosophical Studies*, 71, 163–90.
- McLaughlin, B. and Warfield, T.A. 1994: The Allure of Connectionism Re-examined. *Synthese*, 101, 365–400.
- Peacocke, C. 1986: Explanation in Computational Psychology: Language, Perception and Level 1.5. *Mind and Language*, 1, 101–23.
- Quine, W.V.O. 1972: Methodological Reflection on Current Linguistic Theory. In D. Davidson and G. Harman. *Semantics of Natural Language*. Dordrecht: D. Reidel, 442–54.

- Sebastian, N. (forthcoming): Speech Perception in Catalan and Spanish. In M. Carreiras, J.E. García-Albea and N. Sebastian (eds), *Language Processing in Spanish*. Hillsdale, N.J.: LEA.
- Smolensky, P. 1991: Connectionism, Constituency and the Language of Thought. In B. Loewer and G. Rey (eds), *Meaning and Mind: Fodor and His Critics*. Oxford: Basil Blackwell, 201–27.
- Sopena, J.M. 1994: Connectionism and Language. In M. Bierwisch et al. (eds), *Language, Cognition and Computation*. Barcelona: Institut d'Estudis Catalans.
- van Gelder, T. 1990: Compositionality: A Connectionist Variation on a Classical Theme. *Cognitive Science*, 14, 355–84.