Systematicity: Psychological evidence with connectionist implications

Steven Phillips
Information Science Division
Electrotechnical Laboratory
1-1-4 Umezono, Tsukuba, 305, Japan
stevep@etl.go.jp

Graeme S. Halford
Department of Psychology
The University of Queensland
Brisbane, 4072, Australia
gsh@psy.uq.edu.au

Abstract

At root, the systematicity debate over classical versus connectionist explanations for cognitive architecture turns on quantifying the degree to which human cognition is systematic. We introduce into the debate recent psychological data that provides strong support for the purely structure-based generalizations claimed by Fodor and Pylyshyn (1988). We then show, via simulation, that two widely used connectionist models (feedforward and simple recurrent networks) do not capture the same degree of generalization as human subjects. However, we show that this limitation is overcome by tensor networks that support relational processing.

Distribution of cognitive behaviour

In the search for the essential components of cognitive architecture researchers have looked for concepts, phenomena and principles that help reduce potential candidates to, ultimately, a single architecture that explains cognitive behaviour. Systematicity is one such concept.

Systematicity is the property whereby cognitive capacities are grouped on the basis of common structure. For example, the ability to infer that "John went to the store" given that "John and Mary went to the store", extends to other structurally related inferences such as "Mary went to the store" given that "Mary and John went to the store". These two inferences share the common structure "P and Q implies P".

The concept of systematicity was introduced by Fodor and Pylyshyn (1988) to differentiate two candidate cognitive architectures: classical (symbol based) and connectionist (vector based) on the basis of their distribution of behaviours. In brief, their argument is that:

- Human cognitive behaviour is grouped on the basis of common structure (e.g., from above, it is not the case that one can do the first inference, but not the second).
- Classical architectures capture this grouping of behaviours by positing structure sensitive processes.
- Connectionist architectures, by specifying context-sensitive (structure insensitive processes), distribute behaviour irrespective of structure.
- Therefore, classical (symbol) systems are a better explanation for cognitive architecture, although connectionist architectures may provide suitable implementations of classical ones.

At issue here is not whether an architecture can ultimately exhibit all the observed stimulus-response behaviours, but how these behaviours are distributed over their available resources (e.g., learning trials). For example, an architecture based on simple associations requires two association steps (e.g., 1: A→B; 2: B→A) to support a bidirectional link between events A and B. By contrast, a relation based architecture only requires one step (e.g., R(A,B)), since bi(omni)directionality is built into relational operators (Phillips, Halford, & Wilson, 1995). The two architectures, although supporting the same functionality, distribute that functionality differently. The relevant difference is that there are states of associative based architectures for which representations of events are accessible in one direction, but not the other (e.g., after step 1, but before step 2). If one only ever observes bidirectional behaviour then such observations would be support for the relation based architecture, and not the association based architecture, although the former could be implemented by the latter[1].

Clearly, then, the root of the systematicity argument over cognitive architecture rests on the degree to which human cognition is systematic. Fodor and Pylyshyn take systematicity to be self-evident. Without recourse to specific data they claim, for example, that one can make inferences of the form $P \rightarrow Q, P \vdash Q$, if and only if one can make inferences of the form $Q \rightarrow P, Q \vdash P$. Subsequently, Hadley (1994) characterized systematicity as generalization to novel syntactic position, based on a review of language learning. Researchers have demonstrated networks supporting this definition of systematicity to various degrees (Christiansen & Chater, 1994; Hadley & Hayward, 1994; Niklasson & van Gelder, 1994; Phillips, 1994). However, others[2] question whether the empirical evidence supports this definition either way, given the difficulty of controlling subjects' background knowledge and observing what knowledge they have acquired in the course of an experiment. Furthermore,

[1] Analogously, whereas the architectural components of a database system are typically provided by fourth-generation languages such as SQL, such languages may be implemented in third-generation languages such as C, or Pascal. The point is, of course, that the sorts of behaviours exhibit by "paper-based" information processing systems are better captured (modelled) by relational languages like SQL, rather than procedural languages like C.

[2] Anonymous reviewer of Phillips (submitted).

in the domain of reasoning, van Gelder and Niklasson (1994) argue that the empirical evidence on logical inference (specifically, *modus tollens*[3]) does not support the Fodor and Pylyshyn's claim that human cognition is systematic. When thematic information is supplied in the premises, most subjects successfully perform the inference. Yet, when abstract information is supplied performance drops dramatically. A purely structure-based architecture would not predict this difference since the structure is the same in both cases.

The context-sensitive nature of cognition and its dependence on background knowledge make it difficult to quantify levels of generalization. Is it the case, for example, that no aspect of cognition is purely structure-based, in which case a classical picture is entirely wrong? Or, is it the interaction of two underlying factors (familiarity, and structure-sensitivity) that is being observed?

We point to recent psychological experiments by Halford, Bain, and Maybery (submitted) as evidence for the degree of systematicity consistent with the original claims of Fodor and Pylyshyn. When contextural and background information are controlled (by using materials of equally low association value and devoid of semantic content), subjects consistently make generalizations on the basis of the structural relationships between materials. We show that this purely structure-based generalization is difficult to achieve for two standard connectionists networks. Even though these networks demonstrate generalization, the degree of generalization is not the same as human subjects. However, we provide one alternative to the lack of generalization in terms of tensor networks that support relational processing.

Evidence: Relational schema induction

Recent psychological experiments by Halford et al. (submitted) demonstrated rapid induction and transfer of relational schemas by human subjects on a series of tasks sharing a common structure. In Experiment 1, four-task series were generated from the Klein 4-group using two operators: horizontal and vertical (see below). The task consists of four states and two operators, which map each state to a next state. When the states are depicted as vertices of a square, the two operators can be interpreted as horizontal and vertical transitions (Figure 1(a)). Each task instance consists of four randomly generated strings, and two shapes (corresponding to the horizontal and vertical transitions). Subjects are presented with a string and a shape, and asked to predict the response string. For example, in Figure 1(b), **PEJ** and △ predict **BIP**. In each trial, all eight possible string-shape pairs are presented one at a time in random order. After making a prediction, subjects are informed of the correct response. No reference is made to the structure and underlying meaning of the task. Learning within a task instance continues until all eight pairs are correctly predicted within a single trial, or to a maximum of six trials. After the intra-task learning criterion is reached, or after six trials, the next task instance is presented, until four task instances have been completed.

[3] The inference: if p implies q and not q then not p.

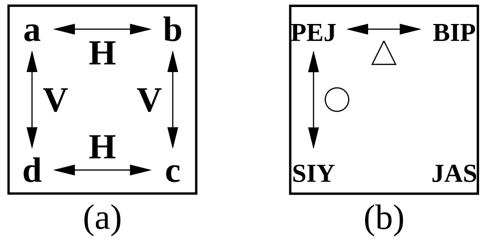

Figure 1: Klein 4-group (a) and task instance (b).

One of the most pertinent results to this paper is that first trial error rate on the fourth task was 2.00 for a group of 12 participants. Halford et al. considered two possible explanations for correct prediction on novel patterns. One can make correct predictions by interpreting elements of the new task as elements of a common task structure, or a previous task. For example, given stimulus pairs and their target responses (from Figure 1):

$$(\mathbf{PEJ}, \triangle) \rightarrow \mathbf{BIP}; \text{ and}$$
$$(\mathbf{SIY}, \bigcirc) \rightarrow \mathbf{PEJ},$$

one can make the following interpretations[4]:

(**PEJ**, **a**);	(△, **H**);
(**BIP**, **b**);	(◯, **V**);
(**SIY**, **d**);	and
(**JAS**, **c**).	

A third stimulus pair (**BIP**, ◯) can be predicted as **JAS** via the aligned elements and task structure. The three steps are:

1. **BIP** → **b** and ◯ → **V**;
2. (**b**, **V**) → **c**; and
3. **c** → **JAS**.

An alternative, non-structure based explanation is to observe that (1) a string never predicted itself, and (2) no string was paired with a shape more than once. Using this statistical knowledge the expected error is 3.34. However, the observed error rate was significantly lower than this value $t(23) = 1.89$, $p < .05$ (Halford et al., submitted).

These results are strong evidence for the sort of structure-based generalization claimed by Fodor and Pylyshyn. When materials are controlled for association value and semantic content (by using pronounceable, but otherwise meaningless three letter strings and shapes), subjects consistently reached the point (4th task instance) of making correct inferences on new tasks conforming to the same structure, independent of the materials used. What then is the support for the same degree of structure-based generalization in connectionist networks? We address this question by examining feed-forward networks on the Klein 4-group task used to test human subjects.

[4] The last interpretation (**JAS**, **c**) comes from the knowledge that they are the only remaining uninterpreted elements.

Connectionist properties

A common method for demonstrating generalization is to partition data into training and testing sets, where the inputs and outputs range over the same vector space. However, in the schema induction typically the stimulus and response materials do not appear in more than one task. Therefore, particular attention must be given to the way input and target vectors are represented in the network. One way is to use a different group of input/output units for each task, with learning transfer on the basis on common connections between hidden units (Figure 2(a)). Hinton (1990) used this style of network to demonstrate generalization between two isomorphic family trees. However, this approach is cumbersome for a longer series of tasks since it adds many additional weights and units that are only updated during one of the tasks. The approach we adopted was to use the same units and weights for each task, but to reset the input-to-hidden and hidden-to-output connections. This approach simulates the use of novel materials across tasks, while allowing knowledge transfer by not resetting the hidden-to-hidden unit connections (Figure 2(b)). Dashed arrows indicate the actual weights reset during simulations (see Method).

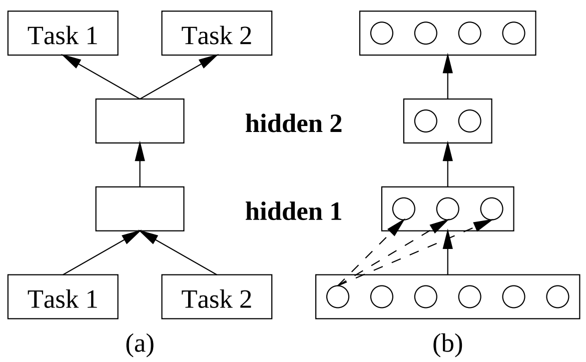

Figure 2: Feedforward network with different (a) and same (b) input/output units for each task.

Another consideration is the number of hidden layers (at least two) and the number of units within each layer. Importantly, the number of weights (free parameters) must be small enough to facilitate generalization, but large enough to support a solution. Preliminary simulations suggested a 6-3-2-4 network, where the 6 input units (4 states plus two operators) are connected to the first hidden layer of 3 units, connected to the second hidden layer of 2 units, connected to the 4 output units. For units with activation functions that form hyperplanes, it can be shown that 2 is the minimum number of units for the second hidden layer, and that 2 is a lower bound for the first hidden layer, under the condition of local input/output vectors (i.e., a single unit with activation 1 and the rest 0). Preliminary simulations with a 6-2-2-4 failed to learn all patterns, so the 6-3-2-4 network was used. Use of more hidden units only decreases the likelihood of generalization as it introduces more free parameters for the same number of examples.

Method

Preliminary simulations showed that the network failed to learn many of the patterns in the second task despite successfully learning all patterns in the first task. This, despite the fact that one of the solutions to the second task is the same set of weights learnt from the first task, some of which were already specified. The inability to learn the second task introduces a methodological problem: how to examine generalization when the network cannot find any of the available solutions to the training set. There are several ways to overcome this problem, for example: use more trainable weights; fix fewer weights common to both tasks; or, use more powerful learning methods. However, a failure to demonstrate generalization under these circumstances is always subject to the "what if you tried ..." response. Alternatively, one can take a upper bound approach by identifying the degree of generalization capable by the network given the most amount of information. If the network fails to meet the generalization criterion under this condition we can say that it cannot support systematicity as there is no further information available to the network.

Accordingly, we adopted the following procedure: (1) train the network to correct prediction on all patterns in the first task; (2) reset only those weights connected from the input unit corresponding to the test pattern for the second task; and (3) retrain the network with all other patterns, and all other weights and biases fixed. Figure 2(b) shows the weights (dashed arrows) reset for learning the second task. The networks were randomly initialized from a 0 mean 0.5 variance normal distribution, and updated using the standard backpropagation algorithm (Rumelhart, Hinton, & Williams, 1986) with squared difference between output and target patterns as the error function, and a learning rate of 0.1. Training continued until the average squared error for each output unit and pattern reached 0.01 training[5]. Networks were examined for transfer in the cases where 1, 2 and 3 weights from the input unit representing the missing stimulus pattern were reset. Since local representations (i.e., a single unit with activation 1, the rest 0) were used, this corresponds to retraining on all patterns for the new task except the single test pattern containing the input string represented by that unit. Results are reported for the second task instance.

Results

The network demonstrated generalization to the test pattern in 7 of 10 trials when only one weight was reset. There was no evidence of generalization when all 3 weights were reset. In 2 trials (with 3 weights reset) the output unit with the maximum activation corresponded to the target string. This rate of success is not better than chance (2.5), and the difference between the largest two activated units, in each trial, was only marginal (< 0.02). In both trials, the network only learnt 6 of the 7 training patterns. In the 4 trials where all 7 training patterns for the second task were learnt, there was no generalization to the test pattern.

[5] Sufficient for correct prediction on all training patterns.

Discussion

The most pertinent result is the 4 trials where all 7 training patterns were learnt without generalization to the single test pattern: since the region of weight space that is a solution to the 7 training patterns is larger than the region of space for all 8 patterns, there is no necessity that training will generalize to the test pattern. One way to reduce the difference between the two regions is to eliminate hidden units to reduce the number of weights (free parameters). However, as we have seen, further reduction in hidden units leaves the task unrepresentable. Therefore, the feedforward network in its standard form cannot be said to capture the property of systematicity as characterized by the empirical data.

The results do not preclude the feedforward network from demonstrating some degree of generalization across structurally isomorphic tasks. For example, Hinton (1990) showed the same type of network (with different numbers of hidden layers and units) exhibiting prediction on isomorphic family trees. However, the degree of generalization was extremely limited (after training on 100 of a possible 104 patterns, the network correctly predicted the remaining 4 test patterns), and in light of evidence discussed here, different to that of humans.

The limitation of the feedforward network on Klein 4-group tasks also extends to the simple recurrent network (Elman, 1990), which others (Christiansen & Chater, 1994; Niklasson & van Gelder, 1994) have shown exhibits (to some degree) the definition of systematicity put forth by Hadley (1994). A simple recurrent network applied to the Klein 4-group task includes all the weights and units of the feedforward network, plus additional weights for mapping the context (hidden unit vector from the previous time step) to the hidden units. Since the additional weights in the simple recurrent network make the solution even less constrained than the feedforward network it is highly unlikely to exhibit generalization for the Klein 4-group task.

Extensions

As well as connectivity, activation and error functions also determine the shape of the error surface. Given the possibly infinite variations one can make in terms of these architectural components it is not feasible to canvas all possibilities. However, we can identify specific properties and examine their capacity to support systematicity.

The effective weight space of the network can be restricted by enforcing fewer activation states for its units (e.g., binary, rather than real valued). In the extreme case, the smallest number of identifiable states for the second hidden layer is 4 (i.e., one state for each possible response). Fewer states means that at least one state must be mapped to two different responses. Suppose, also, that these 4 states are supported by a single hidden unit. This condition requires nonmonotonic activation functions (e.g., gaussian, pulse) at the output layer, since as mentioned above it is not possible to partition 4 points on a line into using single threshold functions. However, gaussian functions for example have two thresholds permitting each point to be separated from every other point.

Under these conditions, how many patterns are required to learn a new task? The lower bound is 4 (i.e., one pattern for each possible response), since each output unit requires training on at least one pattern for which the correct response corresponds to that unit. Assuming we have the appropriate architectural elements (e.g., connectivity, activation and error functions) to enforce such a representation are these components sufficient to support systematicity? We must conclude *No*, since subject response rate is 2.00 (over 12 subjects), which is significantly less than 4. Therefore, such extensions do not support systematicity as defined by the empirical evidence.

One limiting property of the feedforward network is the unidirectional nature of its flow of information. Learning the links between internal representations of the task structure and current task elements proceeds in one direction only: from stimulus to internal representation and then from internal representation to response. These two directions are independent. The consequence is that learning to map a task element as a stimulus to a suitable internal representation (i.e., alignment) does not permit the related generalization of mapping the aligned internal representation to the same task element, but treated as a response.

An obvious extension to the feedforward network is to introduce bidirectional weights and linear units so that the interpretation and its inverse are learnt concurrently. However, Experiment 5 (Halford et al., submitted) showed that subjects also have the ability to predict the missing operator (shape) given the initial and final states (strings). Simple bidirectional links would not support this kind of generalization, since each state is equally associated with both operators.

In general, omnidirectionality between related elements is supported by tensor networks. We can characterize a branching point in the evolution of connectionist networks; one which turns on the property of directionality, which we consider as an additional requirement for systematicity. In the next section, we outline an alternative approach to capturing systematicity: the use of tensor networks that support relational processing.

Support for relational processing

An alternative approach to the continued extensions to the feedforward network is to embody the properties of relations into connectionist networks. Our purpose here is not to provide a complete model of the data from the relational schema induction experiments, but to show how connectionist (tensor) networks capture some essential properties pertaining to systematicity.

Relational architecture

The Klein 4-group task can be supported by a relation-based architecture that assumes relational data structures (i.e., sets of ordered tuples) and processes for accessing relational elements. A relation, as the concept has been adopted for cognition (Halford, Wilson, &

Phillips, submitted; Phillips et al., 1995), consists of a schema (symbol that identifies the name of the relation and its arguments, or roles) and a set of ordered tuples (instances of the relation). In addition, there are the operators: *select*: $\sigma_c R \rightarrow R'$, which returns the relation R' containing all tuples from R satisfying condition c; and *project*: $\pi_a R \rightarrow R'$, which returns all tuple elements from R at argument position a.

For example, given the Klein 4-group as the relation: $R_k(I, O, F) = \{(a, H, b), (a, V, d), ...\}$, then $\pi_F \circ \sigma_{\langle I=a \rangle, \langle O=H \rangle} R_k \rightarrow b$, where $\circ$ is the composition of two operators. That is *select the instances from the Klein relation such that the elements in the initial state (I) and operator (O) positions are a and H (respectively), and project out the element in the final state (F) position of the selected instance.* Since the project and select operators often appear in pairs, we simplify the notation to: $R_k(a, H, -) \rightarrow b$.

Suppose relations $R_k(I, O, F)$ and $R_t(I, O, F)$, representing the Klein 4-group structure and current task (respectively); and $R_i(T, S)$, representing the interpretation between task and structure elements. From the example pertaining to Figure 1, where the first two presented patterns are: $(\mathbf{PEJ}, \triangle) \rightarrow \mathbf{BIP}$, and $(\mathbf{SIY}, \bigcirc) \rightarrow \mathbf{PEJ}$, the response to the third stimulus pair: $(\mathbf{BIP}, \bigcirc)$ is predicted as follows:

1. $R_i(\mathbf{BIP}, -) \rightarrow \mathbf{b}$, $R_i(\bigcirc, -) \rightarrow \mathbf{V}$;
2. $R_k(\mathbf{b}, \mathbf{V}, -) \rightarrow \mathbf{c}$; and
3. $R_i(-, \mathbf{c}) \rightarrow \mathbf{JAS}$.

Although, elements $\mathbf{c}$ and $\mathbf{JAS}$ do not appear in the first two patterns it is assumed that they are added to the interpretation relation R_i as they are the only remaining uninterpreted elements.

Tensor architecture

Based on the work of Smolensky (1990), Halford, Wilson, Guo, Gayler, Wiles, and Stewart (1994) showed how tensor networks can support relations as the sum of the tensor outer products of vectors representing each tuple element. A rank n tensor (T^n) is constructed by taking the outer product[6] of vectors representing each tuple element of an n-ary relational instance. For example, the ternary relational instance $(\mathbf{b}, \mathbf{V}, \mathbf{c})$ is represented by the rank 3 tensor $T^3 = \vec{b} \otimes \vec{V} \otimes \vec{c}$.

In addition, the relational operator pair project-select $(\pi \circ \sigma)$, as we have used them, corresponds to the tensor inner product[7] (Halford et al., submitted; Phillips et al., 1995). For example, $R_k(I, O, F) = \{(\mathbf{b}, \mathbf{V}, \mathbf{c}), \ldots\}$ corresponds to the tensor $T_k = \vec{b} \otimes \vec{V} \otimes \vec{c} + \ldots$, under the assumption that element vectors are mutually orthonormal. Therefore, assuming tensors T_k and T_i, corresponding to relations R_k and R_i (respectively), the third stimulus pair is predicted as follows:

1. $B\vec{I}P \odot \vec{T_i} \rightarrow \vec{b}$, $\vec{\bigcirc} \odot \vec{T_i} \rightarrow \vec{V}$;
2. $\vec{b} \otimes \vec{V} \odot \vec{T_k} \rightarrow \vec{c}$; and
3. $\vec{T_i} \odot \vec{c} \rightarrow J\vec{A}S$.

[6] $S^m \otimes T^n = (S_{i_1 \ldots i_m} . T_{j_1 \ldots j_n})$.

[7] $S^m \odot T^{m+n} = (\sum_{i_1 \ldots i_m} S_{i_1 \ldots i_m} . T_{i_1 \ldots i_{m+n}})$.

Thus, a tensor based network captures the same degree of generalization as subjects at the 4th task by embodying some of the properties of relational systems. Those properties are omnidirectional access to relational elements, and representation of the relational structure via groups of units dedicated to particular relational arguments (roles).

The omnidirectional property observed in Experiment 5 is supported in relational systems by the project and select operators. For example, having learnt the stimulus-response pair $(\mathbf{PEJ}, \triangle) \rightarrow \mathbf{BIP}$ from a single presentation, subjects consistently inferred that $\triangle$ is the missing shape resulting in the transition from string **PEJ** to string **BIP**. In a relational system, the single stimulus-response pair is added as the triple (**PEJ**, $\triangle$, **BIP**) to the relation $R_t(I, O, F)$. Each of the three elements are accessed as: $R_t(\mathbf{PEJ}, \triangle, -) \rightarrow \mathbf{BIP}$; $R_t(\mathbf{PEJ}, -, \mathbf{BIP}) \rightarrow \triangle$; and $R_t(-, \triangle, \mathbf{BIP}) \rightarrow \mathbf{PEJ}$.

The corresponding tensor operations supporting omnidirectionality are: $P\vec{E}J \otimes \vec{\triangle} \odot T_t \rightarrow B\vec{I}P$; $P\vec{E}J \odot T_t \odot B\vec{I}P \rightarrow \vec{\triangle}$; and $T_t \odot \vec{\triangle} \otimes B\vec{I}P \rightarrow P\vec{E}J$.

Where do we stand now?

This paper was motivated by comments from an anonymous reviewer of Phillips (submitted), who questioned whether it is possible to determine the systematic nature of human cognition given the difficulty of observing a subject's internal (representational) state, and their sensitivity to contextual information. Our response has been to point to psychological experiments showing evidence of generalization on the basis of common relations between the stimulus materials, not on the basis of the contents of those materials.

Our subsidiary points concern the connectionist properties that do/don't support the same degree of systematicity as human subjects. The degree of generalization exhibited by human subjects places strong requirements on (connectionist) models of cognition. Those requirements are not captured by specifying standard feedforward or recurrent networks, despite the fact that these networks demonstrate generalization in other domains. We point to tensor networks that support relational processing as one property that supports systematicity as measured by one set of empirical studies.

We do not claim that all of cognition is as equally systematic, only that at least one (significant) part is. Furthermore, the negative results with respect to feedforward and recurrent networks do not rule out these architectures as interesting candidates for other aspects of cognition. For example, McClelland (1995) has demonstrated that the feedforward network captures the important *torque difference* effect in the development of balance-scale. However, we do stress that in light of other evidence, such models will not provide the whole story (at least not without significant extensions).

Now that we have separated out two factors influencing the distribution of cognitive behaviour (i.e., structure based, as pointed out here, and familiarity based, as pointed out by van Gelder & Niklasson, 1994), the question remains as to how to put the two back together under a single architecture. We have briefly outlined two directions: (1) continue the extension of existing models by including additional structural constraints (e.g., bidirectionality, etc), and (2) start with connectionist networks (e.g., tensors) with properties isomorphic to classical (e.g., relational) systems and progressively integrate the context-sensitive properties of other networks. Some work is being done in this direction (Phillips, submitted), although further work is required.

Finally, it should also be noted that the tensor network (as we have presented it here) is not, specifically, a claim for connectionism as an alternative to classicism as a theoretical basis for cognitive behaviour. In fact, as our use of the tensor network was designed to support relational processes, it can be regarded as an implementation of a classical (relational) system. Whether, in fact, connectionism does provide an alternative theoretical basis is still debated (Fodor, 1997; Smolensky, 1995), and given the extensive literature on this issue (see Fodor & Pylyshyn, 1988; Smolensky, 1988; Fodor & McLaughlin, 1990; van Gelder, 1990, among others), it would be inappropriate to address it here. Nevertheless, alternative or not, the problem of determining what connectionist properties support systematicity remains.

Acknowledgements

We thank Yoshiro Miyata for the PlaNet simulator. The first author was supported by an STA fellowship.

References

Christiansen, M. H., & Chater, N. (1994). Generalization and connectionist language learning. *Mind and Language*, *9*(3). Preprint.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Fodor, J. A. (1997). Connectionism and the problem of systematicity (continued): Why Smolensky's solution still doesn't work. *Cognition*, *63*(1), 109–119.

Fodor, J. A., & McLaughlin, B. P. (1990). Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work. *Cognition*, *35*, 183–204.

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

Hadley, R. F. (1994). Systematicity in connectionist language learning. *Mind and Language*, *9*(3), 247–272.

Hadley, R. F., & Hayward, M. (1994). Strong semantic systematicity from unsupervised connectionist learning. Tech. Rep. CSS-IS TR94-02, Simon Fraser University.

Halford, G. S., Bain, J. D., & Maybery, M. T. (submitted). Induction of relational schemas: Common processes in reasoning and learning set acquisition.. http://www.psy.uq.edu.au/Department/Staff/gsh.

Halford, G. S., Wilson, W. H., Guo, J., Gayler, R. W., Wiles, J., & Stewart, J. E. M. (1994). Connectionist implications for processing capacity limitations in analogies. In K. J. Holyoak & J. Barnden (Eds.), *Advances in Connectionist and Neural Network theory*, chap. 7. Norwood, NJ: Ablex.

Halford, G. S., Wilson, W. H., & Phillips, S. (submitted). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology..

Hinton, G. E. (1990). Mapping part-whole hierarchies in connectionist networks. *Artificial Intelligence*, *46*(1-2), 47–76.

McClelland, J. L. (1995). A connectionist perspective on knowledge and development. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modelling*, chap. 4, pp. 157–204. Hillsdale, NJ: Lawrence Erlbaum Associates.

Niklasson, L., & van Gelder, T. (1994). Systematicity and connectionist language learning. *Mind and Language*, *9*(3). Preprint.

Phillips, S. (1994). Strong systematicity within a connectionist framework: The tensor-recurrent network. In A. Ram & K. Eiselt (Eds.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, pp. 723–727. Lawrence Erlbaum.

Phillips, S. (submitted). Systematicity and the nature of internal representations..

Phillips, S., Halford, G. S., & Wilson, W. H. (1995). The processing of associations versus the processing of relations and symbols: A systematic comparison. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*, pp. 688–691.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning Internal Representations by Error Propagation*, Vol. 1 of *Computational models of cognition and perception*, chap. 8. Cambridge, MA: MIT Press.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11*, 1–74.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, *46*, 159–216.

Smolensky, P. (1995). Connectionism, constituency and the language of thought. In C. Macdonald & G. Macdonald (Eds.), *Connectionism: Debates on psychological explanation*, Vol. 2. Cambridge, MA: Blackwell.

van Gelder, T. (1990). Compositionality: A connectionist variation on a classical theme. *Cognitive Science*, *14*, 355–384.

van Gelder, T., & Niklasson, L. (1994). Classicism and cognitive architecture. In A. Ram & K. Eiselt (Eds.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, pp. 905–909. Lawrence Erlbaum.