Book review published in Neural Networks 9(3), 545-548, 1996.

Artificial Intelligence and Neural Networks: Steps toward Principled Integration
Edited by Vasant Honavar & Leonard Uhr, Academic Press, Boston, MA: 1994, $89.95, 653 pp., ISBN 0-12-355055-6.

Are Symbolic Artificial Intelligence (SAI) and Neural Network (NN) approaches toward building intelligent systems complementary? Are they two competing theories trying to arrive "there" first? Do they have such irreconcilable differences that it is a waste of effort trying to forge them together? Do we have a firm theoretical basis for "principled integration"? These are some of the questions that the book sets out to address. The book presents some thought-provoking discussion on all these issues and leaves the reader with enough avenues ('leads') to search for answers. It is a useful compendium of issues and results by some of the leading researchers in the field and in several directions: basic issues in symbol versus connectionist paradigms, representation and inference, vision, language, and learning. As always happens with a book that reports on a fast growing field such as Neural Networks, some of the material becomes dated by the time the book sees the light of publication. In the review we will point out (where we are aware of) some of the latest results. To make the review more interesting we start with a discussion of general issues in the integration of symbolic AI and NN. Specific comments on some of the chapters are organised along the lines of division in the book. We conclude with more general comments.

Newell and Simon's (1980) symbol system hypothesis is a formal statement of the theoretical underpinnings of the AI enterprise. They claimed that human intelligence entails a symbol system and a set of procedures to manipulate these symbols. Combining this idea with Fodor's Language of Thought hypothesis (Fodor, 1976), a typical AI-er tends to reduce human thought (intelligence) to logic, rules, knowledge, and symbol processing. Searle's (1980) "Chinese room problem" highlighted the inadequacy of symbolic AI in explaining human cognition. Harnad (1990) argued that the woes of symbolic AI relate to the "ungrounded" nature of the atomic symbols. He pointed out that the initial set of symbols, being arbitrary, is flexible enough to support compositionality (atomic symbols can be combined and molecular representations can be decomposed according to a formal syntax) and systematicity (the atomic and molecular symbols and the rules of syntax can be systematically assigned a meaning). He asserted however, that the very fact that these symbols are arbitrary, means they become "ungrounded" (their meaning and interpretation reside in the subject's 'head' and are otherwise 'rootless') and give credence to Chinese room-like arguments.
Harnad's solution calls for grounding the categorisation in motor and sensory perception. In other words the atomic symbols are grounded in sensori-motor perception and not arbitrary. We feel that this grounding problem, when addressed properly, would give deeper insight into the feasibility and appropriateness of integration of symbolic and connectionist models of human cognition. Integration then amounts to explaining how intentional reasoning arises from these primitive grounded perceptions. In addition, Aparicio & Levine (1994) stressed that emotion, mood, motivation, and instinct have a greater role to play in perception and cognition than logic alone. Classical AI has always focused on developing models of "what we know" but for a true science of cognition (mind), we also have to answer "how do we know what we know". The thrust of biologically inspired network studies is to address these issues. In the following summary, we have selected only some of the chapters from the book that, in our opinion, have made fruitful contributions on the issues discussed above.

Symbol Processors versus Connectionist Networks:
The editors in their introductory chapter propose a notion that neural nets and symbolic AI are equivalent using the universal Turing machine argument. In our view, arguments of this nature are suitable for general discourse but do not highlight principles for integration. They identify two terms that are in use: 'integrated' and 'hybrid' systems. Integration refers to the gleaning of principles of information processing from both the AI and NN streams and building systems that utilise them. The term hybrid usually refers to the use of NN for some parts of the process and SAI for others. One popular hybrid system comprises a neural net front-end to a rule based expert system. Authors argue rightfully that integrated systems are more desirable in the long run.

In this introductory section, the oppposing forces of Good Old Fashioned AI (GOFAI) and NN are ranged against each other, with Boden and Chandrasekaran & Josephson arguing strongly for logical models of intelligence which rest on the ability to apply knowledge. Arbib, on the other hand, presents the first attempt at integration describing the merging of NN and his own schema theory. He also identifies the key idea of connectionism as the need to understand how local interactions can lead to an overall result without explicit executive control, and places great store on "emergent" behaviour. Chandrasekaran & Josephson (also the mainstream AI researchers) assume separability of intelligence from affective mental states. However, it is common knowledge in neuroscience (for example, Damasio, 1991) that emotion, mood, and motivation colour human perception. In fact, Newell himself concedes in an interview with Chandrasekaran that a theory of emotion needs to be included in the scope of his Unified Theories of Cognition for it to be complete (Chandrasekaran, 1993). Neural net researchers have started exploring these ideas seriously (Levine & Leven, 1992).

Included in this section is a paper which we see as representing another distinct strand in AI research, namely that of computational neuroscience. McKenna argues strongly for a firmer biological basis for artificial models of intelligence, presenting a number of significant success stories to back his claims, and rejects the idea of relying on a merging of NN and SAI to solve problems.

Having presented the reader with these contrasting views, the focus of attention in the remainder of the book appears to switch to the implementation of traditional AI tasks within a connectionist framework.

Knowledge Representation and Inference:

The majority of papers within this rather long section appear to be concerned with addressing Fodor & Pylyshyn's (1988) arguments against connectionsim, namely that compositionality and structure sensitivity are necessary to support cognitive processes and that only symbolic representations are capable of doing so. Sharkey & Jackson argue that these claims reduce to the need to show that connectionist systems are capable of systematicity, compositionality, and grounding. In their paper they demonstrate elegantly the ability of a connectionist system to exhibit systematicity. They consider an example from natural language implemented on a feedforward network with two layers of weights. They argue that the input-to-hidden unit weights (IH-weights, the fan-out from an input unit) enable *context independent* representation of input symbols. After successful training (IH-weights are fixed), these weights determine the extent of participation of an input symbol, independent of the context. In contrast the hidden unit representation is *context dependent*. The fan-in weights to each hidden unit determine the contribution of that hidden unit to the final outcome and this fan-in is determined by the context of the co-occurrence of input symbols. In addition to this interpretation, they propose a novel mechanism of decomposing these fully distributed context-dependent representations into the representations of their constituent parts. With these demonstrations the authors not only address most of the criticisms of connectionist architectures but also exhort the connectionist-camp not to give up hope and settle for premature hybridisation, in the sense defined in the previous section.

van Gelder & Port explore issues in compositionality and identify the importance of dynamic representations. They identify three dimensions each for properties of tokens of primitives and properties of manner of combination. Tokens can be classified as static/dynamic, analog/digital, and aribtrary/nonarbitrary. Manner of combination can be classified into concatenativity, static/temporal, and syntactic conformity. Printed English text, for example, has tokens that are static, digital, and arbitrary, and their manner of combination is concatenative, static, and relatively strict. Whereas, spoken English has dynamic and analog tokens and context sensitive, temporal and loose mode of combination. The authors call attention to the latter type of tokens that require *dynamic* compositionality and mention that classical AI has concentrated on the tokens of the former type that require *static* compositionality. They illustrate an example of dynamic compositionality with a neural network that processes speech sequences.

Barnden discusses the representation of beliefs, or intentionality by connectionist systems and concludes that the need for embedded reasoning may move connectionists closer towards symbolic approaches than they may wish. Harnad, Hanson, & Lubin's paper on categorical perception (CP) is a serious attempt to address the symbol grounding problem that has plagued AI. CP is simultaneous perception of enhanced within-category similarities and between-category differences, and is proposed as the mechanism which underlies symbol formation and grounding.

Servan-Schreiber, Cleermans, & McClelland explore the recognition of context as a means for the interpretation of individual events. They discuss these issues with an example of learning a finite state grammar. Omlin & Giles present an interesting study of insertion and extraction of symbolic information in Recurrent Neural Networks (RNN). Grammar rules can be inserted into an RNN by fixing the initial weights corresponding to known transitions. A trained RNN behaves like a deterministic finite state automata (DFA). Symbolic rule representation of a DFA can be extracted from a trained RNN by a dynamic exploration of the output state of the recurrent neurons. The authors' exposition of both these processes is very readable. They demonstrate that rule checking is an example application of the RNNs in that inserted rules can be verified and can even be corrected by training. The authors confess that their methods can only extract DFAs with no more than 15 states as compared to 800 states with other non-neural methods and urge research emphasis.

Sharkey & Jackson, Harnad et al, Servan-Schreiber et al, and Omlin & Giles provide insights into the role of the hidden units and attempt to show how NNs can represent symbolic information effectively. This is also demonstrated in the paper by Shavlik, to be discussed later. Sun alludes to the equivalence of fuzzy evidential logic and weighted-sum connectionist models and attempts to solve the variable binding problem. Pinkas' fault tolerant logic proof procedure utilises its theoretical relationship with energy minimisation techniques. The algorithm would have been more comprehensible if the examples were discussed more thoroughly. We were left with a nagging feeling that, by embedding all the intermediate steps of the proof procedure into to net, don't we effectively encode the answer in the net ?

The article by Uhr concludes this section. Its placement is rather strange as it has somewhat tenuous links to both the themes of the book and the section. He argues for the use of digital and analog microcircuits as useful building blocks for NNs and presents the interesting concept of computational analogs. However, the discussion of the possibilities for the hardware implementation of NNs is very dated, and the wide availability of numerous NN 'chips' (for example, a special issue of Intelligent Software Strategies (1992) has a survey of neural hardware and vendor information) is not noted. A rather surprising mistake was found in the equation used to describe the perceptron model (p. 341); using the notation given by him, shouldn't h(w,g(f(v))) read g(f(h(w,v)))?

Vision:

Kosslyn & Jacobs discuss a biologically and psychophysically motivated architecture for representing complementary (there is a typo in the title of this chapter which reads 'complimentary') relations in encoding shape and space. They note that the left hemisphere of the brain specialises in encoding categorical spatial relations (for example, the up/down relations of objects in a scene) and "category" representations of shape (for example, rectangular versus circular shape of an object); whereas, the right hemisphere encodes metric coordinate spatial relations (the actual distance between objects) and "exemplar" representations of shape (the identity of objects in a category). They attempt to understand the significance of this split in a computational paradigm. The authors do a commendable job in posing the right questions and analysing them from the results of network simulations. However, their discussion of the connectivity between the receptive fields and the input layer in the neural net is not clear. Also they do not address the themes of the book directly.

Carpenter & Grossberg, with Fuzzy ARTMAP, cover the three AI camps, namely, symbolic AI, connectionism, and biological (computational neuroscience). They identify the primary goal of biological neural network research as the 'design of systems that can self-organise intelligent symbolic processing capabilities'. They present a coherent, comprehensive body of work, more firmly based in theory, than the tinkering, piecemeal approach evident in many of the other papers. They discuss the issues in pattern recognition and prediction and refer to their impressive solutions using fuzzy ARTMAP to various benchmark problems.

Language:

All the papers in this section attempt to address the symbol grounding problem in one way or another. Dyer relates language understanding to perception action sequences, in a very complex system which he states is a first step towards addressing the language grounding problem. Much of the process in this system is hand -coded, but the underlying ideas are firmly based in cognitive models. He concludes with a big list of limitations of his model.

Miikkulainen suggests that the difference between AI and NN may be paralleled by that between reason (sequential, symbolic, rule-based) and intuition (parallel, subsymbolic). He has developed a DIstributed

Script processing and Episodic memoRy Network (DISCERN) for story understanding. This modular system consists of sub-networks for learning word meanings, script processing, and episodic memory organisation and are based on two principles, self-organisation and gradient-descent in error. Based on his experience with DISCERN, he suggests that future research needs to focus on developing methods for control of high-level behaviour and for learning more of the necessary structure in the sub-modules automatically rather than being fixed or specified in advance.

Learning:

Learning is probably the area where the clearest difference lies between GOFAI and NN but the problems identified in learning within both paradigms are not extensively explored in this section. Goldfarb & Nigam take the stand that both the SAI and NN-camps ignore the fundamental importance of inductive generalisation. Newell seems not convinced about this nor are we. Their point that vector-space based learning models such as NNs are inadequate inductive learning models, is well taken. However it is not clear as to what makes the weight space formulation in their example any different from a vector space? Their example does not demonstrate how inductive learning approach will avoid problems they identify for NNs. Shavlik discusses Knowledge-Based Artificial Neural Network (KBANN) system for embedding and extracting rules from an NN. She identifies protection of incorporated prior knowledge from being corrupted by the subsequent learning process as a major problem with these methods.

In sharp contrast to the methods and systems discussed so far in the book, Booker, Riolo, & Holland introduce adaptive production system framework of classifier systems as a way of integrating symbolic and numeric representations. Their exposition assumes familiarity with evolutionary computing and Genetic Algorithms (GA) which are not the main themes of the book. Honavar concludes this section by gathering all the ideas presented in the book and discussing the steps that can be taken to integrate different strategies and representations in learning systems. He argues that the space of possibilities resulting from the systematic integration of the powerful techniques for manipulating symbolic representations, robustness of statistical and connectionist approaches, and exploration capability of large search spaces by evolutionary methods, is boundless and filled with interesting and unexplored alternatives.

In conclusion, the book presents the reader with a large body of work, and includes many stimulating ideas and issues. We suggest that, keeping in view the large variety of topics covered, the book would benifit from short editorial summary for each section, setting the chapters within the context of the book and identifying the contributions they make toward the aims/themes of the book. The broad scope of the book makes it quite demanding to read and it is therefore unlikely to be a ‘must’ for everyone. However, it would be a very useful reference for researchers in the fields of AI and NN. Selections from the book may also provide good background reading for seminar discussions on postgraduate courses.

Raju S. Bapi
Sue L. McCabe
Neurodynamics Research Group
Rajkumar Roy
Engineering Design Centre
School of Computing
University of Plymouth
Plymouth PL4 8AA
United Kingdom
email: rajubapi@soc.plym.ac.uk

References

Aparicio IV, M. & Levine, D. S. (1994). Why are neural networks relevant to higher cognitive function? In D. S. Levine, and M. Aparicio IV (Eds.), *Neural networks for knowledge representation and inference* (pp. 1-26). Hillsdale, NJ: LEA Publishers.

Chandrasekaran, B. (1993). Interview with Allen Newell. *IEEE Expert*, June, 5-12.

Damasio, A. R. (1991). Concluding comments. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), Frontal lobe function and dysfunction (pp. 401-407). New York: Oxford University Press.

Fodor, J. A. (1976). *The language of thought*, Harvester Press, Sussex. (Harvard University Press paperback).
Fodor, J. A. & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, **28**, 3-71.
Harnad, S. (1990). The symbol grounding problem. *Physica D*, **42**, 335-346.
Neural networks: Hot air or hot technology? Part II. *Intelligent Software Strategies*, **VII: 6**, May, 1992.
Levine, D. S. & Leven, S. J. (1992). *Motivation, emotion, and goal direction in neural networks*, Hillsdale, NJ: LEA Publishers.
Newell, A. (1980). Physical symbol systems. *Cognitive Science*, **4**, 135-183.
Searle, J. (1980). Minds, brains and programs. *Behavioral Brain Sciences*, **3**, 417-457.

Grant support from the Engineering and Physical Sciences Research Council (EPSRC, United Kingdom) and Rolls Royce, plc/University of Plymouth to RSB and RR, respectively, is gratefully acknowledged.