Self-Organizing Maps as Traveling Computational Templates

Tarja Knuuttila, Anna-Mari Rusanen, Timo Honkela

Abstract— **In this article we approach neural networks as computational templates that travel across various sciences. Traditionally, it has been thought that models are primarily *models of* some target systems: they are assumed to represent partially or completely their target systems. We argue, instead, that many computational models cannot easily be conceived of in representational terms. Rather, they can be seen as *models for* various epistemic endeavors. Apart from dealing with the question of representation, we discuss also what implications the genuinely cross-disciplinary computational templates such as neural networks have for the organization of science. We use Self-organizing maps as an example through which we study the aforementioned questions.**

I. INTRODUCTION

"Our slogan will be mathematics, not logic; computation, not representation; machines, not mentation."
– Paul Humphreys (2004).

Models play various roles in science. They can be used to make precise predictions and represent some physical systems more or less accurately; some models are used to analyze, summarize and visualize data; and they can be also used as a tool for designing experiments and testing the results. Computational models, and neural networks among them, are good examples of models that are used for all these purposes.

Models typically do not only have various roles within a scientific domain, but they also travel: successful models cross the boundaries of different scientific disciplines. This is indeed an old observation in the history of science. Often the traveling happens through borrowing an already successful formalism from one scientific domain and applying it to another scientific domain. This phenomenon is especially conspicuous in the case of computational models, where the same formal representations are applied across radically different subject matters. Recently, philosopher of science, Paul Humphreys [1] has emphasized, that these computational templates provide a key feature of computational science. In his opinion computational templates, not theory, scientific laws, research programmes

T. Knuuttila is with the Academy of Finland and Department of Philosophy, P.O. Box 9, 00014 University of Helsinki, Finland (e-mail: tarja.knuuttila@helsinki.fi).
A-M. Rusanen is with Department of Philosophy, P.O. Box 9, 00014 University of Helsinki, Finland (e-mail: anna-mari.rusanen@helsinki.fi).
T. Honkela is with Adaptive Informatics Research Centre, Helsinki University of Technology, P.O. Box 5400, 02015 TKK, Finland (e-mail: timo.honkela@tkk.fi)

or models, provide the proper unit of analysis of computational science. With computational templates Humphreys refers to mathematical formalisms that are computationally tractable. According to him "clean, abstract presentations of formal schemata disguise the fact that the vast majority of theoretical templates are practically inapplicable in any direct way even to quite simple physical systems" [1]. Computational science means for Humphreys a turn to tractable numerical mathematics from analytical mathematics, and what is more, the successful computational templates can be used to model completely different subject matters. Inspired by his work we will consider in the following neural networks as *traveling computational templates* that have a variety of applications in different scientific domains.

That neural networks furnish an apt example of traveling computational templates is already shown by their history. Initially, most neural networks were inspired by the idea of looking at brains as a model of a parallel computational device, but nowadays neural networks are applied in several different scientific domains, lying most often far away from the domain of neuroscience. They are, for example, frequently used in various applications of control theory, including fields such as engineering, biology, neurosciences and economics. Or, to mention another flourishing area, they are used in different sort of tasks that require time series analysis and prediction.

There are several reasons for this fertility. One of them is clearly the nature of the neural networks research. The research strategy of neural networks is to develop and create simplified mathematical models of generally brain-like adaptive systems, and to study the properties of these models to understand how various computational problems can be handled by them. These tasks are not necessarily ones that are faced by real biological brains, and neither do they describe realistically the way the actual biological brains would handle them. They have offered, however, a powerful way to model various processes and data. Since research in every field of science can, at least to some degree, be cast as a class of some sort of computational problems, it is understandable why the neural network research has been able to branch out to so many sciences.

From a philosophical perspective, this raises some interesting questions. Especially, philosophers should find it worthwhile to evaluate the discussion concerning the nature of scientific models from this viewpoint. Traditionally, philosophers of science have assumed that scientific models are *models of* something: they have taken models to represent some specific target systems in some relevant respects and aspects. However, granting that neural networks

have several different applications in various scientific domains, it is difficult to understand why they, as a class of mathematical structures, should be seen primarily as *models of* some pre-defined target system, their "brain-like" initial motivation notwithstanding. Instead we argue that they could be seen as computational templates that can be applied to many purposes.

Apart from the questions concerning models and representation, our ability to use and recycle same known computational tools across disciplinary boundaries has also significant consequences for how we envisage the organization of science. Perhaps, as Humphreys suggests, parts of the theoretical science could be reclassified on the basis of which computational templates they use [2].

In what follows, we will explore the aforementioned questions concerning (i) the role of representation in modeling and (ii) the role of computational templates as having a potential integrative role in current sciences. To render our presentation more concrete, we will look more closely at one specific type of neural network, the Self-Organizing Map, (the SOM).

II. SELF-ORGANIZING MAPS

A. *The Architecture of Self-Organizing Maps*

The architecture of a Kohonen self-organizing map can be summarized as follows [3,4]: It consists of a set of laterally interacting adaptive processing elements, adaptive prototypes. They are usually arranged as a two-dimensional grid called *the map*. Every node of the map is connected to a common set of input. Any activity pattern on the input gives raise to excitation of some local group of map nodes. Since there is no need for a priori classifications of the input, the learning of the map is unsupervised. The learning process is based on similarity comparisons in a continuous space. After the learning process, the spatial positions of the excited groups specify a mapping of the input onto the map. The resulting system is a system that maps similar inputs close to each other in the resulting map.

More technically, starting with an initially random set of prototypes, the self-organizing map algorithm gradually adjusts them to reflect the clustering of the training data. In setting up the self-organizing map, one first assigns to each unit in the array a parameter vector $m_i(t) \in R^n$ called the prototype vector. It has the same number of elements as the input vector $x(t)$. The initial values of the parameters can be selected at random [3,4]. The following process changes these parameters. The image of an input item on the map is defined to be in the location of the model vector $m_i(t)$ which matches best with $x(t)$ in some metric. The self-organizing algorithm that creates the ordered mapping can be described as a repetition of the following basic tasks [3,4]:

- An input vector $x(t)$ is compared with all the prototype vectors $m_i(t)$. The best-matching unit on the map, i.e. the unit where the parameter vector is most similar to the input vector in some metric, called *the winner*, is identified.
- The prototype vectors of the winner and a number of its neighboring units in the array are changed incrementally according to the learning principle.

The idea of the self-organizing map can thus be summarized as follows: For each input sample vector $x(t)$ the parameters of the winner and units in its neighborhood are changed closer to $x(t)$. Adaptation of the prototype vectors in the learning process takes place according to the following equation:

$$m_i(t+1) = m_i(t) + \alpha(t)[x(t) - m_i(t)] \text{ for each } i \in N_c(t),$$

where t is the discrete-time index of the variables, the factor $\alpha(t) \in [0,1]$ is a scalar that defines the relative size of the learning step, and $N_c(t)$ specifies the neighborhood around the winner [3,4]. At the beginning of the learning process the radius of the neighborhood is large. It shrinks during learning. This ensures that the global order is obtained already at the beginning, whereas towards the end, as the radius gets smaller, the local corrections of the prototype vectors in the map will be more specific.

B. *The applications of SOMs*

Originally self-organizing maps were regarded as artificial models of experimentally found ordered "maps" in the cortex. There is some neurophysiological evidence to support the idea that the self-organizing map can serve as an abstract model of some of the fundamental adaptive processing principles of the brains [4]. However, since then SOMs have been used to model phenomena in various scientific fields, including real-life applications, such as, speech recognition [3], image analysis [5,6], recognition of handwritten characters [7], decision support [8], financial analysis [9], information retrieval [10] and process monitoring [11,12]. To furthermore indicate the range of applications, the SOMs are being applied by the authors to analyze the behavior of consumers and to visualize data about the conceptual organization of the students of philosophy.

III. THE NON – REPRESENTATIONAL NATURE OF COMPUTATIONAL MODELS

That a model can have so many and various applications, raises in our opinion some significant philosophical issues concerning the nature of models and how they give us knowledge. Both questions have traditionally been answered by philosophers of science by reverting to representation. On one hand models are considered to be representations and, on the other hand, they are thought to give us knowledge

because they represent. To understand how this representational conception of models came to be held by the overwhelming majority of philosophers, let us quickly review the dominant views on models of the last century.

Traditionally, it has been theories and not models that have occupied the center stage in the philosophical discussion. Models have been defined according to what kind of a relation they have towards theory. Thus according to the so-called syntactic view of theories, the first one of the two 20th century dominant views concerning theories and models, a model is an entity that is designed to give an interpretation to a theory. For the proponents of this view, a scientific theory in turn was a partially interpreted or uninterpreted formalism or calculus i.e. a syntactic structure consisting of a set of axioms. The models were relegated the task of interpreting the theoretical structures: to interpret a theory was to specify a model for it. Models were also expected to give more flesh to abstract formalisms in terms of familiar concepts and visualizable materials.

The semantic conception that challenged the aforementioned "received view" becoming itself in turn the received view, was inspired by the mathematical model-theory and emerging computer science. The semantic conception of theories contested the syntactic view by replacing the syntactic formulation of the theory with theory's models which were conceived of as non-linguistic entities in contrast to the syntactic view. For the proponents of the semantic conception, models are structures that are either defined by the use of set-theoretical predicates or by the use of suitable mathematical language. The kernel of this semantic, or structuralist, conception of models, is the conception that models specify structures that are posited as possible *representations* of either observed phenomena, or underlying target systems [for overview, see 13,14]. A scientific theory, according to this view, consists of a set, or a "family" of models. This structuralist conception emphasizes that the representation provided by a model can be understood as a correspondence between the model and its target system. The correspondence was analysed in terms of isomorphism or similarity [for example, 13,15].

In the recent research in the philosophy of science it has been shown, however, that neither isomorphism nor similarity can provide an adequate analysis of scientific representation. They lead to well-known problems. Firstly, the isomorphism view in fact assumes that there is no such thing as false representation, either the model and its target system are isomorphic, or then they are not, in which case there is no representation either. Secondly, both isomorphism and similarity are symmetric relations, which runs counter our intuitions concerning representation: we want a model to represent its target system but not vice versa. Both problems appear to be solved once the pragmatic aspects of representation are taken into account. The users' intentions create the directionality needed to establish a representative relationship; something is being used and/or interpreted as a model of something else. Taking into account human agency introduces also indeterminateness into the representative relationship: human beings as representers are fallible.

Consequently, pragmatic approaches to representation solve many problems of the structuralist notion of representation—but this comes with a price. When representation is grounded primarily on the specific goals and representing activity of humans as opposed to the respective properties of the representative vehicle and its target system, nothing very substantive can be said about representation in general: There is nothing in the *nature of* the representation (the model) and its target system that guarantees the representative relationship between the two. But if nothing substantive can be said about representation in general, then to regard models as representations does not tell us much about their cognitive value either. Whether or not this is an unwelcome consequence depends on to which degree we attribute the knowledge-bearing aspects of models to representation.

Indeed, one way to interpret the recent practice-oriented approaches to scientific models is to regard them to be in the process of breaking away from the too restrictive representational paradigm in favor for other ways to approach models.

We suggest that looking at models from the point of view of formal templates they share and embody provides such a new approach to modeling. We do not of course deny that in many cases models are made to represent some target system or object. However, it seems to us that the representational character of for instance such computational models as neural networks should not be overplayed. There are several reasons for that. Firstly, the representational properties of neural networks do no explain their usefulness and versatility across different domains. Secondly, even if neural networks were taken as representations of some aspects of human or animal brains, there are also many known facts about neural information processing that are not modeled in any given connectionistic model. Thus, to the extent that models are actually *directly* relevant to understanding what goes on in real brains, they are at best coarse approximations of some details of neurophysiological processing [16].

This is true of self-organizing maps as well. In their inception SOMs were based on a few simplified approximations of the neurophysiological processes at a very abstract level of description. Thus they carry some initial analogy to the brains, but it is not so clear, whether they represent the human brain in any particularly revealing or interesting way.

More importantly, even though SOMs were originally inspired by the neurophysiological structures of the cortex, that fact does not explain their success in other domains. What is more, when SOMs are used in the fields of inquiry that lie quite afar from the neurophysiological research of the cortex, to conceive SOMs as representations becomes vague. Instead, the SOM models reveal statistical structure in the

data. To do this they rely on a "neurally inspired" algorithm, but this fact does not really make the SOM a representation of a neural representation of the domain of interest. If it represents anything, then it represents the data in a certain way. In this SOM models are like simulation models in general, since often they are first and foremost appreciated for the *output representations* they produce. There was also a specific reason to consider the SOM as a sample of a neural network model rather than for instance, a backpropagation algorithm. Namely, as the SOM applies unsupervised learning paradigm, the end result of the analysis reflects relatively more the contents of the data than any supervised learning model that imposes the predetermined output categories on the analysis.

When a computational template such as the SOM is applied to a new domain, the modelers' primary goal is usually to get their models to produce proper output, or to reveal something interesting about the modeled data or the underlying system that has produced that data. The application of the SOM can, depending on the task at hand, be considered as (1) an artificial neural network model, (2) a statistical machine learning method, (3) a tool for data analysis and visualization within data and text mining area or (4) a model component for some cognitive functions [21]. Thus it is not enlightening to envisage the applications of SOMs, or any other similar computational models for that matter, as clear cut representations of any specific external target systems. They from rather a class of multipurpose things that can play various roles in scientific endeavor and are flexible enough to be applied to diverse computational tasks in different domains. Maybe it is this flexibility that has lead to the current situation: the SOM is used practically in all scientific disciplines. A bibliography of the research on SOM theory and applications has been collected and it includes currently over 7700 publications [17, 18, 19].

As has been argued by Knuuttila and her colleagues [20, 21], taking seriously this phenomenon is bound to change the way we conceive the epistemic (or knowledge-bearing) nature of models. As we have already mentioned, models have traditionally been seen as ready-made representations of pre-defined target systems. Yet it seems that in many cases their epistemic value could be better appreciated if they were rather conceived of as *epistemic artifacts* that bring us knowledge in many ways [20, 21, 22]. Calling models epistemic artefacts means that they can be conceived of as intentionally constructed things that are materialized in some medium and used in our epistemic endeavor in a multitude of ways. Thus the very same model can variously take the roles of a *tool*, of a *research object* or *inferential device*. These different roles of models are also discernible in the case of SOMs. As a tool the SOM can used to represent data, in which uses the modelers need not be interested in the very characteristics of the SOM method itself using it more or less in a black box manner. This can be contrasted with those uses of the SOM in which the developers of the SOM method are interested of the SOM as an evolving research object in its own right. Also in these uses the SOM may be applied to various kinds of data, but the main focus is on the applicability and the various characteristics of the SOM method itself. The SOM can also be used to *infer* various things about our linguistic and cognitive skills. The inferential role of models comes closest to that of representation, but does not assume that the model would be an accurate or even "right" representation of its target system [23]. (It is a common phenomenon in the sciences that even false models can give us knowledge). The discussions in the cognitive science concerning the implications of the connectionist modeling methods as regards to our cognitive architecture serve as a good example of the use of the neural networks as inferential devices. Thus we suggest that rather than being a model of, or a representation of, anything in particular, neural network models could be fruitfully approached as epistemic artefacts that unfold into many different epistemic tools, objects and inferential devices, depending on the task at hand. Representation is just one aspect of these uses.

This artefactual and multipurpose nature of neural networks becomes even more evident, if one focuses on the role the computational templates play in neural networks. In a nutshell, a computational template is a computationally tractable template, which is the "heart" of a computational model. These computational *templates* are not models, models are too specific in embodying already an interpretation and data to function like templates, viz. as pieces of formalism that can rather easily be transferred from one discipline to another [1,2]. The detailed subject-specific considerations, construction assumptions, approximations and idealizations that make part of the constructing of models make them also context-specific. Computational templates are in contrast general and context independent, at least to some degree. They are not yet models of anything in particular but can be turned into such.

As a consequence, computational templates such as the SOM can be also considered from a purely syntactic point of view without any empirical interpretation [1,2]. Thus syntactically isomorphic templates with different interpretations should not be considered as re-interpretations of the same model, but they are completely different computational models [1,2]. Thus, even if a computational template can be used to model more than one system type, the resulting computational models will be distinct, if the interpretations of the syntactically isomorphic templates are different.

IV. COMPUTATIONAL TEMPLATES AS SOURCES OF UNIFICATION

We also suggest in line with Paul Humphreys that if philosophers of science adopted the genuinely cross-disciplinary computational templates as units of their

analysis, it would have potentially far-reaching implications. It might lead them to rethink the issues concerning the organization of science, the relationship between different disciplines and the nature of scientific knowledge. Our ability to use and reuse the same computational templates across various disciplines could be taken as a starting point for studying the organization of science, because, as Humphreys puts it, "the emphasis on the sameness of mathematical form has significant consequences for how to conceive of scientific domains" [2]. Thus computational templates provide a new perspective to consider the issue of unification across various disciplines.

With the unity of science or unification it has been referred to many things, but the classical accounts of unification typically fall into a two categories, i.e. to the methodological and explanatory unification.

By methodological unification is meant a search for such unificatory methodological principles that could offer a common methodology for all sciences. The classical demarcation criteria suggested by Karl Popper or the verification thesis advocated by the early logical empiricists, are examples of this style of unification.

A classical example of explanatory unification is the idea of inter-scientific reduction, in which the approach is to characterize the unity of science in terms of the ability to explain (the phenomena of) higher level sciences in terms of some lower and fundamental level science (i.e. ultimately physics). The unity of science will be achieved through such reduction of higher-level sciences to lower level sciences. This emphasizes the unity of science as unity of scientific ontology and vocabulary, which is achieved by reinterpreting the idiom of the reduced theory in terms of the reducing theory.

There are many good reasons for not to adopt these conceptions of unification if only because they do not seem to be supported by the actual scientific practice. In consequence, philosophers of science have attempted to develop other models of unification. For example, in the context of neurosciences, Craver has recently defended the so called mosaic model of unification [22]. The central idea of the mosaic model is that the unification is done not by the reduction of all explanations to a fundamental level, but rather by using results from different fields to constrain a multilevel mechanistic explanation [22]. This is similar to the template-based approach in spirit, but differs in that it is the domain-specific *contents* rather than the abstract computational *templates* that are considered to be the vehicles of interdisciplinary cross-pollination.

The computational templates seem to serve neither as a source of common methodology nor as explanatory elements in science. The reason for the former is quite transparent. Templates as such do not have any empirical content and as a consequence they cannot be used as a source of common empirical methodology. Thus the mere templates cannot be considered as the purveyors of methodological unification in the classical sense. Moreover, as bare formalisms computational templates do not serve as sources for explanatory power either.

They are, at best, parts of theories (or models) that do the explanatory job after the interpretation is done. Thus the computational templates as such seem not to have anything to do with the explanatory notions of unification such as the inter-level reduction or the mosaic model of unification. In both cases the explanations, not the computational templates inside them, are the source of unification.

Consequently, there is a need for a quite different kind of conception of scientific unification or organization. Perhaps this kind of unification has some similar features to the unificatory power of taxonomies. Sometimes taxonomies are partially borrowed from the other disciplines, and they offer conceptual tools for reorganizing and slotting items into more and more accurate categories. They seem to bring unity to science without explaining, since – as Carl Craver once put it – "merely slotting is not explaining" [24]. However, if in an attempt to comprehend unification we target *scientific understanding* instead of explanation, rather different kinds of conclusions are seen to emerge. Namely, it is often claimed that explanations explain because they make the phenomena intelligible and thus the weight is shifted to the problem of understanding. Interestingly, one influential view on scientific understanding approaches intelligibility as that of seeing connections and common patterns in different situations. According to Philip Kitcher "science advances our understanding of nature by showing us how to derive descriptions of many phenomena using the same patterns of derivation again and again" [25]. The computational templates *do* bring initially very remote fields of inquiry a bit closer to each other by using the same computationally tractable templates, and if the templates used prove successful that is bound to lead to further questions of how and in virtue of what the various phenomena might be similar. Moreover, the templates, though syntactic structures, are not innocent and neutral. In many cases they are carriers of certain kinds of metaphysical assumptions. Thus for instance the SOMs model their target systems as self-organizing and adaptive. There is ample evidence already that the various computational templates adopted by different disciplines tend to refashion their problems and lead to new conceptualizations, new questions to be asked and even to the emergence of new disciplines.

V. CONCLUSIONS

In this paper we have discussed the nature of computational models and templates. As an example we have used the self-organizing maps. We have argued that being computational models designed to execute certain computational tasks, the various applications of SOMs do not as such represent any target systems in the world, not at least in any straightforward way. Moreover, we have suggested that they should be rather conceptualized as multifunctional epistemic artifacts. More generally, the traditional philosophical view according to which models are first and foremost representations of some pre-defined target systems does not capture what seems to us the characteristic feature of modeling: the use of inherently cross-disciplinary

computational templates. What indeed could the ideal of accurate representation mean in the case when the same computational templates are used in various disciplines dealing with totally different natural or social phenomena. We have also discussed the role of computational templates in organizing and unifying the different disciplines. It seems to us that this question opens up one of the most intriguing topics concerning interdisciplinary research: how do computational templates redesign the fields to which they are applied.

ACKNOWLEDGMENT

REFERENCES

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

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

[3] T. Kohonen, "Physiological Interpretation of the self-organizing map algorithm", *Neural Networks*, 6/7, 1993, pp. 895-905.

[4] T. Kohonen, Self-Organizing Maps. 3rd edition, Springer, 2001.

[5] A. Visa and J.Iivarinen, "Evolution and Evaluation of a Trainable Cloud Classifier", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 5, 1997, pp. 1307-1315.

[6] V. Vuori, J. Laaksonen, E. Oja, and J. Kangas, "Experiments with adaptation strategies for prototype based recognition for isolated hadwritten characters", *International Journal on Document Analysis and Recognition*, vol. 3, 2001, pp. 150-159.

[7] E. Carlson, "Self-Organizing Feature maps for appraisal of land value of shore parcels" in *Proceedings of ICANN'91*, International Conference on Artificial Neural Networks, vol II, ed. Kohonen et al., North Holland, 1991.

[8] B. Back, K. Sere and H. Vanharanta, "Data Mining accounting numbers using self-organizing maps" in *Proceedings of SteP`96*, Finnish Artificial Intelligence Conference, eds. Honkela et al, Finnish Artificial Society, Finland 1996: pp. 35-47.

[9] G. DeBoeck and T. Kohonen, *Visual Explorations in Finance with Self-Organizing Maps*, Springer, 1998.

[10] K. Lagus, T. Honkela, S. Kaski, and H. Lappalainen, "WEBSOM for textual Data Mining", *Artificial Intelligence Review*, vol. 13, 1999, pp. 345-364.

[11] T. Kohonen, E. Oja, O. Simula, A. Visa and J. Kangas, " Engineering Applications of the Self-Organizing Map", *Proceedings of the IEEE*, vol. 84, no.10, 1996, pp. 1358-1384.

[12] O. Simula, J.Ahola, E. Alhoniemi, J. Himberg, J. and J. Vesanto, " Self-Organizing Maps in Analysis of Large.Scale Industrial Systems" in *Kohonen Maps*, eds. Oja et al., Elsevier 1999, pp. 375-387.

[13] B. van Fraassen, *The Scientific Image*. Oxford: Oxford University Press, 1980.

[14] S. French and J. Ladyman, "Reinflating the Semantic Approach", *International Studies in Philosophy of Science* 13, vol 2., 1999: pp.103-121.

[15] R. Giere, *Explaining Science: A Cognitive Approach*. Chicago and London: 1988

[16] D.E. Rumelhart and J.L. McClelland, "PDP models and General Issues in Cognitive Science", in *Parallel Distributed Processing, Explorations in the Microstructure of Cognition*, Vol. 1. Foundations, Rumelhart et al., ed. Mass: the MIT Press, 1986, pp.

[17] S. Kaski, J. Kangas and T. Kohonen (1998). Bibliography of self-organizing map (SOM) papers: 1981-1997. *Neural Computing Surveys*, 1: 102-350.

[18] M. Oja, S. Kaski and T. Kohonen (2003). Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001 Addendum. *Neural Computing Surveys*, 3: 1-156.

[19] M. Pöllä, T. Honkela and T. Kohonen (2007). Bibliography of Self-Organizing Map (SOM) Papers: 2002-2005 Addendum. *Neural Computing Surveys*, forthcoming

[20] T. Knuuttila and A. Voutilainen, "A Parser as an Epistemic Artefact: A Material View on Models", *Philosophy of Science*, vol. 70, pp. S1484-S1495

[21] T. Knuuttila and T. Honkela, "Questioning External and Internal Representation: The Case of Scientific Models", in *Computing, Philosophy, and Cognition*, L.Magnani ed., London: King`s College Publishing, 2005.

[22] T. Knuuttila, "Models, Representation, and Mediation". Philosophy of Science , 72., 2007.

[23] T. Knuuttila, "From Representation to Production: Parsers and Parsing in Language Technology", in *Simulation: Pragmatic Constructions of Reality. Sociology of the Sciences Yearbook*, J. Lenhard, G. Küppers and T. Shinn eds., New York: Springer, 2007.

[24] C. Craver, *Explaining the Brain: What a Science of the Mind/Brain could be*. Forthcoming.

[25] P. Kitcher, "Explanatory Unification and the Causal Structure of the World", in *Scientific Explanation*, Minnesota Studies in the Philosophy of Science, vol. XIII, P. Kitcher and W. Salmon eds., Minneapolis, University of Minnesota Press.