Computational Templates, Neural Network Dynamics, and Symbolic Logic

Otto Lappi

***Abstract*—The relationship between subsymbolic neural networks and symbolic logical systems is discussed from the point of view of the account of computational science by Paul Humphreys [Humphreys, Paul (2004), Extending Ourselves. Oxford: Oxford University Press.] This philosophical account of the organization of scientific knowledge considers the units of analysis constituting scientific knowledge to be computational models, and computational templates. Computational templates are more abstract computational schemata underlying specific models. This view is contrasted with the received views where concepts, theories or laws constitute the units of analysis. Humphreys' examples are mainly from physics, biology and statistics, but neural networks can also be considered computational models/templates in Humphreys' sense. Implications regarding psychoneural reduction vs. emergence, and the symbolic-subsymbolic debate are also discussed.**

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. This change in perspective has ramifications in many issues in the philosophy of science, including scientific discovery, explanation, reductionism and the unity of science.

In developing his framework, Humphreys discusses examples mainly from computational methods in physics and biology. Computational methods are also widely used in cognitive and neurosciences as well, and artificial neural networks would seem like a good example of computational templates as well. Here we take a look at one particular case for which Humphreys' approach seems appropriate, namely, the relationship between subsymbolic information processing devices such as neural networks, and systems symbols and rules (of nonmonotonic symbolic logic).

Looking at the issues from the computational templates perspective not only presents a case study for philosophical analysis, but may also have relevance for discussions on foundational issues in the philosophy of cognitive science such as the connectionist/classicist, symbolic/subsymbolic debates on cognitive architecture, psychoneural reductionism, as well as the nature of computation.

Manuscript received January 31, 2007.

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 and the scientific method? In particular, what would be the best account of the use of computational techniques in scientific reasoning & explanation, discovery, empirical justification of theories, and the growth of scientific knowledge in the neurocomputational sciences?

Some of the more traditional units of analysis in the philosophy of science are research programs/paradigms, concepts, theories and laws.

For sure, neural networks research and connectionist modeling is often called a "new paradigm" for the cognitive sciences [3] - though what philosophical posits are implied is often less clear. But the philosophical accounts of paradigms and research programs [4,5] seem to take somewhat too broad a perspective for our present purposes. A finer grain of analysis is needed.

Concepts are often thought to demarcate a domain of research proprietary to a discipline (physics, neuroscience, psychology etc). However, for analyzing interdisciplinarity, for example the unificatory role of neural networks in neurocognitive science, we need to look at units of analysis that are shared by the disciplines, rather than that which differentiates one discipline from the another.

To the logician, concerned as he is with the maximally general and abstract accounts of scientific knowledge, this knowledge is represented as theories: sets of sentences closed under a consequence relation. But when our concern is to understand the workings of some piece of real science this is not a very helpful starting point, either; simply put, to a working scientist any arbitrary set (finite or infinite) of sentences and its consequences would not seem like a "theory". A theory is *about* a subject matter, and the assumptions that go into a theory must *cohere* in some way and economically describe some phenomena in nature.

Enter laws, as the "core" assumptions of a theory from which special cases may be derived as consequences. Scientific laws bring unity to a discipline. Fundamental first principles (for example Newton's and Maxwell's laws) unify what initially appear disparate phenomena under one coherent theoretical scheme. Here we are already close to the idea of templates. However, the unificatory effect of laws is still to bring phenomena into a single conceptual framework, the domain of a theory - rather than providing *inter*disciplinary unification of genuinely disparate branches of science each with their distinct subject matter.

What is more, in the neural and cognitive sciences there are as yet few, if any, *laws* in the sense physics has them. What cognitive scientists and neural modelers construct can be somewhat more modestly be labeled models of specific phenomena, rather than universal laws.

Humphreys [1,2] suggests of computational science is best seen as organized around computational templates. Computational templates can be considered as abstractions from computational models; they are abstract computational schemata, used to compute very different things in models of highly varied phenomena, and which can therefore be considered in separation from any particular interpretation *qua* models of phenomena of any particular domain. Laws and models intrinsically have an interpretation, they represent a domain and the elements coordinated by the formal structure may be considered as concepts denoting objects of study. The templates are formal structures considered in isolation, schemas that must be supplemented with domain specific interpretation in order to be applied.

Humphreys' thesis is that organization of scientific knowledge can be understood in terms of computational models that are based on templates, which may travel from one discipline to another.

A caveat: This 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 constructed as a template and with an interpretation 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 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 constructed with an intended interpretation in mind. This interpretation provides a 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 are "non-negotiable". A template may then be abstracted (separation of the formal syntax from interpretation) and discovered to apply in a new domain as well. This might happen when templates are explicitly used as a vehicle for analogical reasoning in model construction in the target discipline, or it might happen that the two disciplines develop largely in parallel, independently of one another, and only later is it realized that they have converged on models based on a shared template.

In different fields the same template will be differently interpreted, and used differently to model quite different phenomena. For example, if the predictions of the model fail to fit the data the corrections that the modelers will be disposed to make will generally depend on domain specific content-knowledge of the idealizing and abstracting assumptions that went into the adoption of the template. What remains the same is the abstract (computational) structure of the template.

"Traveling" templates thus create opportunities for inter-disciplinary cross-pollination. They can be seen either as abstract mathematical structure that may be reinterpreted in modeling a new domain, but Humphreys stresses that in computational science they should be seen as *syntactic* vehicles for reasoning enabling scientists to simulate, predict and/or to explain phenomena: "Syntax matters" [2].

III. Neural Networks

A. *Neural networks as templates*

Neural networks can be considered abstract computational templates in the above sense in a very natural way.

One of the original motivations of neural network models as models of parallel distributed information processing in real brains was their neurological plausibility. Compared with most symbolic information processing models the were "neurally inspired". Another strong point was their ability to perform tasks whose execution depends on taking into account multiple "soft constraints" (something that comes naturally to humans, but not, it was argued, to "Good Old Fashioned Artificial Intelligence" [3,6]).

But today neural networks have found applications in many different scientific domains beyond neuroscience and cognitive psychology, where the goal is to create models that can shed light on some some phenomena of interest, where the original neurophysiological and psychological construction assumptions are not an issue, and the network model is not interpreted necessarily as a model of the human mind/brain. In such cases the "neural network" functions more as an abstract template than a model in its own right, as discussed by Knuuttila, Rusanen and Honkela (submitted for presentation in this Special Session).

Still, from a cognitive science point of view the most interesting question is still whether neural networks can shed light on the relation of mind and brain. Might a clearer understanding of the relationship between neural and cognitive computational models be attained if we looked at cognitive modeling from the point of view of templates?

B. Neural Networks and Systems of Symbolic Logic as Models Based on a Common Template

Here we consider as a case study one theory from current neural network research, for which Humphreys approach seems particularly useful. However, the example is at the same time in some interesting ways rather different from the examples Humphreys describes.

Here is the theory: Leitgeb [7] shows how to represent artificial neural networks as interpreted dynamical systems, and to establish a correspondence between the dynamics of the network and a system of logic governing the inferential relations between interpretations of network states. Without going into detail the outline of the relevant properties of the systems can be sketched as follows (for details see [7,8,9]): States of neural networks are assigned interpretations, these interpretations being represented as propositional formulae in a symbolic logical language (not all states need have an interpretation). A logical system of qualitative laws governing defeasible inferences on the information contents is established. (The idea here is that “less informative” 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 may infer that Tweety flies; this belief may need to be revised if one finds out that Tweety is probably an ostrich or a dodo). Finally, and crucially, the dynamics of the state-transitions of the system are established to respect this information ordering and thus embody the system of inferences. Then it can be said that the dynamical system (specified at the level of a connectionist architecture of interconnected neurocomputational elements) embodies or represents the logical system of inferences, (specified at the symbolic level, by a small set of qualitative laws).

To be a bit more specific, a qualitative law $\varphi \Rightarrow \psi$ (where φ and ψ are sentences of the propositional language in terms of which the interpretation is given and $\Rightarrow$ is a defeasible conditional) is said to hold of the interpreted dynamical system when for every state representing exactly the content φ it holds that the dynamics of the system are such that they will take the system to a stable state (taken to be the outcome of the inference/computation) whose interpretation will contain ψ. (Conversely, the information content ψ is said to be contained in φ just when, with respect to the information ordering, the state whose interpretation φ larger-than-equals the state whose interpretation ψ is).

In summary, then, the parallelism between the dynamics of the neural network and the consequence relation of the system of logic thus defined is such that the dynamics of state transitions conform to patterns of (nonmonotonic) reasoning. In other words, the dynamics of the network represents a system of logic (and *vice versa*).

What, here, is the template? What is the use the template is put to? It seems the best way to characterize the situation is to say that the shared template is neither the network architecture nor the rules of logic. Instead, it is the abstract structure governing both instances. But if this is so, note that establishing the existence of the template is a theoretical result, not part of the initial process of model construction. One may say that the template does not have a canonical description. (Indeed, it does not have any model independent representation until it is shown that there is a more abstract similarity). Also, the point of constructing the template is not to produce a vehicle for prediction – simulation value of such highly simplified models is probably quite meagre. Instead the point of constructing the template is to establish in a precise manner the correspondence between two classes of systems that initially seem rather different.

If this analysis is correct, it suggests uses for Humphreyan computational templates and template construction for the purpose of explanation and the understanding interdisciplinary unification, rather than prediction.

IV. Psychoneural Reduction vs. Emergence

The ultimate goal of cognitive science, arguably, is to understand *how* the cognitive organization and the neurobiological organization of the mind/brain relate to each other.

Generally, cognitive organization is inferred from observations of behavior, using the theories and methods of cognitive psychology, while neural organization usually abstracted from neurophysiological and anatomical observations. But the relation between the cognitive and the neural is considered to be one between ontological levels: the psychological (or behavioral) level and the neural (or physical) level, and the interlevel relations among *properties* at each level. Unification of cognitive and neuroscience would then seem to call some kind of psychoneural identity. Running a neural network simulation would model the brain's “information processing mechanisms” at both levels. In some sense the cognitive process *is* the neurocomputational process (sometimes with the further qualifier “nothing but”).

The traditional approach to the unity of science recognizes two main angles of attack: methodological unification (unity of the scientific method) and explanatory unification, based on intertheoretic reduction. The computational template-approach cross-cuts these familiar alternatives in the following way: The use of the same computational templates across disciplines resembles methodological unification, but no pretense is made to the effect that the same templates should be used in all, or even very many, sciences.

Templates provide explanatory unification in the sense that the same sort of computational means can be used to model and explain phenomena in different disciplines. However, use of the same template in modeling different domains does not mean that one domain would be reduced to the other. No pretense is made to the effect that the unity of science would require the explanation of (phenomena of) higher level sciences in terms of some more fundamental lower level science. No reductionist scientific ontology is assumed, whereby one discipline (physics, neuroscience) would have an priori preferred status vis-à-vis some other (psychology), whose vocabulary would need to be reinterpreted as merely convenient “dramatic idiom”.

Organization of neurocognitive science around templates gives a picture of unification that is neither reduction nor really emergence either.

The cognitive level does not, ontologically speaking, seem to contain “novel” emergent properties that would be unspecified at the lower level. The cognitive level is just a particular kind of description of the neural networks’ dynamics. One that it is particularly straightforward to relate to psychological theories.

But as far as unity of science is concerned, template based unification is not really reduction of higher level theories or phenomena to lower level ones, either. In fact, the question does not arise since the templates do not belong to any particular “level of organization”. Unlike for a reductionist, the neural network level is not “more basic” (in terms of explanation, ontology or causation), and the cognitive level is not reducible to the neural level (the qualitative laws account gives a sort of multiple realizability). The cognitive level is not derivable from the neural level, either, except via template construction - but the template (and the existence of one) is scientific knowledge in its own right.

V. Conclusion

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 a case theoretical cognitive science from broadly Humphreyan point of view. It was noted that the use of and motivation for constructing the template seems to differ somewhat from the features emphasized by Humphreys, but in a way that nevertheless seems to shed light on intertheoretic unification in the neurocognitive sciences. 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 [10] for a discussion on the asymmetry of explanation and prediction).

The ultimate goal - only tentatively pointed at here - would be to see whether, empirically, the approach generalizes to the computational architecture of cognition and real brains, and whether, philosophically, the template-based view of the unity of science applied to empirical theories in computational and systems neuroscience.

Acknowledgment

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

References

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

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

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

[4] T. Kuhn, The Structure of Scientific Revolutions, 2nd. ed. Chicago: University of Chicago Press, 1970.

[5] I. Lakatos, "Falsification and the Methodology of Scientific Research Programs," I. Lakatos and A. Musgrave, Eds., 1970.

[6] T. Horgan and J. Tienson, Connectionism and the Philosophy of Psychology. Cambridge, MA: MIT Press, 1996.

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

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

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

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