The Properties of Higher Cognitive Processes and How They Can be Modelled in Neural Nets

William H. Wilson
Computer Science & Engineering
University of New South Wales
Sydney NSW 2052 Australia
billw@cse.unsw.edu.au

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

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

Abstract

It is proposed that the distinction between basic and higher cognitive processes can be captured by the difference between associative and relational processes. Properties of relational processing include reification of the link between entities, so higher-order relations have lower-order relations as arguments, whereas an associative link per se cannot be a component of another association. Therefore relational processes can be hierarchical and recursive, whereas associative structures are flat. Relations, unlike associations, also have the properties of omni-directional access and systematicity. Relational processes support reasoning and content-independent transfer, and have many of the properties of symbolic models. Typical feedforward neural nets do not implement these properties in a natural way, but they can be implemented with tensor product nets. The requirements for neural nets to model higher cognitive processes are considered.

The debate about the nature of higher cognition has included claims that it is symbolic, compositional and systematic, and that it cannot be modelled by associative architectures (Fodor & Pylyshyn, 1988). On the other hand it has often been shown that human reasoning does not readily conform to logic, and is better modelled by more content-specific processes such as mental models, that function as analogues (Johnson-Laird & Byrne, 1991; Niklasson & van Gelder, 1994). There has been a parallel debate about whether learning is conscious or unconscious, and whether what is learned is rule-based or instance-based (Shanks & St. John, 1994). These dichotomies also overlap to some extent with the implicit-explicit distinction (Clark & Karmiloff-Smith, 1993). As Hadley (1994) points out, we still lack a clear definition of the nature of higher cognition and, consequently, we are unable to clearly distinguish it from more basic processes. One consequence of this situation is that the criteria that neural net models of higher cognition need to fulfil have not been defined. In this paper we propose that higher cognitive processes entail representing and processing explicit relations, whereas basic processes can be identified with associations. This in turn constrains the type of neural net models that are required.

Higher cognitive processes depend on a set of structure-sensitive rules for operating on representations. In language syntactic rules define relations between words (e.g. parts of speech or case roles), but reasoning also depends on rules that relate entities, partly independent of content; e.g., the idea that fruit includes apples and non-apples is an instance of complementation or (in the Psychological literature) inclusion, but we can also understand inclusion as nonempty sets *a* and *a'* being included in *b*, independent of specific instances.

A structure is a set on which one or more relations is defined. Because relations are the essence of structure, the structural properties of higher cognitive processes can be captured by systems that process relations. However, despite its importance, and intensive study in disciplines such as computer science (Codd, 1990) the theory of relational knowledge has received little attention in Psychology (Smith, 1989).

A relation that relates *n* entities, or *n-ary relation* is a subset of the cartesian product of *n* sets: ie. $R(a_1,a_2,\ldots,a_n)$ is a subset of $S_1\times S_2\times\ldots\times S_n$. A relation must be identified by the relation symbol, R, and the entities by argument symbols, $a_1,a_2,\ldots,a_n$. For example the relation "larger" identifies a specific subset of a cartesian product, that subset in which the first entity is always larger than the second; i.e. $a_1 > a_2$.

The properties of cognitive processes based on relations, or relational schemas, are considered in detail by Phillips, Halford & Wilson (1995) and Halford, Wilson & Phillips (submitted). We propose that relations are links between entities that are identified by a symbol - the *symbolisation* property. The need for labelled links is recognised in models of higher cognitive structures such as propositional networks, in which links between nodes carry labels such as "agent", "object", "location". By contrast, though the point is rarely made explicitly, there seems to be agreement that associative links are all of the same kind, varying only in strength, and are unlabelled. In part because of symbolisation, *higher-order relations* can be defined, that have lower-order relations as arguments.

Association does not share this property. Associations can be chained, so that the output of one association is the input to another: $E_1 \rightarrow E_2 \rightarrow E_3 \ldots\ldots \rightarrow E_n$, and may converge, so that E_1 and E_2 elicit E_3, or diverge, so that E_1 elicits E_2 and E_3. However associations are not identified by a symbol as relations are, and the associative link *per se* cannot be an argument to another association. Therefore the recursive, hierarchical structures that can be formed using higher-order relations do not appear to be possible with associations.

Fodor and Pylyshyn (1988) argue that cognition depends on symbols that retain their identity when composed into complex symbols (compositionality), and have the property of systematicity, meaning that certain relations imply other relations. Relations have the *compositionality* property, because the symbols for the relation, and representations of arguments, retain their identity when bound into a structure. For example, in the relational instance larger(whale,dolphin), the components "larger", "whale" and "dolphin" retain their identity when bound into the relation. The *systematicity property* can be captured by *higher-order relations*, which have lower-order relations as arguments. Thus the fact that >(a,b) implies <(b,a), can be written as the higher-order relation IMPLIES(>(a,b),<(b,a)).

The *omni-directional access* property of relations gives flexibility to higher cognitive processes. It means that, given all but one of the components of a relational instance, we can access (i.e. retrieve) the remaining component. For example, given the relational instance mother-of(woman,child), and given mother-of(woman,?) we can access "child", whereas given mother-of(?,child) we can access "woman", and given ?(woman,child) we can access "mother-of".

Propositions, which are the core of some models of higher cognitive processes, can be treated as relational instances (Halford et al., submitted, section 2.2.2). For example loves(Joe,Jenny) is a proposition in that it has a truth value, but it is also a relational instance.

We will develop our argument by reference to a task that has longstanding importance in the psychological literature, conditional discrimination, an example of which is shown in Figure 1.

Conditional discrimination makes the correct response dependent on background (e.g. triangle is positive on a black background, but square is positive on a white background). It cannot be learned by association between stimulus elements and responses, because each response is equally associated with both stimulus elements, and with both backgrounds (e.g. R+ is associated with both triangle and square, and also with black and white). Therefore associative interference is high, and effectively blocks learning based on elemental associations.

Conditional discrimination can be learned by fusing or chunking each stimulus into a unique *configuration*, such as triangle/black, as shown in Figure 1. The task can be learned this way because each configuration is distinct, and associative interference is reduced. The problem however is that, for each configuration to become unique, the elements must lose their identity. Thus if "triangle" and "black" become fused into a unique configuration *black/triangle*, which is distinct from (say) *black/square* or *white/triangle*, then the components of black/triangle lose their identity. An element within a configuration is, of necessity, no longer recognisable as the element it was. The problem then is that the structure of the task cannot be represented. The observable effect would be that the person would not be able to transfer to a problem isomorph.

Conditional discrimination task

background	cue		response
black	triangle	→	R+
black	square	→	R-
white	triangle	→	R-
white	square	→	R+

Configural learning

configuration		response
configuration triangle/black	→	R+
configuration square/black	→	R-
configuration triangle/white	→	R-
configuration square/white	→	R+

Isomorphic conditional discrimination task

background	cue		response
green	circle	→	R+
blue	cross	→	?
green	cross	→	?
blue	circle	→	?

Figure 1. Conditional discrimination tasks.

In general configural associations cannot be transferred to isomorphs, because configurations do not represent structure. Figure 1 shows an isomorphic conditional discrimination task. Notice that, once the first item is known, responses to the remaining 3 items can be predicted, and this would be true for any order of presentation. This can be done only if the relations in the task are represented, and it is not possible if configurations are learned. Configural learning is implicit, in that it enables the tasks to be performed, but without understanding of the principles it entails. Learning the relations in the task corresponds to explicit knowledge, because it confers understanding of the structure.

Neural nets

Variations in the capabilities of neural nets in some ways parallel the contrast between associative and higher cognitive processes. Elemental association can be implemented by 2-layered nets, but conditional discrimination has the same structure as exclusive OR (XOR) and requires 3-layered nets (Minsky & Papert, 1969; Schmajuk & DiCarlo, 1992). However Phillips and Halford (in press) found that although a three-layered net learned specific instances of a conditional discrimination, there was no evidence of transfer to problems with the same structure, but different input stimuli. A three-layered net was unable to transfer to an isomorphic problem. Furthermore Phillips (1994) has shown that three-layered nets cannot exhibit strong systematicity.

It appears therefore that two- and three-layered nets (multi-layer perceptrons, MLPs) can model basic processes, but are not structure sensitive. Simple recurrent nets (Elman, 1991), and nets with multiple hidden layers, are more complex but they do not represent relations explicitly, and their representations are arguably more like configurations. Essentially, the problem with MLPs is that learning is tied to specific inputs and outputs. Although MLPs demonstrate generalisation, it relies on input/output pattern similarity. So, tasks involving completely novel stimuli force relearning.

Transfer across isomorphs imposes particular constraints on feedforward networks. A single hidden layer will not suffice since the weights from units representing novel input stimuli will not have been trained in previous tasks. Hence, no sensible mapping from input to hidden unit representations can be expected (see Phillips, 1994, for parallel arguments on strong systematicity). Yet, the use of two hidden layers, to separate inter-task knowledge (hidden to hidden weights) from intra-task knowledge (input to hidden and hidden to output weights) does not solve the problem. Either, there are too many weights and so, no generalisation, or too few weights and no solution to individual problems (Phillips & Halford, in press).

Transfer between isomorphs requires that task structure be represented. This can be effectively represented by a more structured (specifically connected) network that computes the outer product of symbol and argument vectors (Halford et al., 1994). A collection of relational instances can be superimposed on the same representation, by adding up the outer products. Thus representations of loves(John,Mary) can be represented as $v_{loves} \otimes v_{John} \otimes v_{Mary}$ and loves(Tom,Wendy) can be represented as $v_{loves} \otimes v_{Tom} \otimes v_{Wendy}$. These representations can be superimposed by summing the outer products, yielding $v_{loves} \otimes v_{John} \otimes v_{Mary} + v_{loves} \otimes v_{Tom} \otimes v_{Wendy}$.

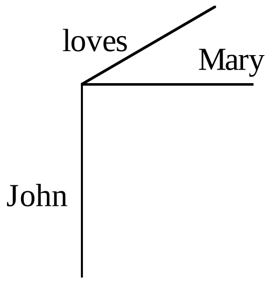

Figure 2. Tensor product representation of loves(John,Mary).

The resulting sum of outer products is referred to as a *tensor*. Thus the relational instance $r(a_1,a_2,\ldots,a_n)$ would be represented in a tensor product space $V_R \otimes V_1 \otimes V_2 \otimes \ldots \otimes V_n$, where V_R represents alternative relation symbols including *r*, and v_i (i>0) represents concepts appropriate to the *i*th argument position.

The tensor product network architecture of which this is an example is due to Smolensky (1990). A tensor product network can be used to represent the structure of tasks such as conditional discrimination (Phillips & Halford, in press; Phillips et al., 1995). For example, the structure of the first conditional discrimination task in Figure 1 can be stored in a rank 3 tensor as:

T_3 =

black ⊗ triangle ⊗ yes +

black ⊗ square ⊗ no +

white ⊗ triangle ⊗ no +

white ⊗ square ⊗ yes,

Transfer to an isomorph can be performed by analogy, that is by mapping the structure of the first conditional discrimination task into the isomorphic task. This can be done by the Structured Tensor Analogical Reasoning (STAR) model (Halford, Wilson, & Phillips, 1996). The resulting mapping can be stored in a rank 2 tensor for interpreting between task elements as in:

T_2 = green ⊗ black + blue ⊗ white + circle ⊗ triangle + cross ⊗ square

Transfer to the isomorph would require having seen, for example, the instance green,triangle → yes. Then, the stimulus blue,cross is predicted by interpreting the inputs as blue • T_2 = white, and cross • T_2 = square; and computing the response based on knowledge from the previous task: black × square • T_3= no. Here "•" signifies the product of a vector by a matrix.

The essential difference between the tensor and MLP approach is that task knowledge is explicitly represented in the former, but not the latter. That is to say, that knowledge is accessible by other processes within the network, without having to reproduce the same context within which that knowledge was acquired. In other words, knowledge of the previous task was elicited without having to be in the context of solving the previous task.

This approach to neural net modelling of relations is based on symbol-argument-argument bindings. Roles are not represented explicitly, as the role-filler models of Hummel and Holyoak (in press), Plate (in press), Shastri & Ajjanagadde (1993) or Smolensky (1990), but are determined positionally, a type of coding also used in language. One difficulty with the role-filler approach can be illustrated by considering how two relational instances might be represented. For example loves(John,Mary) can be represented as:

loves + lover.John + loved.Mary

Here the "." symbol signifies the role-filler binding and the "+" serves to concatenate the bindings to the relation-symbol. Suppose we now represent loves(Tom,Wendy) as:

loves + lover.Tom + loved.Wendy

When we put the representations of both relational instances together we have:

loves + lover.John + loved.Mary + loves + lover.Tom + loved.Wendy

This represents the fact that John and Tom are lovers and that Mary and Wendy are loved, but it does not distinguish between John loving Mary and John loving Wendy, and is similarly ambiguous with respect to Tom. This problem is discussed in more detail by Halford et al. (submitted).

Our approach, based on symbol-argument-argument bindings is able to implement all the properties of relational knowledge (Halford et al., submitted) and therefore can represent structure. Tensor products seem naturally adapted to representing relations in this way because their structure is analogous to the cartesian product space in which relations are defined. We can think of an N-ary relation as a set of points in N-dimensional space, and each dimension corresponds to an axis of the tensor. The proposed model takes advantage of this natural correspondence.

Robustness properties

Tensor product representations have the property of graceful degradation (Wilson & Halford, 1994). More recent simulations in our laboratory have extended this finding. For example, a rank 5 tensor of side 16 (i.e. n=16, k=5) with up to 93.75 percent of the binding units deleted, reliably distinguished stored facts (relational instances) from non-facts. Such a tensor has the same number of active binding units as an intact rank 4 tensor with side 16. Thus it appears to be possible to simulate a rank k+1 tensor with the number of binding units available to an (intact) rank k tensor, over part of the range of k, but with processing becoming progressively poorer at successively higher ranks.

Cognition, structure and neural nets

A major challenge of contemporary cognitive science is to account for the way higher cognitive processes represent and process structured knowledge. The theory of relational knowledge provides a way of defining the properties of cognition that are essential to this task. Relational representations are reified in the sense that they are objects within the domain of other cognitive processes. They are accessible to other cognitive processes, and constitute knowledge to the system as well as knowledge in the system. This makes possible hierarchical, recursive structures. The omni-directional access property gives relational processes the flexibility which is characteristic of higher cognition. Relational representations can also be transferred across isomorphs, which gives content independence, and permits unknown relational instances to be inferred. Associative links, by contrast, cannot be components of other associations, so associative structures are flat and non-recursive. Omnidirectional access is not inherent in associations, so they lack some of the flexibility of higher cognition.

These properties constrain the type of neural net that is required to model higher cognition. The components, symbol and arguments, should remain accessible and identifiable, at least in principle, when bound into a representation of a relation. There must be provision for higher-order relations, systematicity, and omnidirectional access. Feedforward nets do not typically implement these properties in a straightforward or natural way. One approach to their implementation has been illustrated using tensor products of vectors.

References

Clark, A., & Karmiloff-Smith, A. (1993). The cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind and Language*, **8**(4), 487-519.

Codd, E. F. (1990). *The Relational Model for Database Management: Version 2*: Addison-Wesley.

Elman, J. L. (1991). Distributed Representations, Simple Recurrent Networks, and Grammatical Structure. *Machine Learning*, **7**, 195 - 225.

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.

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 connnectionist and neural computation theory, Vol. 2: Analogical connections* (pp. 363-415). Norwood, NJ: Ablex.

Halford, G. S., Wilson, W. H., & Phillips, S. (1996). Human analogical reasoning capacity: Toward a neural net model. *International Journal of Psychology,* **31**(3/4), 443.

Halford, G. S., Wilson, W. H., & Phillips, S. (to appear). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology, *Behavioural and Brain Sciences.*

Hummel, J. E., & Holyoak, K. J. (in press). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Minsky, M. & Papert, S. (1969). *Perceptrons: an introduction to computational geometry*. Cambridge, MA: MIT Press.

Niklasson, L., & van Gelder, T. (1994,). Can connectionist models exhibit non-classical structure sensitivity? *Sixteenth Annual Conference of the Cognitive Science Society, Atlanta, Georgia,* 664-669.

Phillips, S. (1994). Strong systematicity within connectionism: The tensor-recurrent network. *Sixteenth Annual Conference of the Cognitive Science Society,* 723-727.

Phillips, S., & Halford, G. S. (in press). Systematicity: Psychological evidence with connectionist implications. *Nineteenth annual conference of the Cognitive Science Society*.

Phillips, S., & Halford, G. S. (submitted). Modeling schema induction within relational and connectionist frameworks: A first approximation.

Phillips, S., Halford, G. S., & Wilson, W. H. (1995, July). The processing of associations versus the processing of relations and symbols: A systematic comparison. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society, Pittsburgh, PA,* 688-691.

Plate, T. A. (in press). Estimating analogical similarity by vector dot-products of holographic reduced representations. *Cognitive Science*.

Schmajuk, N. A., & DiCarlo, J. J. (1992). Stimulus configuration, classical conditioning, and hippocampal function. *Psychological Review,* **99**, 268-305.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences,* **17**, 367-447.

Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences,* **16**(3), 417-494.

Smith, L. B. (1989). From global similarities to kinds of similarities: The construction of dimensions in development. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 146-178). Cambridge, MA: Cambridge University Press.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence,* **46**(1-2), 159-216.

Wilson, W. H., & Halford, G. S. (1994,). Robustness of tensor product networks using distributed representations. *Proceedings of the Fifth Australian Conference on Neural Networks, Brisbane,* 258-261.

Acknowledgments: This work was supported by grants from the Australian Research Council and the Science and Technology Agency (Japan).