Computational Templates, Neural Network Dynamics, and Symbolic Logic

Otto Lappi

Abstract— **This paper looks at the relationship between subsymbolic neural networks and symbolic logical systems from a philosophy of science perspective. More specifically, the point of view is that of Paul Humphreys' philosophical account of the organization of scientific knowledge [Humphreys, Paul. Extending Ourselves – Computational Science, Empiricism and the Scientific Method. Oxford: Oxford University Press, 2004.] Humphreys considers the units of analysis constituting scientific knowledge (in computational science) to be computational models, and computational templates. Computational templates are abstract syntactic schemata underlying domain specific models. Humphreys' own examples are mainly from computational research in physics, biology and statistics. Neural networks research in cognitive science can also be fruitfully considered from this computational models/templates perspective, illustrated here by two examples from recent neural networks research.**

I. Introduction

HOW is scientific knowledge organized? Paul Humphreys [1,2] has presented a framework for looking at the organization of scientific knowledge where computational models and templates - rather than theories, concepts, laws, or research programs or paradigms – constitute the units of analysis of scientific knowledge. Adopting this perspective on computational reseach, Humphreys argues, has ramifications in many issues in the philosophy of science, including scientific discovery, explanation, reductionism and the unity of science.

Humphreys' [1,2] main thesis is that computational science is best seen as organized around computational templates, which can be considered as abstractions from computational models. They are abstract computational schemata, the common syntactic core of a diversity of models, used to compute very different things in models of different phenomena. As purely syntactic abstractions, they can be considered in separation from any particular interpretation. Templates are not models of phenomena, but are instead necessary but not sufficient constituents of computational models in any given domain.

In developing his framework Humphreys discusses examples mainly from computational methods in physics and biology. Computational methods are of course also widely used in cognitive science and neuroscience. Artificial neural networks research, in particular, would seem like a good example of modeling based on common computational templates.

In this paper I present as case studies two examples of recent neural networks research. I will first illustrate how they fit into the overall view of computational science proposed by Humphreys [1,2], and then close with a brief discussion of implications for neurocognitive explanation, psychoneural reduction and the unity of science.

Otto Lappi is with the Cognitive Science Unit, Department of Psychology, PO BOX 9, University of Helsinki, FINLAND (phone: 358-9-19129674; fax: 358-9-19129542; e-mail: otto.lappi@ helsinki.fi).

II. Computational Models And Computational Templates

What are the appropriate units of philosophical analysis of scientific knowledge? In particular, what would be the best philosophical account of the use of computational techniques in the neurocomputational sciences? What is the use of computer simulation in scientific reasoning and argumentation and how do computational models work in scientific explanation of cognition? Is there some special contribution computational methods have made to the growth of empirical scientific knowledge of the mind/brain?

Humphreys [1,2] has put forward a view whereby the organization of scientific knowledge is understood in terms of computational models that are based on computational templates. (This is meant to contrast with some of the more traditional units of analysis in the philosophy of science, such as research programs or paradigms, concepts, theories, and laws).

Templates, as understood here, are formal structures considered in isolation from their use in modeling particular phenomena. They are schemata that must be supplemented with a domain specific interpretation in order to be applied, or to be considered representational models of a specific class of phenomena. The *very same* template may therefore surface in domains that do not overlap in subject matter, i.e. the same syntax may turn out to be useful in computational models that are otherwise quite unrelated.

A caveat is in order. The *logical* analysis of a model into a template and an interpretation is not to be read too literally, as saying that in the context of discovery models would be *historically* constructed first as a template with an interpretation then bolted onto it, almost as an afterthought. According to Humphreys, as far as the historical and psychological aspects of model construction are concerned:

"Although one can view the computational templates as consisting of a string of formal syntax together with a separate interpretation, this is a misinterpretation of the construction process. The computational language is interpreted at the outset and any abstraction process that leads to a purely syntactic computational template occurs at an intermediate point in the construction." [1].

One may think of the process of model construction in the following way: A model is first constructed with an intended interpretation in mind. This interpretation provides initial plausibility and justification for the model (even before it is tested against data). It also gives the scientists a clue as to which parts of the model are the first to be revised or refined, and which parts on the other hand are "not negotiable". A template may then be abstracted (a process of separation of the formal syntax from the interpretation), and subsequently discovered/shown to apply in some new domain, as well.

In different fields the same template will be interpreted differently, and used differently to model different phenomena. This means that if the predictions of the model fail to fit the data the corrections that the modelers in any particular field will be disposed to make will generally depend on domain specific knowledge, such as knowledge of the idealizing and abstracting assumptions that went into the adoption of the template in the first place. Justification and negotiability of different aspects of a model will depend on judgments of explanatory leverage and overall theoretical coherence.

In summary, templates traveling from one field to another create opportunities for interdisciplinary cross-pollination. They can be seen as abstract mathematical structure that may be reinterpreted for the purposes of modeling different domains. (Although Humphreys [2] stresses that in computational science they should be seen not just as abstract structure but as *syntactic vehicles* for reasoning, enabling scientists to simulate, predict and/or to explain phenomena).

III. Neural Networks

While Humphreys draws his examples mainly from physics and engineering, it seems natural to consider neural networks research in cognitive science from the point of view of computational templates as well. I will next present two cases from contemporary neural network research for which Humphreys approach seems particularly appropriate. Both cases reviewed here are research attempting to integrate the symbolic and connectionist approach to cognition. After reviewing the main points, I will briefly discuss at a more general level how the perspective of templates may turn out to be useful for understanding cognitive and neural theories of information processing, and the relationship between them.

A case study of the use of template construction in neural networks research is provided by the analysis of coherence problems in Thagard and Verbeurgt [3]. They propose that many phenomena in psychology, philosophy, law, and the social sciences can be represented as coherence problems, and present a formal computational characterization of coherence problems in terms of constraint satisfaction. They then review several algorithms (both symbolic and connectionist) for computing coherence. See also [4] and references in [3,4], for related research.

A coherence problem is defined as the task of dividing a set of elements – beliefs, concepts, representations generally - into two mutually exclusive and jointly exhaustive subsets (a set of accepted elements, and a set of rejected elements). This division will be based on the elements' coherence relations, so that pairs of elements that cohere will tend to be accepted together or rejected together, and pairs of elements which do not cohere with each other will tend to be accepted if and only if the other one is rejected. Coherence and incoherence are here treated as soft constraints, positive and negative, on a partitioning, and the computational problem is to maximize satisfaction of such constraints [3,4].

Let E be a finite set of elements $\{e_i\}$, and let C be the set of constraints between these elements, which are represented as pairs of elements of E: $\{(e_i,e_j)\}$. Each constraint is either positive or negative in sign, dividing C into C+, the positive constraints, and C-, the negative constraints. Each constraint also has a weight w_{ij}. The task is then to partition E into mutually exclusive and jointly exhaustive subsets A (for "accepted") and R (for "rejected"), such that the following coherence conditions are respected:

1. if (e_i,e_j) is in C+, then e_i is in A if and only if e_j is in A.
2. if (e_i,e_j) is in C-, then e_i is in R if and only if e_j is in R.

It may not be possible to simultaneously meet these conditions for *all* elements in E. In that case, the goal is to find the partitioning that maximizes satisfaction of these constraints. Letting W be the weight of a partition, defined as the sum of all constraints satisfied by the partitioning, the problem is then to maximize W.

Thagard et al. [3,4] connect the computational analysis above to connectionist modeling by showing how neural networks can be constructed so as to compute coherence. Indeed, they acknowledge neural network research as the *historical* origin of their account of coherence as constraint satisfaction: "Connectionist algorithms can be thought of as maximizing the goodness-of-fit or harmony of the network (…) The characterization of coherence given (…) is an abstraction from the notion of goodness-of-fit." [3] Note here the close correspondence to what Humphreys says about the process of template construction, above.

What is the benefit to be gained from such template construction? According to Thagard and Verbeurgt, "The psychological contribution of this paper is that it provides an abstract formal characterization that unifies numerous psychological theories. We provide a new mathematical framework that encompasses constraint satisfaction theories of hypothesis evaluation, analogical mapping, discourse comprehension, impression formation, and so on. Previously, these theories shared an informal characterization of cognition as parallel constraint satisfaction, along with use of connectionist algorithms to perform constraint satisfaction. Our new precise account of coherence makes clear what these theories have in common besides connectionist implementations.". And, "The value of the abstraction is that it provides a general account of coherence independent of neural network implementations." [3]

Abstracting from the specifics of neural networks makes it possible to investigate different algorithmic approaches to computing coherence, including symbolic algorithms more closely related to the abstract definition of coherence itself.

Thagard et al. also point out [3] that non-connectionist algorithms (for the same computational problem) can be used to benchmark the performance and resource requirements of the connectionist systems. What is more, they note that establishing the correspondence between the symbolic algorithms and connectionist networks provides insight into the behavior of the network - as the behavior of the symbolic algorithm is often more readily interpretable in terms of the computational problem than are the dynamics of the corresponding network. (I here interpret correspondence to mean that both the network and the symbolic algorithm are implementations of the same computational template).

To take another example, Leitgeb [5] shows how to represent artificial neural networks as interpreted dynamical systems, and how to define precisely the correspondence between the dynamics of a network and a system of logic; the idea is that the dynamics governing the state transitions parallels the inferential relations between interpretations of network states.

The logical system of allowed inferences is specified with a small set of qualitative laws, the neural network by structure and dynamics of a network of formal neurons. The parallelism between the dynamics of the network and the consequence relation of the system of logic is such that the dynamics of state transitions conform to patterns of (nonmonotonic) reasoning, allowing the network to be considered an interpreted dynamical system that represents a system of logic (defining logical relationships between interpretations of systems states).

Without going into too much detail, the relevant properties of the systems can be outlined as follows (for details see [5,6,7]):

First of all, states of neural networks are assigned interpretations. These interpretations, the information contents of network states, are represented as propositional formulae in a symbolic logical language (not all states need have an interpretation). At the symbolic level, a logical system of qualitative laws of defeasible inference is formulated, whereby less informative (with respect to an information ordering of the propositions) conclusions may be defeasibly inferred from more informative premises. For example from the premise that Joe is a rat one may infer the less informative conclusion that he is a mammal, or from the premise that Tweety is a bird one might infer the less informative conclusion that Tweety flies. (This inference is defeasible in that the belief in the conclusion may need to be revised if one finds out that Tweety is an ostrich, or a dodo). The dynamics of the state-transitions of the network is designed to respect this information ordering, thus embodying the system of inferences.

This establishes that the dynamical system, specified at the level of connectionist architecture represents the logical system of inferences, specified at the symbolic level: A (symbolic) qualitative law $\varphi \Rightarrow \psi$, where φ and ψ are sentences of the propositional language in terms of which the interpretation is given and $\Rightarrow$ a defeasible conditional, is said to hold of the interpreted (subsymbolic) dynamical system when, for every state representing exactly the content φ, the dynamics of the system is such that it will take the system to a stable state whose interpretation will contain ψ. This interpretation (stable state) can thus be taken to be the outcome of the inference (computation). Conversely, the information content ψ is said to be contained in φ just when, with respect to the information ordering, the state whose interpretation is φ larger-than-equals the state whose interpretation is ψ.

How should this theory be analysed from Humphreys' point of view, then? What, in particular, is the template here and what use is the template put to? The shared template is neither the network architecture nor the rules of logic as such, but the more abstract information-ordering structure governing both. The use it is put to is unification of symbolic and subsymbolic approaches to cognition.

Here the whole point of constructing the template is to establish, in a logically precise manner, the correspondence between two classes of systems that initially seem rather different - *not* to produce a vehicle for prediction. Indeed, the *simulation* value of such highly idealized models in real world contexts is probably quite meagre. (Humphreys stresses the predictive use of templates in simulation. But if the model is suitably abstract, or the system is tremendously complex, and most parameters and variables affecting behavior are unknown, as is the case with the human brain, templates can have unificatory (and explanatory) value even

without predictive value). See [8] for a discussion on the asymmetry of explanation and prediction.

Note also that in this case establishing the existence of the template is a result of purely theoretical/foundational inquiry, not part of the process of constructing a model for a particular phenomenon. It seems, then, that besides modeling Humphreyan computational templates and template construction is useful approach for the purpose of explanation and interdisciplinary unification, too.

IV. Computational Templates Approach as an Aid to Understanding Multilevel Explanation in the Neurocognitive Sciences

Templates provide explanatory unification in the sense that the same computational structure can be used to model and explain phenomena in a diversity of disciplines. Computational modeling in the neurocognitive sciences is concerned with phenomena spanning what are commonly seen as different levels of explanation. In very broad terms, the traditional picture of the unity of science recognizes two main approaches to *interlevel* integration.

First, there is methodological unification (based on the idea of unity of the scientific method). The use of the same computational template across disciplines perhaps has an air of methodological unification about it – it is after all a case of using the *same means* to model different phenomena. But surely the *methods* of computational modeling are the same regardless of template. Therefore, the additional integration achieved by two fields sharing a template must be something over and above methodological unification. (Also, surely it is not the case that adhering to both the templates perspective as well as the doctrine of the unity of science would commit one to assume that the same templates should be used in all, or even very many, sciences).

Second, there is explanatory unification, usually thought of as based on intertheoretic reduction of "higher level" laws, theories or concepts – in other words, the domain specific ontology of one field - to those of another field, which is seen as ontologically more fundamental. In terms of psychoneural integration, *psychology* is typically considered here the field to be reduced, and neuroscience the reducing, more fundamental, field. However, use of the same template in modeling these different domains does not mean that one needs to think of one domain thereby being reduced to the other. From establishing the fact of a symbolic and subsymbolic system being describable in terms of a shared a template, no implications follow as to the explanation of phenomena of, say, psychology in terms neurophysiology. No reductionist ontology need be assumed, whereby one discipline (physics, neuroscience) would have an *a priori* preferred status vis-à-vis some other (psychology), whose vocabulary would then be considered as merely convenient "dramatic idiom".

In fact, the question does not arise at all, since templates are by design ontologically neutral; they do not belong to any particular "level of explanation" as such. Therefore, interlevel integration in terms of templates can be considered likewise ontologically neutral. The neural network level need not be seen as "more basic" in terms of template construction (or explanation).

V. Conclusions

Humphreys introduces the idea of computational templates in conjunction with physical laws and core computational schemata of biological models. The prominence of computational methods and computer simulations in both cognitive and neural modeling make the computational model/template based approach particularly natural as an account of the organization of knowledge and interdisciplinary relations in the cognitive sciences. Here we have looked at two cases in theoretical cognitive science from the point of view suggested by Humphreys.

For the philosopher of science, Humphreys' account is useful in spelling out what the two studies reviewed here have in common. Both represent neural networks research that attempts to integrate the symbolic approach - cognition as rule governed logical systems - with the connectionist approach to cognition - subsymbolic elements interconnected into dynamical systems. The way that they go about doing this fits well with Humphreys' view of interdisciplinary integration in computational science generally – that is, the construction of shared computational templates that can be applied across disciplinary boundaries.

Neural networks are usually seen as contributing a brain-based, neurologically inspired, and biologically plausible approach to cognitive modeling. One of the original motivations of neural network models as models of parallel distributed information processing in real brains was their neurological flavor [9]. Compared to most symbolic information processing models, they are more "neurally inspired". Correspondingly, the intertheoretic unification provided by connectionism is often seen as bringing the brain back into the picture – which in turn naturally leads to issues of reductionism, (lack of) autonomy of scientific psychology and eliminative materialism.

The template perspective gives a rather different view of these issues. The contribution of the emergence of the subsymbolic paradigm is not just the domain-specific ("neuro") content that can be brought in (to interpret the template in *neuro*computational terms), but also the unificatory job carried out by the process of template construction itself. The unificatory power of templates resides not in that the neural mechanisms would explain or reduce cognitive phenomena, nor in that the cognitive

interpretation would explain the neural mechanism. Instead, it resides in the correspondence of the two established by constructing a template common to both. Explanation and unification is neither a bottom-up nor a top-down affair, but a bidirectional one.

From a philosophy of mind point of view, the most interesting question is of course whether neural networks can shed light on the relation between mind and brain. Perhaps a clearer philosophical understanding of the relationship can be attained by looking at cognitive modeling from the point of view of shared computational templates in neural and cognitive models. Be that as it may, from a philosophy of science point of view it seems that besides providing a nice analysis of the use of computational methods in physics and engineering, Humphreys' theory of computational templates is a useful way to look at the issue of psychoneural integration in the neurocognitive sciences as well.

ACKNOWLEDGMENT

To Anna-Mari Rusanen for her generosity, and useful comments concerning both content and presentation.

REFERENCES

[1] P. Humphreys, "Computational Models", *Philosophy of Science*, vol. 69, pp. S1-S11, 2002.

[2] P. Humphreys, E*xtending Ourselves - Computational Science, Empiricism and the Scientific Method*. Oxford: Oxford University Press, 2004.

[3] P.Thagard & K.Verbeurgt, "Coherence as Constraint Satisfaction", *Cognitive Science*, vol. 22, pp. 1-24, 1998.

[4] P. Thagard, C.Eliasmith, P.Rusnock and C.Shelley, "Knowledge and Coherence". In R. Elio (Ed.), *Common Sense, Reasoning, and Rationality* (Vol. 11), 104-131. New York: Oxford, 2002

[5] H. Leitgeb, "Nonmonotonic Reasoning by Inhibition Nets", *Artificial Intelligence*, vol. 128, pp. 161-201, 2001.

[6] H. Leitgeb, "Nonmonotonic Reasoning by Inhibition Nets II", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 11, pp. 105-135, 2003.

[7] H. Leitgeb, "Interpreted Dynamical Systems and Qualitative Laws: from Neural Networks to Evolutionary Systems", *Synthese*, vol. 146, pp. 189-202, 2005.

[8] R. A. Wilson and F. Keil, "The Shadows and Shallows of Explanation", *Minds and Machines*, vol. 8, pp. 137-159, 1998.

[9] J. McClelland and D. Rumelhart, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1. Cambridge, MA: MIT Press, 1986.