Explicitness With Psychological Ground

FERNANDO MARTÍNEZ and JESUS EZQUERRO
Department of Logic and Philosophy of Science, and Institute for Logic, Cognition, Language and Information (ILCLI) Universidad del País Vasco/EHU. Apdo. 1249. 20080 San Sebastián, Spain.
(e-mail: ylbmamaf@sf.ehu.es & ylpezmaj@sf.ehu.es)

Abstract. Explicitness has usually been approached from two points of view, labelled by Kirsh the structural and the process view, that hold opposite assumptions to determine when information is explicit. In this paper, we offer an intermediate view that retains intuitions from both of them. We establish three conditions for explicit information that preserve a structural requirement, and a notion of explicitness as a continuous dimension. A problem with the former accounts was their disconnection with psychological work on the issue. We review studies by Karmiloff-Smith, and Shanks and St. John to show that the proposed conditions have psychological grounds. Finally, we examine the problem of explicit rules in connectionist systems in the light of our framework.

Key words: connectionism, explicit, implicit, process, representation

1. Introduction

A traditional approach to explicitness maintains that it is all a matter of representation. Information can be represented in a system as part of the system's data (i.e. it can be a data structure), or it cannot; in the first case it is explicit, while in the second we can speak of implicit information. This approach, which considers explicitness as a matter of all or nothing, can be summarized in the following quotation from Cummins (1989, p. 158n):

> "Inexplicit representation" is, I take it, a contradiction in terms. For some content to be explicit in a system is just for it to be the content of some representation – i.e., (in the Computational Theory of Cognition) the content of some data structure. In orthodox computational systems, whatever isn't the content of some data structure isn't explicit

The exploration of non-classical information-processing systems, especially of connectionism, brought the necessity to rethink the issue. A possibility was to keep the traditional distinction and judge the new systems in its light, but this proves difficult, as structure and process go so intimately joined in systems like networks. The alternative, thus, was to develop a new account which served to define explicitness for a more comprehensive range of systems. This is, for instance, the task undertaken by Kirsh (1990), who proposes a process approach, instead of the traditional (that he labels 'structural') perspective.

An undesirable consequence with Kirsh's account is, as noted by Hadley (1995), that it leads to regard as inexplicit a good portion of everyday sentences which seem to convey explicit information. Hadley's proposal is to retain traditional intuitions,

Minds and Machines **8:** 353–374, 1998.
© 1998 *Kluwer Academic Publishers. Printed in the Netherlands.*

and to refine the notion of explicitness with narrow and wide definitions of explicit and implicit.

There seems to be a clash of intuitions behind the structure versus process dispute. However, a common problem to both kinds of analysis is that they do not ground their respective intuitions on corresponding psychological work. However, when applied to psychological phenomena, things do not appear so neat as a dichotomy explicit/implicit suggests. In this paper we want to advance a view of explicitness that, we contend, is both psychologically motivated and conceptually well-founded. After exposing the main assumptions that underlie the structural and process approaches, we will show how to retain some assumptions in a single framework that takes into account both structure and process. Then we will examine two recent psychological works and conclude that they are better accounted by our framework. We treat explicitness as a continuous dimension; that is to say, information will not be explicit merely if it is represented, but it will be more or less explicit depending on how it is represented and how its representation is processed. Thus, we will present different kinds of evidence to judge the extent to which we can say that a piece of information is explicit in a system. Finally, we will apply this analysis to the case of explicit rules, specially to their presence in connectionist networks.

[A preliminary clarification: We take that it is *information* which is explicit or inexplicit. So if we use the expression ‘implicit representation’ it must be understood in a derivative sense as ‘representation which conveys information **I** implicitly’. This is common usage (e.g. Cummins 1986, Kirsh 1990), though it should be contrasted with Hadley (1995), who takes ‘implicit information’ as elliptical for ‘information conveyed by an implicit representation’.]

2. Two approaches to explicitness

2.1. Structure versus process

In Kirsh (1990) we find the distinction between two different perspectives of explicitness: the so-called structural and process points of view:

> “From a process perspective information is explicit only when it is *ready to be used*. No computation is necessary to bring the content into a usable form. From a structural perspective information is explicit when it has a definite location and a definite meaning. The confusion arises when a representation that seems to be in a usable form when viewed structurally turns out to be in a non-immediately usable form procedurally.” (p. 344, his emphasis)

Kirsh carries on saying that standard views on explicitness are bewitched by the analogy of a printed word. The consequence is that explicitness is usually attributed to properties which belong properly to words. This treatment is distinctive of the sentential paradigm, that is to say, the view that conceives mental representations as sentences written in some language of thought. Kirsh spells out four of those properties and then tries to replace them with four conditions to judge explicitness from a process perspective.

This commitment to the sentential paradigm, however, is not necessarily characteristic of the structural perspective. Being explicit is being represented, but if representations were conceived differently (for instance, images instead of sentences) the analogies to be drawn would be different. The features that we consider to define each approach are the following ones:

In the structural approach a piece of information is explicit if there is some structure in the system that represents the information. Structures that represent the information are generally known as data structures. As explicitness is a matter of being part of the data structures existent in a system, this view is patently dichotomic: something is either included among the data or it is not, and it is explicit only in the first case. This approach is consistent with the classical view, which distinguishes strictly between structure and processing. We can find statements supporting it (though differing in important respects) in Cummins (1986), Fodor and Pylyshyn (1988), Adams et al. (1992), or Hadley (1995).

In the process approach a piece of information is explicit in a system if it is readily accessible by means of the processing mechanisms of the system. In contrast with the dichotomic structural approach, this view allows a continuum of explicitness: something can be more or less explicit inasmuch as it is more or less accessible. This approach would be favoured by connectionist proposals, since they blur the structure/processing distinction. More specifically, it is endorsed by Kirsh (1990),[1] Clark (1993), or Plunkett (1993).

One of the problems of the structural approach is the nature of implicit information. Several treatments have been proposed (see specifically Cummins 1986), and we could say that, in general terms, implicitness refers to information that is present in the system but that is not represented in the form of a data structure. In the process view, in contrast, there is not such a radical division between implicit and explicit information. Both kinds of information are, in fact, *made* explicit. The difference between them lies in the amount of processing needed to extract information.

There are many cases in which judgements coming from either approach will coincide (e.g. when reading the sentence ‘The cat is on the mat’, one forms an internal representation whose information is explicit from both the structural and process views). But there are many others in which we find disagreement. For instance, consider the case of a particular information you acquired in the past (e.g. Chucky’s phone number is 1234567), but that you are actually unable to remember. Let’s assume, for the sake of argument, that it was coded in some data structure of your cognitive system. Is that information explicit? From a structural view, it is certainly so: the information is explicitly represented somewhere in the system even if it is not accessible at the present time. It belongs to what Hadley (1995) calls ‘data which are temporarily inaccessible for *accidental* reasons’ (as opposed to data which are *intrinsically* inaccessible, and which will require further processing *even after the data is located*). From a process approach, however, the information in question does not count as explicit if it is difficult (or impossible) to access. The time you spend racking your brains in search of Chucky's number has to be taken into account. This is the view advanced by Kirsh (1990): it does not matter whether you spend time and effort locating information (in space) or decrypting it; what matters is the amount of time and effort required. In other words, we can say that from the process perspective explicitness is *explicitation*.

The dispute between structural and process views should not be regarded as a pointless terminological controversy. To begin with, we have already pointed out that neural networks are reluctant to be analyzed structurally, while they are amenable to a process treatment. So if we want to reconcile connectionism with a theory of explicitness, we cannot adopt a thoroughly structural approach. On the other hand, for artificial modelling purposes we find that in the structural view the designer can make relevant decisions in order to determine if something will count as explicit or not: if we include a rule among the data, we have rendered it explicit; if we hardwire it, it will be implicit. Knowledge representation is very much a matter of finding the right formalism to render information explicit. But in the process approach explicitness escapes to great extent the designer's control. One could construct a system so that its processing mechanisms make certain kinds of information more easily accessible than others, but whether a particular piece of information is explicit or not (for the system) cannot be generally determined in advance. Finally, and most important, there are features of each account that seem to motivate (or underlie) different psychological hypotheses, as we will see in section 3.

2.2. STRUCTURAL PROCESS

Structural and process approaches, as we have presented them, hold up opposite intuitions about explicitness. We want to introduce now a view that retains some features of each approach, while rejecting others. For lack of a better name, we will call this view *structural process*.

From the structural view we want to retain the relation between explicitness and data structures. For a piece of information to count as explicit it must belong to the pool of data that constitute the system's knowledge. This implies that we want to maintain the distinction between data and processing of those data (i.e. the structure/process distinction). The need of structure is, in fact, involved in the process view; if we base explicitness on accessibility we may ask: accessibility of *what*? What is this "something" we must access to access information? The answer has to be: a representation, a structure that bears the information in question. Of course, the structure does not need to exist in advance, it can be created in run time, using whatever representational resources are available to the system. But it must be an identifiable structure.[2]

The process view, on the other hand, supports an important insight: the idea of explicitness as a continuous, instead of dichotomic, dimension. There are two aspects that we have to distinguish behind this idea. Firstly, by insisting on the idea of accessibility, the process view suggests that for a given system information can be more or less easily accessible. This gradability upholds the notion of explicitness as something to be achieved by degrees. Secondly, as many authors observe, explicitness is *agent-relative*. This means that a data structure that represents explicitly some information for a system S, may not do so for another system S′. Now, if we conceive an agent as subserved by a multiplicity of (sub)systems (a picture highly coherent with what we currently know about mental architecture) it is perfectly plausible that a piece of information which is explicitly represented to some of the (sub)systems, is only implicit for some of the others. So information that is explicit to a greater number of (sub)systems is arguably more explicit (for the system as a whole) than information that is only explicit to, say, a couple of them.

Thus, what a structural process view contends is that to determine if some information is explicit for a given system we have to take into account (1) its representational status within the system, and (2) its accessibility. To state it in the form of a definition: Information **I** is *explicit* for system *S* iff

(i) there is a structure *R* that represents **I** in *S*,

(ii) there are processing mechanisms in *S* such that if a structure like *R* is processed **I** is immediately obtained, and

(iii) information **I** is usable by *S*.

Condition (i) is obviously the structural requirement. There is no information explicit to a system if there is not a physical structure that bears the information. Structure should be understood here in a broad sense, covering conventional representational means (e.g. productions, logical expressions) as well as non-conventional ones (e.g. networks' representations). In addition, structure can refer to a set of representations.

Condition (ii) asserts, on the one hand, that if there is a data structure in system *S* which represents some information, but there are no processing mechanisms in the system that allow it to access the information, then the information in question is *not* explicit for system *S*; in other words, it may be information *in S* but not *to S*. On the other hand, it demands that the information is obtained from *R immediately*. Kirsh (1990) tries to be precise about immediacy by appealing to accessibility in constant time. We agree with Hadley (1995) that Kirsh's constant time condition is too strict.[3] However, we are not going to provide another operational version of this admittedly vague expression. Establishing a cut-off point to tell explicit from implicit will be always arbitrary if, as we contend, we are dealing with a continuous dimension. We are content with Hadley's observation that "a representation conveys 'immediately usable' information if the representation *need not* be transformed into an informationally equivalent representation in order to be used **as data** in the reasoning of the agent, or as a basis for the agent's action." (Hadley 1995, p. 233, his emphasis).

Condition (iii), finally, states that explicit information must be usable by the system. This means that the information must be capable of influencing the functioning of the system, and therefore of playing a role in guiding the system's subsequent behaviour. Information which is always inert in a system cannot be regarded as explicit.

One would like to say that only information which meets the three conditions can receive the label of explicit, whereas all the rest would be implicit. We prefer to say that information is more explicit as it meets more conditions. In addition, the notion of immediacy in condition (ii) is flexible enough to allow a gradation. It can be concluded that information will be less explicit (or, if you prefer, non-explicit) inasmuch as it fails (i), (ii), (iii) or a combination of them. Let's consider the cases in turn.

That information fails condition (i) means that there is no structure in *S* that bears the information in question, even though *S* possesses mechanisms which, in principle, allow it to extract the information from such a structure. This is closely the case of the typical structural approach: if it does not belong to the representational pool, it is not explicit.

If information **I** fails (ii), *S* lacks the appropriate processing mechanisms to obtain **I** immediately (or to obtain **I** at all). It can occur that a representation *R* does not lead immediately to information **I**, but that by processing *R* we obtain, at some point, a representation R' that bears explicitly information **I** (e.g. an inferential device could perform a logical derivation from *R* to R'). From a structural perspective information in R' is not initially explicit, no matter how quickly (or slowly) R' is obtained. Our point, however, (and here is where a gradation of explicitness enters the picture) is that a piece of information which requires greater processing effort to be accessed will be less explicit than information which requires little computational effort to be accessed. For example, we could consider an expert logician for which many easily derivable theorems are practically as explicit as the axioms she starts from, even though inferentially the former are derived from the latter. (This is the sense in which some demonstrations are said to be trivial in logic texts; compare to novices for whom those "trivial" derivations are often not trivial at all!).

If information fails condition (iii) it is causally inert in the system. It might seem that this condition is not independent from the other two, because if a system has a representation *R* and means to obtain **I** from *R*, then information **I** must be able to play some causal role in the system. Because obtaining **I** from *R* amounts to understanding **I**, that is, to extracting the relevant meaning. But if **I** is not usable by the system then we cannot say that it is understood: understanding and usability are thus intrinsically connected. From our point of view, however, this is at most an empirical finding, i.e. in general we will not find systems in which (i) and (ii) hold while (iii) does not. But there is not a logical connection between the conditions: we could conceive a system that possesses isolated pieces of (represented) information which do not play any role whatsoever. (A different question would be its relevance for cognitive theorizing).

If neither (i), (ii), or (iii) are met[4], we find a limiting case for implicit information. There is no representation encoding **I**, the system lacks appropriate processing resources, and **I** is not usable. It seems that in this case **I** is not present at all, even implicitly. There would be a sense, however, in which **I** is implicit, as far as we dropped the idea of implicitness as agent-relative. This is, in fact, the way in which Hadley (1995) considers it (though his definition of explicit is indeed agent-relative). For instance, A is implicit in B if A is logically derivable from B, independently of whom will perform the derivation. As we are concerned with information which is interesting for psychological purposes, and as we are considering a single gradable dimension, we prefer to say that to judge in what degree information is explicit in a system will be always relative to the processes and representational capacity of the system. Moreover, if we adopt, with Hadley, an agent-relative perspective for explicitness, and an absolute point of view for implicitness, we are missing the close relationship that exists between both notions, and falling in the dichotomy entailed by the structural view.

3. Psychological import of explicitness

In the previous section we have offered an analysis of explicitness from a purely conceptual perspective. We have tried to provide a definition that embraces aspects that we consider fundamental to explicit information: its being data to the system supported in some representational structure, and its being a continuous dimension. However, the question of explicitness can be also approached from a more psychological point of view. In particular, there are two questions that we are going to address in this regard:

1. What is the role of the explicit/implicit distinction for psychological theorizing?
2. What sort of psychological evidence could be gathered to judge that information is explicit in a subject (or system)?

To answer these questions we will examine two different psychological studies that deal with the issue of explicitness. One of them is Karmiloff-Smith's (1992, 1994) theory of Representational Redescription; the second is Shanks and St. John's (1994) review and critique of the work on explicit versus implicit learning systems. Although each work is concerned with different problems, and uses the distinction for different purposes, we will see that in the end there is a common picture of explicitness that emerges from their discussions, a picture that lends itself to the be analyzed by the structural processing view we have advanced. Let's review briefly the basics of both studies.

3.1. THE ROLE OF EXPLICITNESS IN PSYCHOLOGY

According to Karmiloff-Smith, humans are endowed with a mechanism for the repeated redescription of their own knowledge. She names this process *representational redescription* (RR). RR is developed from childhood in an internally-driven way. It seems that young children represent their knowledge mostly implicitly and activate it in response to external stimuli.[5] The ability to redescribe their representations allows them to build a fast primitive level of explicit representations, which begins to be knowledge *to* the system instead of merely knowledge *in* the system. This knowledge is subsequently refined to get increasing degrees of explicitness. The RR process does not only occur along development, but also across any given moment of it. That is, an adult will use RR processes to improve performance in many domains. So in this theory the explicit/implicit distinction plays the role of distinguishing between a multiplicity of representational formats that information can adopt in a system.

Shanks and St. John's (1994) paper is aimed to test the existence of different learning systems in two dimensions: explicit *versus* implicit learning, and rule-based *versus* instance-based learning. We should realize that in this context explicit means 'concurrent with awareness'. S and SJ review a good number ofexperiments to conclude that the results provide no basis for talking of unconscious learning, but they support the idea of different learning systems in the rule/instances dimension (more on the latter in section 4). This is not only observed in experiments involving certain complexity, such as learning artificial grammars, but even in simple tasks, such as Pavlovian conditioning, some awareness of the relation is required. To make their point they introduce two criteria which must be attended to effectively assess whether the subject is aware or not. One is the *Information Criterion*, which refers to the match that must exist between the information the experimenter is looking for in the awareness test and the information responsible for performance changes due to learning the task. The other one is the *Sensitivity Criterion*, which appeals for means to show that the test of awareness is sensitive to *all* of the relevant conscious knowledge. Applying these criteria, S and SJ argue that in all the experiments they examine either the information that is searched for is not relevant for the task, or the information that is relevant is not searched by sensitive means. Hence, the role of the explicit/implicit is, in this context, to distinguish between information of which the subject is aware from information which is not present in consciousness.

What sort of conclusion should be drawn from these different treatments about the role of the explicit/implicit distinction? A first temptation is to consider that they reflect different usages of a term for different purposes. Thus explicitness would be a label used simply to refer to different phenomena in different contexts (i.e. consciousness *versus* representational format). However, if we examine Karmiloff-Smith's RR theory it turns out that both usages are related. KS distinguishes several levels of representation which individuals progressively attain. First we find a level-I (implicit) in which “knowledge is represented and activated in response to external stimuli, but it is not available for use by any other part of the system”. From it we move to a level-E1 in which “via a process of redescription, condensed representations are explicitly defined”. Yet, these representations are not accessible to consciousness nor to verbal report. This access is gained by successive redescription which produces level-E2 and E3. It is plain, thus, that access to consciousness is a defining characteristic of explicitness in RR theory, just the same as in the context of explicit learning.[6]

Another hasty, erroneous conclusion to draw is that Karmiloff-Smith’s account is akin to structural approaches to explicitness, while Shanks and St. John’s could be analyzed from a process perspective. In RR theory what underlies the capacity to redescribe representations can be formulated as the ability to turn information into data structures. The problem for the system is in this case to find the appropriate format to represent information: being explicit is a matter of being encoded in a suitable structure. On the other hand, Shanks and St. John are clearly speaking about the accessibility of certain information. The point of the Sensitivity Criterion is to detect what internal information can be accessed by the system to perform a particular task. This fits the process perspective in which information is explicit inasmuch as it is more easily accessed.

In fact both conceptions can be accounted for from a structural process view. First, if KS’s theory leads to a structural approach it would entail a dichotomical notion of explicitness. But by positing different representational levels information is considered more or less explicit as it is formatted in an upper level type. In fact, the dichotomy between explicit and implicit is expressly rejected in several places and replaced by a continuous approach.[7] Nevertheless, the theory cannot be accounted by a purely process view either, due to its insistence on the representational structures that are constructed along the RR process. In this theory, explicitness amounts to (1) being represented (redescribed) in a format that (2) conveys its information in a more accessible way to the system. By the same token, in S and SJ learning context explicitness usually involves, in addition to the involvement of consciousness, the construction of a representation of the task; that is, it is by constructing this representation that relevant information is usually accessed (this is to make information explicit).

Hence, psychological theorizing shows that explaining explicitness requires appealing both to structure and process. A conceptual framework that omits either aspect will not be appropriate to deal with a psychological notion of explicitness. We contend that our structural process framework captures the relevant features of both aspects.

3.2. Psychological evidence for explicitness

Depending on the different treatments of explicitness, the way to probe the existence of explicit information in a system will vary. There are two pertinent questions: how to probe that there is some information at all, and how to probe that the information is explicit for the system in the sense we have defined (i.e. that it is represented and easily accessible). We contend that the chief evidence on this respect belongs to four different groups: speed, generalization, verbal reportability, and consciousness.

3.2.1. *Speed*

From a process perspective, that rates explicitness from its accessibility, speed turns out to be a crucial parameter. The assumption here is that explicit information is *more rapidly* accessed than non-explicit information. This is Kirsh's (1990) position. As we said above, in his account information is explicit when it can be accessed in constant time. Speed is also taken as a piece of evidence by Hadley (1993). In this earlier paper he focuses on the phenomenon of rule acquisition. From his structural view, the fact that people are able to acquire and retain novel rules almost instantaneously suggests that they assign explicit representations very rapidly to the instructions they are told.

However, if we consider matters from the structural process view we are advocating, speed is only part of the story. It is obviously a way to test the immediacy of information: the less time it requires to activate some information, the more immediately usable this information can be. But speed cannot tell us much about the issue of representation. It can be the case that the information is not represented but nevertheless quickly used (e.g. a hardwired rule). Hadley's point (about the rule's being represented by subjects) cannot be made only by appealing to the *acquisition* of the novel rule but by its *retention* and subsequent reproduction. Thus, speed of processing, by itself, can merely show that a system uses immediately some information, not that it has any representational abilities whatsoever. The information in question could count as non-explicit. (We should notice, however, that it would be more explicit than information which is not even accessed rapidly).

3.2.2. *Generalization*

A method to determine if a system possesses some information is to design a test in which the system must employ the information in question. A common way to do this is by testing if the system is able to generalize to unknown instances in which the information is relevant. To put a quick example, in a conditioning experiment we present an animal a red square and a blue circle and it is reinforced with food every time it picks the former. Now we present a red circle and a blue square. If the animal chooses consistently the red circle we may infer that the information which is likely behind its behavior is (loosely speaking) 'redness is related to food'; if it chooses the blue square, the information can be 'squareness is related to food'. As Shanks and St. John Information Criterion states, the experimenter must look for the information which is responsible of the behavioural changes observed, as successful learning depends on the system's ability to profit from the right information provided by the domain. In our oversimplified example, the animal may be attending to colour-related information, or it may attending to shape-related information. So the experirnenter must test if it generalizes to colours or to shapes.

On the other hand, the ability to generalize is also behind Karmiloff-Smith's level-I. This level is characterized by the achievement of "behavioural mastery", i.e. successful performance in a domain. In other words, when a system is able to make successful (and fast) generalizations from known to unknown stimuli of a domain, it can be said that the system masters the domain.

So generalization is, as speed, evidence for judging if information is explicit. In this case the point is not that information is immediately at hand, but that it is put into use for the system. Again, it could be that the system has not formed a representation of the information in question, it might be accessed through other representations, which do not convey the information explicitly. Turning to our conditioning example, it could happen that the animal has encoded representations of objects that it is able to associate with representations of other objects, but the link (i.e. the rule) which relates classes of objects may not be represented anywhere.

3.2.3. *Verbal Reportability*

How can we know that subjects are employing explicit representations in some task? A rule of thumb would say: Ask them! As a matter of fact, the logic behind most explicit learning experiments relies on something similar: to discover if subjects are aware of some information the psychologist investigates if they are able to produce a *verbal report* of the information in question. When this is not the case it is concluded that learning was implicit. In contrast to speed or generalization, verbal report is a piece of evidence that information is represented. Someone may like to object that what we observe is an external representation, and that it does not warrant the existence of a corresponding internal representation. The report, for example, might be a canned response. This hypothesis, however, is not consistent with psychological theorizing with respect to complex systems. The best working hypothesis is currently to assume that there are internal representations which underlie one's reports, though they do not need to be of the same nature (linguistic, pictorial) as the external format of the report.

What one can report depends, obviously, on what one's representational resources can express. For instance, there is information expressible in predicate logic which cannot be expressed in propositional logic. A system whose representational elements were taken only from the latter would not be able to make explicit certain information.

To possess a representation of some information does not involve, to say it again, that the information was immediately accessed. In fact, the process of framing the right representation requires time and it is part of the accessing of information, that is to say, information is *made explicit* by accessing and representing it. In normal conditions reportability is usually enough as it is assumed that the subject has accessed and is able to use the information that is represented in the report.

But a common problem is that the external observer (the experimenter) ascribes to subjects more information than they really have. The observer may draw the information she is seeking from the report by an inferential process (in which the observer inserts her own information). Thus the need of strict control of the information that is really behind the behavior, which is what Shanks and St. John demand in their criteria.

When a subject produces a verbal report it is usually assumed that the information she communicates meets conditions (i), (ii) and (iii). It is not surprising, thus, that it constitutes the strongest evidence to determine that a particular piece of information is explicit to a subject. Of course, the external representation (the sentence/s) may not convey its information explicitly for the hearer (or even for the speaker), but its great advantage is that it allows such explicitation.[8]

3.2.4. *Consciousness*

Subjects can be conscious of more information than they are able to report. In this account, S and SJ contend that verbalization is not always a sufficient or suitable method to measure people's awareness during the task: (1) because it can lead subjects to consider only the most evident source of information at hand, namely, that provided by the experimental condition, disregarding information which actually causes their performance, and failing thus the Information Criterion; (2) because there are other tests (recognition, prediction) that can detect information left undetected by verbal report, meeting thus the Sensitivity Criterion. Furthermore, one of the levels postulated by Karmiloff Smith, level E2, is hypothesized to consist of representations which are available to consciousness though not to verbal report. (An example would be spatial representations[9]).

When information **I** is present in consciousness in non-linguistic format, **I** will not usually be conveyed in such an explicit way as in linguistic format. (Recall note 8). The reason is that **I** is not so easily accessed, thus failing our condition (ii) in its aspect of immediacy. The need to frame more accurate means to detect this information (S and SJ's Sensitivity Criterion) suggests that it is not readily at hand, that is, it requires more processing effort by the subject.

3.3. THE PROBLEM OF UNDETECTABLE EXPLICIT INFORMATION

In conclusion, speed can be regarded as evidence that certain information is being accessed *rapidly*, which is related to condition (ii); generalization is evidence about the *kind* of information that the system is using, which is related to condition (iii); verbal reportability is not only evidence about kind of information being used but also that the information can be formatted in certain representational structure, which is related to condition (i); finally, consciousness (without verbal report) is also evidence that the information is *used* and may point at the representational structure doing the job. Therefore, it seems that our conditions on explicitness have a correspondence in different pieces of psychological evidence to determine when something is explicit.

By combining the four sources of evidence above, we can conclude that information which is fastly retrieved to consciousness, assigned a verbal format, and immediately usable for a task by a system will be *highly* explicit for this system; inasmuch as the evidence from any source is weak (i.e. slow retrieval, non-verbalizable, unable to use) it will be possible to say that information is not explicit for the system. This suggests a gradation of explicitness from the strongest, verbal form toward the weakest, information merely present (somehow) in a system, passing through the intermediate levels in which information is accessed but not represented, or is represented and accessed slowly, etc. This is consistent with Karmiloff-Smith's theory, in which the most explicit information corresponds to level E4, and the least to level-I; it is also consistent with Shanks and St. John's approach, in which the information underlying the subjects' performance can be verbalized or not, being the former, in our account, more explicit than the latter.

There is an objection that we would like to consider before proceeding. According to this objection, explicitness has nothing to do with the alleged psychological evidence.[10] The problem is that there might be explicit information that was represented but processed slowly, with a narrow scope of generalization, not verbally reportable, and not available to consciousness. It is conceivable that such information would meet our conditions in the following manner. The information in question is represented (condition (i)), but not available for linguistic format. When the representation is processed, the information is achieved immediately, that is, without intermediate representations (condition (ii)). However, this information is not obtained quickly, it requires time to extract the information from the representation. Finally, this information plays such a specific role that leaves it undetected by any generalization test, but it is used anyway (condition (iii)). Due to these features we will call it *undetectable explicit information* (UEI). So it seems that we can have explicit information for which there is no psychological evidence (in fact, this might be the case of much information which a conventional computer includes among its data structures). Hence, this evidence is irrelevant to the problem of explicitness.

Our response to this objection is doublefold. First, it must be noted that the aforementioned psychological evidence does not belong to the set of conditions on explicitness. It only points out means to ascertain that the conditions have been met. There could be other methods to determine when a piece of information is explicit that have escaped our attention, and it might be the case that some of the sources we have considered constitutes poor evidence. But inasmuch as they have been used in psychological research to determine what information a subject have, and in what format, they should not be discarded as irrelevant to the issue. Second, one should postulate the existence of undetectable explicit information, as far as this theoretical construct plays some role in our cognitive theories. From the questions that we posed at the beginning of this section, (2) is ruled out: we cannot ask for psychological evidence of something which is defined as psychologically undetectable. But we can still ask (1), that is, what role would UEI play in psychological work? It certainly is of no use in implicit *versus* explicit learning, or in the RR paradigm, since in both contexts it is characteristic of explicit information to be available to awareness. Maybe the possibility of UEI could be intended as a criticism to both psychological theories: they would be using a wrong conceptualization of explicitness. But then the proponent of UEI should offer a better alternative. In consequence, the issue of what kind of psychology is related to this treatment of explicitness must be faced.

Consider the following research strategy (which is, roughly, the general procedure in implicit learning research). We postulate that a subject must be using some information to perform a task. Then we verify with our best available methodology that the information is not present in consciousness, is unreportable, and so on. Then we conclude that the information is not explicit. Is there anything wrong with this procedure and conclusion? What other considerations should be taken into account to conclude that we are before a case of UEI? A possible answer is to appeal to the potential role of UEI in conventional computers. In these computers there are pieces of data which are presumably unconscious, which the system cannot report, which support little generalization, and that may be processed slowly. These data would be UEI. Thus, as far as an artificial model can be regarded as a (good) model of some cognitive process, and as far as it needs UEI, then UEI must be postulated in cognitive theorizing. So even if the information is unconscious and the like, our artificial model tells us that it has to be explicitly represented.

However, the decision of what, if any, information must be represented (a requirement to be UEI) is precisely what is at stake when such a cognitive model is designed. In order to make this decision one should not consider computational requirements alone, but ponder on the psychological ground of the decision. If our model requires structures without a psychological counterpart, they could be regarded as implementational detail (and hence the role of UEI would have a doubtful relevance to psychological level) or as *ad hoc* additions (and hence in need of further justification). To do otherwise would be to put the cart before the horse. We contend that both psychological works (KS, S and SJ) described above provide some ground to assert that undetectable[11] information should be considered implicit. To be sure, we grant that there can be information which is represented but unconscious, unreportable, and so on. What we reject is that the information in question can be regarded as explicit.

We might be charged of mislabelling the view of explicit information we are criticizing. The point, it could be argued, is not that it cannot be detected. We could detect it by inferential means. Imagine that we are studying some mental process **M**. We postulate that some information is involved, and that it is unconscious, unreportable and narrowly generalizable. In addition, **M** is slow, so that whatever representations mediate it we can think that they are processed slowly. We infer, thus, the presence of information which is not conscious, not reportable and so on, but explicit nonetheless.

Again, what is the point of regarding this information as explicit? The only basis is its presumed representational character. But, what is the basis to infer that the information is gathered from the representation without mediating structures in which the original representation is transformed (condition of immediacy)? The very fact that the process is slow could lead to guess that the information is not extracted immediately from the representation. One would like, then, to reject the condition of immediacy. However, this move leads to the undesirable consequence that any representational element would count as explicit even though it had to suffer a long transformational process for its information to be used. This is the standard structural approach, in which to be explicit is just to be represented. We have been trying to show that this approach offers a notion of explicitness too poor for most psychological purposes: we should preserve a difference between represented information ready for use, and information contained in a representation but that the system cannot use unless such a representation is transformed into another. This is, we contend, what Karmiloff-Smith's theory tries to do.

Hence, as we have been arguing, the psychological role of explicitness, and the psychological evidence for explicit information, fit more smoothly a notion of explicitness that takes into account both structure and process, than either the structural or the process view alone. From this framework we can revise problems in which the notion of explicitness is involved. One of these problems is the symbolic/connectionist dispute on the issue of explicit rules. So our next step will be to examine the issue in the light of our previous discussion.

4. Explicit rules and connectionist systems

The debate on explicit rules hinges on two questions: one is if models of cognition require explicit rules, the other is whether connectionist models can be said to have explicit rules. It is plain that the relevance of the second question depends on the answer given to the first one: if explicit rules are not necessary then it will not be, in general, very important if some systems have them or not, as far as they comply with other features that are considered necessary. Both questions, however, demand an analysis of explicitness to determine when a rule can be said to be explicit. Of course, the discussion in the previous section was not directed to provide arguments for the need of explicit *rules* in cognition, but to offer some psychological ground on the notion of explicitness in general. Nevertheless, we will assume that both symbolic and connectionist researchers assume that explicit rules have a function for at least *some* aspects of cognition (arguments for explicit rules can be found, e.g., in Clark 1993, or Hadley 1993). So it will be worthwhile to examine the connectionist capacities in this respect.[12]

In this section we do not want to provide an exhaustive account of rules. We take that a rule is a particular kind of information, namely, information of a relation that holds between X and Y (where X and Y are some kinds of objects, in the broad sense of the word). In information-processing systems rules usually have a behavior-guiding role. A rule as information must be distinguished from the representation of such a rule. Rule R may be contained (as information) in the set of (pairs of) instances in which R holds. For instance, in the set S={42→even, 13→odd, 66→even, 51→odd, 28→even, 75→odd} we can say that the rule R = 'if the rightmost numeral is even, the number is even; otherwise it is odd' is contained. However, usually sets may contain more than one rule (i.e. it may be undetermined what rule holds, if any). For instance, in the set above the following rule R′ can also be said to hold: 'if the leftmost numeral is even, the number is even; otherwise it is odd'. But neither R nor R′ is represented in the set: it only contains pairings of numbers with properties.

Hence, it can be the case that a system represents a group of instances in which a rule holds *without representing the rule itself*. The system will contain the rule as information, but not as a representation. But the system may have access to the rule-information through the instances in which the rule holds. As the rule(s) may be undetermined, it is not always clear which rule (information), if any, is guiding the behavior of the system. A useful way to ascertain this by an external observer is to introduce a novel input that discriminates between rules. For instance, a system that represents the instances of set S can be offered as input the number '12'. If the system's output is 'even' we can conclude that it may have access to rule R; it the output is 'odd' it may access rule R′; if the system is unable to produce an output, we can conclude that it may have access both to R and R′ (and thus it has reached a kind of impasse) or to none of them (the system may be processing the information conveyed individually by each instance, without accessing the abstract information which constitutes the rule).[13] Note that what is assessed by this method is the system's generalization ability: the judgement is limited to infer the information the system is handling, not the representation of this information. Having access to information R (R′) is not the same as having a representation of information R (R′). The system may have access to R (R′) *by processing representations of something else*.

Connectionist systems, from this perspective, can be seen as systems that do not represent the rules, though their behavior is guided by rules. This is not only to say that their behavior merely *complies* with the rule, but that they have access to the information that constitutes the rule.[14] What they represent, however, is not the rule but the instances in which the rule holds. In support of this distinction we can consider psychological research again. As we noted in section 3, the second dimension Shanks and St. John (1994) examine in their work is rule-based *versus* instance-based learning. In this case they conclude that there is well-founded evidence to posit the existence of two distinct learning strategies: one in which subjects test rule-like hypotheses, another in which subjects memorize instances. According to S and SJ, these strategies demand different mechanisms: a rule-representing mechanism, and an instance-encoding mechanism. Moreover, it is often possible to acquire a representation of the rule having learnt instances alone; this suggests that the information which constitutes the rule is already contained and accessible from the instances. S and SJ also maintain that connectionist systems, which learn by exposition to examples, encode individual instances in their weights. Networks are able to generalize to the global set of instances of a domain because of their distributed style of representation, which allows extracting statistical features and matching patterns between instances. But in this process no representation of the rule as something different from the instances is constructed. S and SJ do not think, therefore, that networks are good models of hypothesis-testing, as there is no representation within the models of rules or of the testing of hypothesis. On the other hand, Karmiloff-Smith (1992) contemplates networks as a possible implementational mechanism of her RR theory, to conclude that they could be the basis for the implicit level, but not for either of the explicit ones.

Concerning our structural process framework for explicitness, the conditions connectionist systems fail are (i) and (ii): they do not possess structures that represent the rule, and they do not have means to extract the information from a representation of the rule, but from the processing of other representations. Thus they do not possess explicit rules. They meet, however, condition (iii), as they are able to use the information that conforms the rule. Imagine a network that was trained with a set of pairs similar to set S mentioned above (more extensive, if you want to be realistic). Depending on the way the network was designed it could be "attending" to information related to the left number, the right number, both, or none. (This could be determined, as we said, by presenting the network a novel discriminative input). The network may, thus, have access to some abstract information which guides its behavior, but all it encodes is the instances. Contrast this with a system based on a look-up table that also encodes the instances. When presented with a novel input the system does nothing. This is because it does not have access to any abstract information based on the instances. So this system would not meet any of the three conditions with respect to explicit rules. A rule would be implicit in it, in the sense of the limiting case mentioned in section 2 (and the sense of implicit contended by Hadley (1995)): it would be information inferable by someone with the right processing and representational resources.

We will consider now two different objections that would refute this view about connectionist capacities for explicit rules. The first says that both classical and connectionist devices can represent rules explicitly (Adams et al. 1992); the second states that, in fact, many networks make use of explicit rules (Plunkett 1993).[15]

For Adams et al. to use an explicit rule is to consult a rule. Hence, explicit rules are part of the data used (i.e. consulted) to process information. (They contrast them to instructions, which are “the primitive changes that are to be made as defined by the basic ontology of a computational device”.) So their account is very close to the structural approach. According to them, there is no reason to suppose that a network cannot consult rules, provided that we feed it with a pattern of activation constituting the rule. This pattern would be a representation of the rule. One could, for instance, feed a pattern representing the rule R = 'i before e, except after c'. The pattern would meet, at least, condition (i), contrary to what we hold.

Although Adams et al. affirm that how and why this can be achieved is open to question, there are two main options. The first is to feed the network with a number of instances of R, (e.g. [si:lIη], [rIsi:v], [fi:ld], [rIli:v] associated to 'ceiling', 'receive', 'field' and 'relieve', respectively). The network would usually be said to generalize R from the instances. However, this cannot be what Adams et al have in mind, since they demand that the pattern that represents *R itself* is feeded, not merely the patterns that represent the instances. So this leaves us only with the second option, which is to construct a network such that it can possess abstract rules like R previous to (or independently to) learning any instance of R. But to this end it would be necessary to solve the problem of variable binding in networks. Though the problem has been attacked by different authors (e.g. Smolensky 1990, Sun 1995), its solution is far from clear. Moreover, it is not obvious if this solution is what Adams et al. demand either: the networks are not generally "feeded" with the rules; rather, they are constructed so that the rules are embedded in the architecture. It is uncertain if these rules are part of the data, to meet Adams et al.'s requisite, or they are more similar to hardwired rules. In the latter case they would not have representational character. In fact, we can suspect that this is the case, given that these rules are not processed themselves, but are used to process other inputs. In addition, they are not transportable or easily removable, as data usually are (or should be).

Hence, even if it we conceded that, as Adams et al. affirm, whether networks can represent rules is open to empirical corroboration, current evidence shows that it is, to say the least, problematic to achieve it. The results show to the date that they function more like processing elements than like data. Nevertheless, the second objection we want to consider is the claim that, even playing this role, rules could be regarded as explicit in networks. In fact, even in the first option we have presented, that is, in the typical learning method in networks,[16] the networks handle explicit rules. According to Plunkett (1993) even the simplest network carries explicit representations provided that (1) its weight matrix was explicitly adapted to the mapping demanded by the task, no matter if the representations are not explicit for the observer, and (2) it can exploit its information to learn a new task with an analogous structure. An example is a network that is trained to learn the English past tense, and then it is tested how well it learns the plural system of English, which has structural similarities with the former.

It is plain that for Plunkett explicitness is a relative notion, dependent on the context of the task performed, that should be judged from the point of view of the system that is carrying out the task. Therefore, the fact that weight matrix representations are often opaque to the modeller does not imply that they are not explicit. To judge if the network has explicit rules Plunkett considers two things. One is the information supplied by the task: if the system is correctly adapted to task demands he assumes that it is using the relevant information. The other is the system's ability to generalize: if the system is able to use the information obtained in a task to perform a structurally-related task, then the information is something more abstract than the instances it has encoded. In other words, information abstracted from a set of instances to solve a task is used to solve another task in which a different set of instances is involved. This abstract information is what we can call the rule.

We agree up to this point. But we do not agree with Plunkett's subsequent conclusion that the network represents the rule. The problem is not, as he supposes, that the presumed representation is opaque for the observer. The problem is that the information used by the network is not used in a representational format. There is not a structure in the net that represents the rule as something different from the instances. Plunkett's network exploits a regularity from a task to master a related task, but this regularity is not represented explicitly, only the instances that support it. This characterization would turn practically any information-processing device that complies with a rule into a rule-representing device. For instance, hardwired rules in a classical system could count as explicitly represented given that (1) they have been explicitly designed to perform certain task, and (2) they could be employed to perform an analogous task, as far as the variables involved are conveniently arranged. Or, even worse, a look-up table with canned responses for both analogous domains (say, past tense and plural) could be judged as a device that represents explicit rules.

At this point it may seem that we are sustaining an inconsistent view of representation. It is true that a system can receive more information than it is able to represent, that is, there is information in a system which is not represented. However, it can be objected that with respect to information a system *processes* it must necessarily be represented, since the way a system has to process information is by processing representations of that information. Now, we are holding that there can be systems that access information that they do not represent; hence, that these systems process information by non-representational means. Thus these systems cannot be information-processing systems, or at least they are outside the representational paradigm. To express the matter in different terms: how can a network access the rule-information if it is not by processing a rule-representation? The representation must be there somehow, so Plunkett must be right.

Our answer to this objection can be easily inferred from our previous discussion. The conception of information-processing we have in mind is, indeed, the processing of representations. Connectionist systems are representational systems too. They usually represent objects and properties. The weight matrix represents, in general, how different objects ane associated to different objects. But they do not represent a rule that stands independently and above the particular representations. The so-called prototype that networks extract is not a representation of an abstract rule, but a combination of superposed features of different instances. The network could be feeded with novel combinations of objects and properties (or microfeatures of objects and properties) taken among the pool of representational resources allowed for the network, but neither of these combinations constitutes a rule. The information that constitutes the rule is thus accessed by processing representations of things other than the rule itself. This is the hallmark of implicit information: it is present and accessible by the system through the processing of something else, not through the processing of a structure that represents it. To turn to the example of hardwired rules again, the rules are present in the system (they are information in the system) but they are not processed: the information is accessed when something else is processed, namely, the data. In connectionist systems the rule-information is accessed when the instances are processed. Thus, it is the instances what function as data for networks, not the rules.

5. Conclusions

Our definition of explicit information was conceived to account for two properties that we consider to be essential in it: its representational character, and the existence of a continuum of explicitness depending on the amount of processing required. There is empirical work that fits this notion, and that provides means to test different aspects of explicit information. With respect to connectionist systems it must be emphasized that they do, indeed, use rules as abstract information that guides their behavior. What they do not use is representations of rules. Thus, when it comes to rules networks meet partially the conditions we proposed for explicitness. If explicit rules turn out to be necessary for cognition this would be a negative consequence for connectionist cognitive theorizing. However, even though we think that there are good reasons to demand from our cognitive models the capacity to deal with explicit rules, we reject that this entails a "no explicit rules, no cognition" conclusion. The view we offer avoids such dichotomy. If explicitness is considered a matter of degree, as we contend, the consequence is not so negative, since networks could always be directed to model those aspects of cognition where less explicit rule-information is involved. This could be a vast portion of our capacities.

6. Acknowledgements

We would like to thank especially Kenneth Aizawa for his detailed comments on an earlier version of this work. A previous version was presented at the IIIth Intemational Colloquium on Cognitive Science, San Sebastian, 1995. We also thank Fernando Broncano, Enric Casabán, Andy Clark, Toni Gomila, Bill Lycan, Kim Plunkett, Stuart Silvers and an anonymous referee for different comments and suggestions. This work has been supported by the Research Project Number PS95-0098 of the *Ministerio de Educación y Ciencia (DGICYT)*. Fernando Martínez's work is funded by the *Departamento de Educación* of the *Gobierno Vasco*.

7. Notes

[1]Though motivated by the discovery of new computational mechanisms like networks, in which informational trajectories are more difficult to track, Kirsh does not offer a connectionist account of explicitness in his paper, only a process account. In addition, some of the requirements his conditions retain could be label structural, as the condition of separability. So his proposal could be reworked in the structural process approach we will defend.

[2]Kirsh (1990, p. 351) establishes a separability condition which functions as a structural requirement. It says: "The states structures or processes – henceforth symbols – which explicitly encode information must be easily separable from each other".

[3]The condition is: "Symbols explicitly encode information if they are either: readable in constant time; or sufficiently small to fall in the attention span of an operator." (Kirsch, 1990, p. 358).

[4]A system may fail just two of the conditions. Though it may be difficult to find a particular system that fits a combination, we'd like to stress again that our point is not empirical, but conceptual. Anyway, to illustrate briefly our point, we propose the following examples: networks fail conditions (i) and (ii) with respect to rules (as we will argue in section 4); failure of (i) and (iii) could exist in young children in a stage in which they have not formed a right representation of some information, are unable to use this information, but possess the processing mechanisms that would allow them to obtain the information from the representation; failure of (ii) and (iii) would be a system that has an information encoded in memory (i.e. Chucky's phone number) but cannot retrieve nor use the information in question.

[5]Although she speaks of implicit and explicit representations, we think that it is not a misinterpretation to treat them in the derivative sense mentioned above, i.e. as representations which convey their information implicit or explicitly (more or less explicitly).

[6]The main difference between both approaches in this respect is that while for Shanks and St. John all learning is concurrent with awareness, for KS there is level-I and level-E1 learning which does not require any involvement of consciousness. But for subsequent degrees of explicitation access to consciousness is indeed a relevant feature.

[7]See, e.g. Clark and Karmiloff-Smith 1993, p. 496, Karmiloff-Smith and Clark 1993, p. 573, or Karmiloff-Smith 1994, p. 736.

[8]As Sperber and Wilson (1986, p. 60) nicely state: "Non-verbal communication tends to be relatively weak. One of the advantages of verbal communication is that it gives rise to the strongest possible form of communication; it enables the hearer to pin down the speaker's intentions about the explicit content of her utterance to a single, strongly manifest candidate, with no alternative worth considering at all. On the other hand, what is implicit in verbal communication is generally weakly communicated: the hearer can often fulfil part of the speaker's informative intention by forming any of several roughly similar but not identical assumptions."

[9]Another good example are some of the representations involved in hearing and understanding instrumental music. Diana Raffman (1993) uses this case against some philosophers that claim that consciousness is exhausted by verbal reportability. We would like to thank Bill Lycan for this remark.

[10]This objection has been posed to us by K. Aizawa (personal communication).

[11]This must not be confused with undetected.

[12]We consider connectionism because it is the most contentious of both approaches. As Fodor and Pylyshyn (1988) point out, whether explicit rules ane necessary to explain cognition is a discussion that arises in the classicist camp. According to them, the symbolic systems permit explicit rules, though they do not need them; while networks do not even permit explicit rules. It is this last statement that we are going to judge.

[13]Another example is the animal experiment referred in the previous section: we must determine if the animal is paying attention to colour-related or shape-related information. Of course, it will not be always easy to frame such a "crucial experiment" to determine what information the system is accessing.

[14]Davies (1990) proposes an intermediate knowledge of rules between explicit representation of a rule, and mere conformity with a rule. This notion appeals to the idea of causally systematic processes. Our argument takes a different stance, but points at the same intermediate position.
[15]There is a third objection that might be interesting to consider. It is directed to undermine certain conception of inexplicit information in digital computers (Skokowski 1994). However, it is based on a compiling relation between high-level languages and machine-language, and it is dubious if it can be applied to connectionist systems.
[16]We do not mean that there are no differences between learning methods in connectionist systems. There are, for instance, important differences in the number, kind, and rate of presentation of instances to the network. What we mean is that all of them respond to the same general strategy: to obtain a central tendency (or prototype, for some authors) from particular instances.

References

Adams, F., Aizawa, K. and Fuller, G. (1992), 'Rules in Programming Languages and Networks.' in Dinsmore (1992), pp. 49–68.

Clark, A. (1993) *Associative Engines*, Cambridge, MA, MIT Press.

Clark, A. and Karmiloff-Smith, A. (1993), 'The Cognizer's Innards: A Psychological and Philosophical Perspective on the Development of Thought', *Mind and Language* 8, 487–519.

Cummins, R. (1986) Inexplicit Information. in M. Brand and R. Harnish (eds.), *The Representation of Knowledge and Belief*, Tucson, AZ, Univ. of Arizona Press. pp. 116–126.

Cummins, R. (1989) *Mental Representation*, Cambridge, MA, MIT Press.

Davies, M. (1990), Knowledge of Rules in Connectionist Networks, *Intellectica*, 9-10: 81–126.

Dinsmore, J. (ed) (1992), *The Symbolic and Connectionist Paradigms: Closing the Gap*, New Jersey, Erlbaum.

Fodor, J. and Pylyshyn, Z.W. (1988), Connectionism and Cognitive Architecture: A Critical Analysis, *Cognition*, 28: 3–71.

Hadley, R.F. (1993), Connectionism, Explicit Rules, and Symbolic Manipulation, Minds and Machines, 3: 183–200.

Hadley, R.F. (1995) The Explicit-Implicit Distinction. *Minds and Machines*, 5: pp. 219–242.

Karmiloff-Smith, A. (1992), *Beyond Modularity*, Cambridge, MA: MIT Press.

Karmiloff-Smith, A. (1994), Précis of *Beyond* Modularity, *Behavioral and Brain Sciences*, 17: 693–745.

Karmiloff-Smith, A. and Clark, A. (1993) What's Special about the Development of the Human Mind/Brain? *Mind and Language*, 8: 569–581.

Kirsh, D. (1990), When is Information Explicitly Represented?, in P. Hanson (ed.), *Information, Language and Cognition: Vancouver Studies in Cognitive Science* Vancouver: University of British Columbia Press, pp. 340–365.

Plunkett, K. (1993), Making Nets Work Hard. *Mind and Language*, 8: 549–558.

Raffman, D. (1993), *Language, Music and Mind*, Cambridge, MA: MIT Press.

Shanks, D.R. and St. John, M.F. (1994) Characteristics of Dissociable Human Learning Systems, *Behavioral and Brain Sciences*, 17: 367–447.

Skokowski, P.G. (1994), Can Computers Carry Content 'Inexplicitly'? *Minds and Machines*, 4: 333–344.

Smolensky, P. (1990), Tensor Product Variable Binding and the Representation of Symbolic Structures in Connectionist Systems, *Artificial lntelligence*, 46: 159–216.

Sperber, D. and Wilson, D. (1986), *Relevance: Communication and Cognition*, Oxford: Blackwell.

Sun, R. (1995), Schemas, Logics, and Neural Assemblies. *Applied Intelligence*, 5.