Cognition, 34 (1990) 93–107 **Discussion**

Autonomy, implementation and cognitive architecture: A reply to Fodor and Pylyshyn*

NICK CHATER
MIKE OAKSFORD
Centre for Cognitive Science, University of Edinburgh

Chater, N., and Oaksford, M., 1990. Autonomy, implementation and cognitive architecture: A reply to Fodor and Pylyshyn. Cognition, 34: 93–107.

1. Introduction

Fodor and Pylyshyn's (1988, this journal, henceforth "F&P") recent defence of the *Classical Symbolic Paradigm* against the emergent *Connectionist Paradigm* in cognitive science depends on the assumption that Connectionism eschews structural representation. However, this assumption is belied by the numerous attempts of Connectionists to implement structured representations in neural networks (Derthick, 1987; Hinton, 1981; Rumelhart, Smolensky, McClelland, & Hinton, 1986; Touretzky & Hinton, 1985). Thus, the issue of structured representation cannot be the principal point of disagreement between Classicist and Connectionist. We contend that although F&P are right to argue that Connectionism is an *implementational theory*, this does not detract from Connectionism's relevance to psychological explanation. F&P's contention that implementational considerations are irrelevant to psychological explanation only follows on the assumption that cognitive and implementational levels are computationally and hence explanatorily *autonomous* (F&P, p. 66). We argue that in attempting to account for the various alluring properties of Connectionist systems, F&P are systematically forced to abandon the autonomy assumption, thereby assuring the relevance of Connectionism to psychological explanation.

*We acknowledge a special debt of gratitude to Mike Malloch, who introduced us to PDP and has continued to be a major influence. Thanks are also due to Jerry Seligman and Robin Cooper for much fruitful discussion. We would also like to thank the members of the PDP Workshop at the Centre for Cognitive Science, University of Edinburgh, especially Richard Rohwer, Keith Stenning and David Willshaw. Nick Chater was funded by ESRC grant C00428622023; Mike Oaksford was funded under ESRC research contract R000231282. Requests for reprints should be sent to Nick Chater, Centre for Cognitive Science, University of Edinburgh, 2 Buccleuch Place, Edinburgh EH8 9LW, U.K.

2. The "lures" of Connectionism

We now re-examine the various "lures" of Connectionism. We argue that F&P's objections stem directly from their stand on autonomy. There are two readings of the claim that the cognitive level is implementation-independent.

The first reading is that the cognitive level can be formally specified. This formal specification must be implementable, but is wholly independent of the particular implementation employed. We can understand the behaviour of a PROLOG program independently of the layers of software on which it runs, and the hardware realisation in the VAX. That such independence of higher levels is *possible* is a central result in computability theory. Establishing the notion of a Universal Programmable Machine demonstrates that hardware places almost no constraints on the class of implementable virtual machines. The second reading is that the implementational level cannot affect higher-level processes. F&P are aware that this is an absurd position. Pylyshyn (1984) points out that implementation affects complexity profile, the effects of damage, reliability and so on. Since the second reading is agreed to be absurd the real nature of the dispute is whether the cognitive level can be formally specified in an implementation-independent way. The Classicist believes in formal autonomy; the Connectionist does not. F&P take formal autonomy to imply that implementation is irrelevant to cognition. To explain the lures we will consistently urge that the cognitive level must *interact* with properties of the implementation and so cognitive performance cannot be explained implementation-independently, *pace* formal autonomy. We will also argue that, in any case, implementational considerations severely constrain the class of cognitively plausible architectures, even if autonomy can somehow be preserved. Hence we will conclude that Connectionism *is* relevant to cognition.

The "lure" of connectionism consists of a series of properties shared by Connectionist devices and the human cognitive system. F&P argue that the lures are consistent with an appropriately implemented formally autonomous Classical architecture. By contrast the Connectionist is not convinced that this is possible. To resolve this issue we now discuss the lures in detail.

To begin, we must reclassify F&P's breakdown of the lures since certain phenomena are cross-classified. They provide the following list:

(1) Speed
(2) Content Addressability and Pattern Recognition
(3) The Blurring of Rule Governed and Rule Exceptional Behaviour
(4) Non-verbal or Intuitive Processes
(5) Resistance to (i) Damage and (ii) Noise

(6) Active versus Passive Storage
(7) All or None Processing, including:
 (i) Partial Rule Matching
 (ii) Non-determinism
 (iii) Graceful degradation
(8) Brain Style Modelling

Some of these issues cluster together. Connectionist approaches to massively parallel soft constraint satisfaction purchase the alluring properties of graceful degradation (7 iii), content addressability (2) and a property F&P do not mention, *automatic generalisation*. F&P group noise and damage tolerance (5) together, but ignore the former. Noise tolerance and partial pattern recognition (7 i) are special cases of graceful degradation. Damage tolerance and rule governed and rule exceptional behaviour (3) will be dealt with separately. We thus invoke five clusters:

(i) Speed (F&P's 1)
(ii) Tolerance of Damage (F&P's 5 i)
(iii) Massively Parallel Soft Constraint Satisfaction (F&P's 2, 5 ii, 7 i, 7 iii)
(iv) The Blurring of Rule Governed and Exceptional Behaviour (F&P's 3)
(v) Brain Style Modelling (F&P's 8)

Some of the issues which F&P raise are peripheral to Connectionism. The distinction between active and passive memory (6) concerns whether the control regime is completely distributed throughout the system (active, no CPU, no interpreter) or completely centralised (passive, CPU and interpreter). It is not about memories "doing" or "not doing" things. Thus, their discussion of Kosslyn and Hatfield (1984) is not germane (pp. 52-53). While many connectionist systems possess active memory in this sense, some do not (e.g., Shastri, 1985; Derthick, 1987).

Non-verbal and intuitive processing (4) are not addressed by F&P, and so we will not discuss them further. We do not know the origin of the alleged "lure" of non-determinism (7 ii). The macroscopic non-determinism of human behaviour seems equally compatible with Classicism or Connectionism.

2.1. Speed

It has been argued that there is an upper bound of about 100 serial steps on any cognitive process lasting less than a second (Feldman & Ballard, 1982). The Connectionist claims that the Classicist cannot account for this fact. However, F&P contend that "All [The hundred-step constraint] rules out is the (absurd) hypothesis that cognitive architectures are implemented in the brain in the same way as they are implemented on electronic computers" (p. 55). Yet the hundred-step constraint *does* severely limit the class of cognitively plausible algorithms. For example, it rules out this algorithm for addition: subtract 1 from the second number and add 1 to the first until the second is 0; the sum is the final value of the first. We suspect that most people can add 1,000,000 and 1,000,000 in less than 1 second. However, our algorithm would require 1,000,000 steps. It is excluded by the hundred-step constraint.

F&P suggest that since "... it is not even certain that the firing of neurons is invariably the relevant implementation property ... the 100 step 'constraint' excludes nothing" (p. 55). So, perhaps we can push up n (the number of steps that can be computed in less than a second). Connectionists themselves have suggested how this may be achieved by probabilistic coding (e.g., Hinton & Sejnowski, 1986) and have speculated that fast neuronal changes other than firing may be computationally important (von der Marlsberg & Bienenstock, 1986). Only by *doing PDP* will we discover what neural properties are computationally relevant. At best F&P may hope that n may be significantly greater than 100. However, even if n is higher by one or two orders of magnitude, the n-step constraint will still significantly restrict the class of cognitively plausible algorithms. The substantive practical issue is whether the Classicist can implement his favourite cognitive algorithms without violating the n-step constraint. Since Classicist algorithms typically require many millions of machine operations, the n-step constraint presents a non-trivial challenge to the Classicist.

2.2. *Tolerance of damage*

The Connectionist claims that Classical symbolic computation does not have the damage tolerance characteristic of human cognition. Connectionist systems achieve damage tolerance by using distributed representations. However, F&P argue that "... neural distribution of representations is just as compatible with Classical architectures as it is with Connectionist networks" (p. 56). All the support they adduce for this claim is that "... *all you need* [our italics] are memory registers that distribute their contents over physical space" (p. 56). However, distributed representations are damage tolerant not simply because they have spatially non-localised coding but because the internal structure of the symbol reflects its semantic properties, for example PEACH and APRICOT will have similar bit vectors. Arbitrary redundancy is like storing the same piece of information in many places. If all copies are damaged, however partially, there is a catastrophic loss of performance. In a non-arbitrary distributed scheme, similar objects are represented by similar bit vectors. Therefore, even if the representation is so damaged that PEACH cannot be reconstructed, a semantically related item will be selected, for example APRICOT (Hinton, McClelland, & Rumelhart, 1986: p. 102). This permits an understanding of graded semantically systematic error.

2.3. *Massively parallel soft constraint satisfaction*

This lure falls under four subheadings: (a) memory is content addressable and pattern recognition easy; (b) memory is noise resistant; (c) rules can be partially satisfied giving (d) graceful degradation. These issues are distinguished in F&P, but are all direct consequences of the Connectionist approach to massively parallel soft constraint satisfaction. Connectionists contend that standard symbolic computation does not have these properties. To maintain autonomy F&P must believe that an appropriate implementation of a Classical architecture can capture (a) to (d).

In F&P's reply to the lures, they say nothing about (a). So arguments that conventional methods such as "hash coding" are inadequate remain unchallenged (Hinton et al., 1986).

F&P identify the problem of noise (b) to be tolerance of "spontaneous neural activity" (p. 52). However, it is usually seen as the problem of achieving reliable computation with a degraded input. The input may be degraded for many reasons: for example, noisy background conditions (in ordinary speech recognition); errorful memory retrieval cues (Hinton et al., 1986); stimuli in peripheral vision; internal noise, whether generated by spontaneous neural activity or physical damage. The section "*Resistance to noise and physical damage*" does not mention noise but it *is* raised briefly in the discussion of "soft" constraints: "The soft or stochastic nature of [Classical] rule-based processes arises from the *interaction* [our italics] of deterministic rules with real-valued properties of the implementation, or with noisy inputs or noisy information transmission" (p. 58). This does not constitute an autonomous solution to the problem of noise since interaction between implementation and cognitive architecture simply concedes autonomy.

Similarly, in discussing (c), F&P seem to immediately concede autonomy: "One can have a Classical rule system in which the decision concerning which rule will fire resides in the functional architecture and depends on continuously varying magnitudes [thus abandoning autonomy]. Indeed, this is typically how it is done in practical "expert systems" which, for example, use a Bayesian mechanism in their production-system rule-interpreter" (p. 58). *Contra* F&P the statistical processes of Bayesian inference in practical expert systems are defined at the level of cognitive architecture, not functional architecture (Charniak & McDermott, 1985: p. 460). So the traditional solution is an autonomous Classical model. This approach seems unpromising in the light of complexity results for standard Bayesian techniques (Charniak & McDermott, 1985: Chapter 8). So the concession to non-autonomy is well motivated.

Autonomy is again apparently conceded in discussing (d). F&P argue that a Classical rule system may capture graceful degradation: "... rules could be activated in some measure depending on how close their conditions are to holding. Exactly what happens in these cases may depend on how the rule-system is *implemented* [our italics]" (p. 58). To have the activation of a cognitive level rule dependent on properties of the implementation completely abandons autonomy. F&P suggest that it is possible *in principle* that some implementation of Newell's (1969) hierarchy of weak methods or Laird, Rosenbloom, and Newell's (1986) universal sub-goaling may capture graceful degradation. But the Connectionist doubts that it can be achieved *in practice* in an autonomous architecture. In contrast, many Connectionist systems have been held to achieve graceful degradation quite naturally (McClelland, Rumelhart, & Hinton, 1986). F&P deny this claim (pp. 58–59) but in the absence of specific criticisms we must assume that the lure of graceful degradation stands.

Although F&P separate the above points, they are all direct consequences of the fact that PDP systems "provide an efficient way of using parallel hardware to implement best-fit searches ... Each active unit represents a 'microfeature' of an item, and the connection strengths stand for plausible 'microinferences' between microfeatures. Any particular pattern of activity of the units will satisfy some of the microinferences and violate others. A stable pattern of activity is one that violates the plausible microinferences less than any of the neighbouring patterns" (Hinton et al., 1986: pp. 80–81). Given this intuitive picture we can see how the various lures emerge.

When an arbitrary sufficiently large fragment of a pattern is presented the microinferences produce the nearest possible completion (stable state). Hence, any sufficiently large part of the content of the memory will access the whole memory. That is, memory is content addressable (a). If part of the presented fragment is wrong, the microinferences will still find the best available fit. Hence a degree of noise can be tolerated (b).

This intuitive picture can be generalised from *within* layer interactions to *between* layer interactions generating (c) and (d). Consider a network trained to map each of a set of input patterns onto a corresponding output pattern. We may treat each input–output pair as a rule with a "condition" (input) and an "action" (output). Each bit of the output pattern is a function of *all* the elements of the input pattern. Thus the information about which output should be chosen is distributed throughout the input. Suppose that the input is a slight distortion of one of the learnt patterns. Since the output is a function of the entire input, the loss of any particular part of the input does not lead to a catastrophic failure to produce any particular bit of the correct output. Rather it leads to a slight distortion of the whole output vector. Hence, as the fidelity of an input is smoothly reduced, the fidelity of the output smoothly reduces. This is *graceful degradation* (d). This behaviour contrasts with that of conventional schemes, where either the input error is detected, corrected for, and the right output chosen, or the error is not detected, and a totally inappropriate output is produced (or none at all).

Suppose that the presented input pattern is a blend of the learnt input patterns (suppose we have learnt A → X; B → Y; etc.; a blend of A : 1111100010 and B : 1110000001 might be simply C : 1111000011). What is the pattern Z that C is mapped onto? Since the presented input pattern C is a slight distortion of A, the output it produces, Z, is a degraded form of the corresponding output X (by graceful degradation). However, similarly, Z will be close to the output corresponding to B. Thus if the input C is a blend of A and B, the output Z will be a blend of X and Y. Thus an input pattern may partially match several different rules, to a graded extent. This is *partial rule matching* (c).

2.4 The blurring of rule governed and exceptional behaviour

F&P claim that Connectionists are committed to a common aetiology for rule governed and rule exceptional behaviour (p. 51). They appear to be adverting to Pinker and Prince's (1988) criticisms of Rumelhart and McClelland's (1986b) Past Tense Learning Model which attempts to learn regular and irregular past tenses with a single mechanism. From a detailed consideration of the past tense system in English, Pinker and Prince argue that the model is unlikely to generalise. However, F&P characterise the rule governed versus rule exceptional distinction in terms of the surely unrelated competence–performance distinction. The use of *went* as the past tense of *go* is to be attributed to linguistic competence yet it is rule exceptional. The Connectionist may wish to blur the aetiology of rule governed and rule exceptional behaviour while maintaining a sharp distinction between competence and performance.

F&P further conflate the rule-governed/exceptional distinction with the rule-implicit/rule-explicit distinction (pp. 59–60). While the issues surrounding the latter distinction are important and unresolved, we agree with F&P that they do not *decide* between Connectionist and Classicist. Since we advocate the implementation of high-level cognitive architectures in Connectionist hardware we are committed to the need for explicit rules in the explanation of cognitive phenomena. However, with respect to particular behaviours, the

Connectionist and Classicist may differ as to which sort of explanation is appropriate. If one retains autonomy, there are only two explanatory avenues open: rule governed or errorful. Linguistic exceptions are either simply *mistakes* or are governed by *explicit* exceptional *rules*. It seems that all regularities must be encoded explicitly. Later on this fact will return to dog the Classicist's attempts to model human reasoning.

2.5 Brain style modelling

F&P characterise this lure as the claim that Connectionist models, in contrast to Classical models, are constrained by the facts of neuroscience (pp. 53–54). However, they claim that biology constrains cognitive architecture very little (p. 62) and further that the biological plausibility of PDP models is in any case problematic (p. 58). We argue that there are good reasons to work with PDP models that do not map directly onto neural structures.

A given cognitive architecture running a particular algorithm will possess radically different real time processing characteristics when implemented in different hardware. So although biological hardware cannot determine the high-level architecture it severely constrains the class of possible cognitive architectures. It is hard to implement many features of standard architectures in PDP systems (Touretzky & Hinton, 1985, for example). Hence, it is unclear whether tolerably efficient implementations of any standard symbolic architecture are possible. Adversion to Turing machine power is of no avail here, since we are concerned with real time processing. We can only discover what cognitive architectures are compatible with biology by doing PDP computer science.

F&P hold that "brain style" modellers expect biology to *specify* properties of the cognitive architecture and counter that "the structure of 'higher levels' of a system are rarely isomorphic, or even similar, to the structure of the 'lower levels' of a system" (p. 63). This assumes that lower-level properties can only specify higher-level properties in virtue of an isomorphism. The relationship between atomic physics and chemistry seems to be an appropriate counter-example. In any case, the Connectionist need only claim that the facts of Biology *constrain* rather than *specify* cognitive architecture.

2.6 The lures: Concluding remarks

Finally, we present some general remarks which argue against F&P's response to the "lures" of Connectionism. They argue that the Classicist can deal with the lures *in principle*. To satisfy the Connectionist the Classicist will have to demonstrate this *in practice*.

The Classicist has work to do! F&P present no arguments to show that Classical architectures are compatible in practice with (2.1) speed; (2.2) tolerance of damage; (2.3) massively parallel soft constraint satisfaction; or (2.5) brain style modelling. The onus is on the Classicist to persuade us that a *single* Classical architecture/implementation can have all these properties.

The lures are desirable in standard computer science. If standard Classical architectures can be implemented to use just 100 steps, to tolerate hardware failure, to implement rapid noise resistant memory search and pattern matching, and to degrade gracefully under noisy conditions, then this is how they *should* be implemented. Such properties would be advantageous in everyday computational applications. If such implementations are possible, they cannot be obvious, or we would be running PROLOG on them!

Learning. The most persuasive lure is not mentioned by F&P. This is that Connectionist systems need not be handwired, but can *learn*. Current learning methods such as back-propagation are not the last word in learning theory. However, the PDP approach to learning gives some insight into how genuinely new structures can spontaneously emerge (Rumelhart & Zipser, 1986; Rumelhart, Hinton, & Williams, 1986; Pineda, 1987; Almeida, 1987; Hinton & Sejnowski, 1986; but see Minsky & Papert (1988) for a general critique).

By contrast, standard learning models cannot develop new structures (see Fodor, 1975, 1981), since Classical learning is just hypothesis generation and confirmation. Everything that can be learnt must be representable innately. Such considerations lead to the conclusion that all concepts (e.g., PROTON) are innate (Fodor, 1981; though see Chater, 1986). PDP promises a theory of learning which sidesteps these difficulties.

What was the real nature of the dispute? Throughout their discussion of the lures, F&P make no reference to structured representation. Thus even for F&P this issue does not bear on the dispute between Classicist and Connectionist. The lures challenge the Classicist to implement some standard architecture which meets each lure. We will only know whether this is feasible by attempting to implement standard architectures in brain-style hardware. And this will involve doing PDP computer science.

The Connectionist hunch is that this project will prove impossible and that many computational properties should be directly purchased from the implementational level. The lures do not decide the issue, but we have independent grounds to suppose that the Classicist project will prove unworkable. We argue that only by rejecting autonomy will we understand the computational characteristics of the mind.

3. Cognition as proof theory

F&P argue that Classical Cognitive Science amounts to "... an extended attempt to apply the methods of proof theory to the modelling of thought" (pp. 29–30). They seem to be proposing that we think in a high-level logic programming language (like PROLOG?) whose domain is the everyday world. Proof theory guarantees truth preserving inference. However, most everyday inferences are not guaranteed to preserve truth; that is, they are *plausible* inferences. These have been discussed under the banners of inductive inference, abductive inference and default inference.

3.1 Inductive inference

Classical inductive reasoning involves hypothesis generation and confirmation (Fodor, 1975). Hence, Classical inductive learning models, for example Winston (1977), can only learn new concepts by combining elements of an innate primitive basis. Fodor (1981) observes that the primitive basis may have to be as large as the lexicon of a natural language. Clearly the claim that, for example PROTON, is innate is close to a *reductio ad absurdum* of the Classicist theory of induction (but again see Chater, 1986).

3.2 Abductive inference

F&P (p. 58) observe that non-demonstrative inferences like abduction (inference to the best explanation) *may* be accommodated by supplementing proof theory with Bayesian inference techniques (cf. Charniak & McDermott, 1985). However, these are generally computationally intractable. In medical diagnosis, heuristic techniques are used to deal with multiple diseases (cf. Caduceus, in Charniak & McDermott, 1985: p. 474). These heuristics cannot be justified semantically within the formal system. For F&P, the heuristics must be implementational details, for example the search strategy of the interpreter. This amounts to computational non-autonomy. This implementational detail *explains* Caduceus' ability to deal with multiple diseases. This amounts to explanatory non-autonomy.

3.3 Default reasoning

Just about any everyday generalisation succumbs to indefinitely many counter-examples. If I see Fred going past my window at 9.00 a.m., I know he's about to buy his morning paper. But not if it's Christmas day, since there are no papers; not if he's being mugged; not if he's already reading *The Times*. These possibilities override our generalisation that Fred buys a paper just after passing my window every morning. To preserve autonomy, we must encode the various conditions that might override our rules, and check that none of them apply in any specific case. This is the standard approach to default reasoning in knowledge representation. Unfortunately we have reason to suspect that it is unworkable.

Most standard logical schemes are monotonic. If on seeing Fred go past the window I infer he will buy a newspaper, then if I reason monotonically no additional premise can invalidate my conclusion. In non-monotonic reasoning you can add premises and *lose* conclusions.

Reiter (1985) attempts to extend standard logic to incorporate non-monotonicity. McDermott (1986) notes that there are two problems with Reiter's approach. First, Reiter's logic is undecidable in principle and intractable in practice. Deciding whether a default rule applies involves consistency checking, which is an NP-hard problem. Second, the conclusions drawn are usually too weak. Although p is the conclusion desired, all that follows is $p \vee q$, where q is some arbitrary proposition.

This technical problem need not decide against autonomy in knowledge representation. However, there are more general difficulties for the Classicist approach. We can invent indefinitely many conditions which override my inference about Fred buying his morning paper. For the Classicist, each possibility must be explicitly encoded in the appropriate rule. To avoid an infinite list of default clauses we must appeal to a finite taxonomy which captures the infinitude of specific cases. Perhaps Fred will not get his morning paper in distracting situations, dangerous situations and so on. However, what counts as a distracting situation is relative to what rule we are considering. A road accident might count as a distracting situation for Fred's getting his paper, but not for his getting to work. So the categories in our taxonomy must be spelt out in detail *in each rule*. It is unlikely that such specifications will be forthcoming. This difficulty with the context dependence of categories is endemic in concept combination (Lyon & Chater, in press). The problem of defaults infects lexical inference as well as structural inference.

These problems with the Classical account of knowledge representation and inference do not argue for a non-autonomous account unless we indicate how the implementation can help. Hinton et al. (1986: p. 82) discuss an implementation of semantic nets in PDP hardware (originally in Hinton, 1981):

> If ... you learn that chimpanzees like onions you will probably raise your estimate of the probability that gorillas like onions. In a network that uses distri-

> buted representations, this kind of generalization is automatic. The new knowledge about chimpanzees is incorporated by modifying some of the connection strengths so as to alter the causal effects of the distributed pattern of activity that represents chimpanzees. The modifications automatically change the causal effects of all similar activity patterns. So if the representation of gorillas is a similar activity pattern over the same set of units, its causal effects will be changed in a similar way.

The similarity metric used in automatic generalisation is induced by pattern similarity and need not be specified by the programmer, but is learnt by the network (Hinton, 1987). Gorilla has "likes onions" as a default which may be overriden by explicitly storing information to the contrary. The default may also be overridden if "gorilla" has a similar pattern to "orangutan" and orangutans do not like onions. The similarity metric gives us default rules for free, and the autoassociative mechanism finds the best fit to the soft constraints. Soft constraints, the very fabric of PDP implementations, just *are* default rules. This is a paradigm case of the value of non-autonomous implementations of structured representations. This toy example is suggestive of how implementation may unburden the cognitive architecture of the problems created by non-demonstrative inference.

4. Psychology as proof theory

If the domain of psychology is a proof-theoretic cognitive level, then the following are apparently not psychology: Marr's (1982) models of low-level vision; Anderson's (1983) spreading activation models of semantic memory; any of the work on the capacity limitations of human memory (cf. Fodor, 1983); the whole of neurophysiology, neuropsychology and physiological psychology; all the work on semantic priming; the Trace model of speech perception (McClelland & Elman, 1986), etc. The only experimental work we know of which explicitly addresses the logical characteristics of the cognitive architecture is that on deductive reasoning (Wason & Johnson-Laird, 1972; Evans, 1982, 1983). On F&P's demarcation principle, the domain of psychological concern is unexpectedly limited. What remains is also problematic for the Classicist, since no existing logical regime is capable of capturing more than an insignificant fraction of the experimentally observed inferences (Oaksford, 1988; Oaksford & Stenning, 1988).

5. Conclusions

On a representational theory of mind the central problem for psychology is providing a semantics for mental states and a mechanism whose causal sequences are semantically coherent: that is, cognition is computation (Fodor, 1975, 1980, 1983, 1987; Fodor, Bever, & Garrett, 1974; Pylyshyn, 1973, 1980, 1984). F&P claim biological computations are autonomous; that is, mental processes are simply an implemented formal system and Cognitive Science is proof theory. F&P adduce evidence for structured representation and take this to decide against Connectionism because of the autonomy assumption. We believe that this assumption is the real locus of the dispute between Classicist and Connectionist. This diagnosis is borne out in the discussion of the lures, which provide empirical adequacy criteria on an autonomous architecture. It is unclear whether these criteria can be met without violating autonomy. Further, autonomous architectures may fail in principle to handle non-demonstrative inference. Admittedly, non-autonomous (Derthick, 1987) Connectionist approaches are embryonic. However, to borrow a Fodorian phrase, they seem to be the only straw afloat. So we must take seriously the possibility that cognitive architecture is not autonomous from its implementation.

References

Almeida, L.B. (1987). A learning rule for asynchronous perceptrons with feedback in a combinatorial environment. In *IEEE*, San Diego.

Anderson, J.R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Ballard, D.H. (1986). Cortical connections and parallel processing: Structure and function. *BBS*, *9*, 67–120.

Charniak, E., & McDermott, D. (1985). *Introduction to artificial intelligence*. Reading, MA: Addison-Wesley.

Chater, N.J. (1986). The present status of the innateness controversy: A reply to Fodor. Unpublished manuscript, Department of Psychology, University of Cambridge.

Derthick, M. (1987). *A connectionist architecture for representing and reasoning about structured knowledge*. Research report no. CMU-BOLTZ-29, Department of Computer Science, Carnegie-Mellon University, Pittsburgh.

Evans, J.S.B.T. (1982). *The psychology of deductive reasoning*. London: Routledge & Kegan Paul.

Evans, J.S.B.T. (1983). *Thinking and reasoning: Psychological approaches*. London: Routledge & Kegan Paul.

Feldman, J., & Ballard, D. (1982). Connectionist models and their properties. *Cognitive Science*, *6*, 205–254.

Fodor, J.A. (1975). *The language of thought*. New York: Thomas Crowell.

Fodor, J.A. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioural and Brain Sciences*, *3*, 63–109.

Fodor, J.A. (1981). The current status of the innateness controversy. In *Representations*. Cambridge, MA: MIT Press.

Fodor, J.A. (1983). *Modularity of mind*. Cambridge, MA: MIT.

Fodor, J.A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.

Fodor, J.A., Bever, T.G., & Garrett, M.F. (1984). *The psychology of language*. New York: McGraw Hill.

Fodor, J.A., & Pylyshyn, Z.W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3–71.

Hinton, G.E. (1981). Implementing semantic networks in parallel hardware. In *Parallel models of associative memory* (Ch. 6). Hillsdale, NJ: Erlbaum.

Hinton, G.E. (1987). Learning distributed representations of concepts. In *Eighth Annual Conference of the Cognitive Science Society*, 1987.

Hinton, G.E., McClelland, J.L., & Rumelhart, D.E. (1986). Distributed representations. In *Parallel distributed processing* (Vol. 1, Ch. 3). Cambridge, MA: MIT Press.

Hinton, G.E., & Sejnowski, T.J. (1986). Learning and relearning in Boltzmann machines. In *Parallel distributed processing* (Vol. 1, Ch. 7). Cambridge, MA: MIT Press.

Kosslyn, S.M., & Hatfield, G. (1984). Representation without symbol systems. *Social Research*, *51*, 1019–1954.

Laird, J., Rosenbloom, P., & Newell, A. (1986). *Universal subgoaling and chunking: The automatic generation and learning of goal hierarchies*. Boston: Kluwer.

Lyon, K., & Chater, N. (in press). Localist and globalist theories of concepts. In K.J. Gilhooly, M. Keane, R. Logie, & G. Erdos (Eds.), *Lines of thought*. Chichester: Wiley.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

McClelland, J.L., & Elman, J. (1986). Interactive processes in speech perception: The TRACE model. In *Parallel distributed processing, Vol. 2: Psychological and Biological models* (Ch. 15). Cambridge, MA: MIT Press (a condensed version of the *Cognitive Psychology* paper).

McClelland, J.L., Rumelhart, D.E., & Hinton, G.E. (1986). The appeal of parallel distributed processing. In *Parallel distributed processing: Exploration in the microstructures of cognition, Vol. 1: Foundations*. Cambridge, MA: MIT Press.

McDermott, D. (1986). *A critique of pure reason*. Technical report, Computer Science, Yale University, June.

Minsky, M., & Papert, S. (1988). *Perceptrons: An introduction to computational geometry* (2nd edn.). Cambridge, MA: MIT Press.

Newell, A. (1969). Heuristic programming: Ill-structured problems. In *Progress in operations research* (Vol. III). New York: Wiley.

Oaksford, M.R. (1988). *Cognition and inquiry: The pragmatics of conditional reasoning*. PhD thesis, Centre for Cognitive Science, University of Edinburgh.

Oaksford, M.R., & Stenning, K. (1988). *Process and pragmatics in reasoning with conditionals containing negated constituents*. Presented to the International Conference on Thinking, University of Aberdeen, August, 1988.

Pineda, F. (1987). *Recurrent backpropagation*. Technical report no. S1A-63-87, Applied Physics Laboratory, Johns Hopkins University, July, 1987.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*, 73–193.

Pylyshyn, Z.W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, *80*, 1–24.

Pylyshyn, Z.W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Montgomery, VT: Bradford.

Reiter, R. (1985). On reasoning by default. In *Readings in knowledge representation*. Kaufman.

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning internal representations by error propagation. In *Parallel Distributed Processing* (Vol. 1). Cambridge, MA: MIT Press.

Rumelhart, D.E., & McClelland, J.L. (1986a). *Parallel distributed processing: Exploration in the microstructures of cognition, Vol. 1 Foundations*. Cambridge, MA: MIT Press.

Rumelhart, D.E., & McClelland, J.L. (1986b). On learning the past tenses of English verbs. In *Parallel distributed processing* (Vol. 2). Cambridge, MA: MIT Press.

Rumelhart, D.E., Smolensky, P., McClelland, J.L., & Hinton, G.E. (1986). Schemata and sequential thought processes in PDP models. In J.L. McClelland & D.E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 2: Psychological and biological processes* (Ch. 14, pp. 7–57). Cambridge, MA: MIT Press.

Rumelhart, D.E., & Zipser, D. (1986). Feature discovery by competitive learning. In *Parallel distributed processing* (Vol. 1, Ch. 5). Cambridge, MA: MIT Press.

Shastri, L. (1985). *Evidential reasoning in semantic networks: A formal theory and its parallel implementation*. TR166, Department of Computer Science, University of Rochester, September, 1985.

Touretzky, D.S., & Hinton, G.E. (1985). Symbols among the neurons: Details of a connectionist inference architecture. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, University of California at Los Angeles, Los Angeles, 18–23 August, 1985, pp. 238–243.

Von der Malsburg, C., & Bienenstock, E. (1986), Statistical coding and short-term synaptic plasticity: A scheme for knowledge representation in the brain. In *Disordered Systems and Biological Organization* (Vol. F20). Berlin: Springer-Verlag.

Wason, P.C., & Johnson-Laird, P.N. (1972). *Psychology of Reasoning: Structure and Content*. Cambridge, MA: Harvard University Press.

Winston, P.H. (1977). Learning to identify toy block structures. In P.N. Johnson-Laird & P.C. Wason (Eds.), *Thinking: Readings in Cognitive Science* (Ch. 12, pp. 199–211). Cambridge: Cambridge University Press.