Nersessian, N.J. & Chandrasekharan, S.: “Hybrid analogies in conceptual innovation in science,” *Cognitive Systems Research*, Special issue: Analogies - Integrating Cognitive Abilities, in press, 10:3, 2009.

Hybrid Analogies in Conceptual Innovation in Science

Nancy J. Nersessian,
School of Interactive Computing,
Georgia Institute of Technology, Atlanta, USA
nancyn@cc.gatech.edu

Sanjay Chandrasekharan,
Cognitive and Motor Neuroscience Lab,
Faculty of Kinesiology,
University of Calgary, Alberta, Canada
schandra@kin.ucalgary.ca

Corresponding Author:
Nancy J. Nersessian,
School of Interactive Computing,
Georgia Institute of Technology
801 Atlantic Drive,
Atlanta, GA, 30332-0280 USA
nancyn@cc.gatech.edu

Abstract

Analogies are ubiquitous in science, both in theory and experiments. Based on an ethnographic study of a research lab in neural engineering, we focus on a case of conceptual innovation where the cross-breeding of two types of analogies led to a breakthrough. *In vivo* phenomena were recreated in two analogical forms: one, as an *in vitro* physical model, and the other, as a computational model of the first physical model. The computational model also embodied constraints drawn from the neuroscience and engineering literature. Cross connections and linkages were then made between these two analogical models, over time, to solve problems. We describe how the development of the intermediary, hybrid computational model led to a conceptual innovation, and subsequent engineering innovations. Using this case study, we highlight some of the peculiar features of such hybrid analogies that are now used widely in the sciences and engineering sciences, and the significant questions they raise for current theories of analogy.

Keywords: Conceptual innovation, hybrid analogies, simulation, visual reasoning, engineering sciences

1. Introduction

It has long been held in cognitive science that scientific problem solving lies on a continuum with the way people use reasoning to solve more ordinary problems. Much research drawing from the history of scientific practices provide studies of how scientists extend and refine basic cognitive strategies, in explicit and critically reflective attempts, to devise methods for probing and understanding nature.[1] Until recently, however, not much study has been devoted to studying the cognitive practices of scientists *in vivo*.[2] Given the sophisticated cognitive requirements for creative scientific research; the ill-formed, evolving nature of the problems; and the consciously reflective nature of scientific problem solving, studying science practices provides valuable data for developing an integrative understanding of high-level cognition. Further, studying the science end of the continuum provides a novel window on the mind, and can lead to a deeper understanding of the potential of human cognitive capacities. Since creative scientific practices are of far greater complexity and sophistication than those customarily studied in current cognitive investigations, we cannot assume that they can fully be explained by means of current psychological theories or computational implementations, which are based mostly on experimental studies of mundane cognition. Rather, these

[1] See, for example Andersen et al. 2006, Darden 1991; Gentner et al. 1997; Giere 1988; Gooding 1990; Griesemer 1991; Griesemer and Wimsatt 1989; Nersessian 1984, 1992, 2008; Rudwick 1976; Shelley 1996; Spranzi 2004; Thagard 1992; Tweney 1992, 2004.

[2] Kevin Dunbar (1995) pioneered what he called the In vivo/In vitro method of observing science in practice and then conducting traditional psychological experiments on findings of interest. More recent in vivo research include Christensen & Schunn 2007; Nersessian 2005, Nersessian et al. 2003; Trafton et al. 2005; Trickett & Trafton 2007.

practices provide novel data on which to both test the adequacy of those theories and support further theory development. Our primary objective within the confines of this paper is to point to processes that need to be incorporated into accounts of analogy, in order to explain the cognitive practices in science. Our focus here is on conceptual innovation, since self-generated, sophisticated, and reflective use of deep analogies figures prominently in many instances of conceptual innovation across the sciences. Such usage often involves processes of building models (conceptual, physical, or computational), visualization, and dynamic mental simulation.

Understanding scientific cognitive practices will doubtless require integration of research in fields currently treated as separate areas of research; for instance, research on analogy, imagery, mental modeling and conceptual change, which are not yet treated in an integrated fashion. Science provides many instances for which such division is artificial, and calls into question that separation. Take, for instance, the simple analogy between the motion of a stone dropped from the mast of a moving ship and the motion of a ball dropped from a tower on the earth that Galileo made in developing the notion of relative motion. This analogy also provides the basis of a dynamic mental simulation ("thought experiment"). Or consider Newton's analogy between the motion of a projectile thrown with successively greater velocity from a mountain rising high above the surface of the earth and the orbits of planets and the moon. The analogy makes use of a sketch related to a thought simulation of the successive paths of the projectile from the mountain to the earth, to the escape point where it, too, would orbit the earth under the effect of centripetal force.

There has been considerable research on the various processes which analogy is thought to comprise, with most research focusing on retrieval, mapping, and transfer.[3] The customary idea of problem solving by analogy is that one recognizes some similarities between the problem situation under consideration (target) and something with which one is familiar and is better understood (retrieval of source). One then creates a mapping between the two that enables solving the original problem (mapping and transfer). In the process, it is likely that the source and target need to be re-represented in order for the comparison or mapping to be made. Take, for example, the solar system (source) - atom (target) analogy, usually - though incorrectly - attributed to Rutherford. In order to posit that similar forces are keeping the entities in orbit, the sun and nucleus need to be re-represented more abstractly along the lines of “centrally located entities” and the planets and electrons as “entities that revolve around centrally located entities.” Once each phenomenon is understood in terms of the generic representation of “centrally located entities with entities revolving around them,” one can reason about whether they actually do belong to the same class of phenomena by testing how inferences from one might apply to the other. Relevant dissimilarities might also be exploited, such as that electrons would lose energy while in orbit.

There are many instances of direct use of ready-to-hand analogical sources by scientists, such as Galileo’s analogies between the moon and a wall made of stone, which he used to cast doubt on the assumed perfect nature of the heavenly bodies. But there are also instances in both historical (Nersessian 1984, 1992; Spranzi 2005) and protocol (Clement 1989, Nersessian 2008) studies of creative scientific problem solving leading to conceptual innovation which show that creative use of analogy can involve constructing

[3] See Holyoak & Thagard 1996 for an overview of empirical findings, and also Gentner & Holyoak 1997 for an overview of current theories.

intermediate, hybrid representations bridging the target and source domains, and that such construction and reasoning involve analogical, imagistic, and simulative processes. Here we examine such a case in what have come to be known as the "engineering sciences" – fields in which both basic research and engineering applications are pursued. Our case is drawn from an ethnographic investigation of a neural engineering research laboratory. It captures how a group of researchers developed, over the span of a year, a new way of conceptualizing the propagation of "bursts" of activity in a "dish" of living neurons. Our exemplar of analogy use provides a diachronic description – the constructing, visualizing, and running of a computational model over time. The model performs much the same function in conceptual innovation as the imaginary simulations discussed in the historical and protocol studies noted above.

To our knowledge, the small literature (primarily philosophical) on the nature of computational modeling in science has not considered the practice across time and through the lens of analogical reasoning. But the built models involve many iterations and are analogue systems, constructed so as to enable relational comparisons between structures, behaviors, or functions in the model and the target system. On our interpretation of this case, the computational model of a target system is constructed using constraints from intra-domain analogical sources. It is a hybrid construction of source, target, and model (computational) constraints, in which problems under investigation are solved, and these solutions are mapped to the target problem eventually, and the implications explored. Several conceptual innovations were derived by means of analogy from the model. There are many intriguing features of this case that require a more extended treatment than we can offer here. Specifically, we note that this is a case of what might be called "nested analogy." That is, the computational model provides an *in silico* analogy to an *in vitro* model-system (dish of neurons) that in turn is an analogue model of *in vivo* cortical neural processes (originally, of learning, and now also, disease processes). Further, although the model was constructed initially by one researcher, its interpretation, development, and use in problem solving was by a community of researchers, who together arrived at the conceptual innovations in neuroscience and neuro-engineering. These have led to significant progress in the lab's research, resulting in several prestigious journal publications in the field, and have the potential to be of historical significance.

The case study will outline five features of built analogies, which are not currently addressed by theories of analogy. These are:

1) Role of building: The process of building an analogical model contributes significantly to conceptual change.

2) Creeping mapping: At the point of starting to build the model, there is an analogy with the target domain, but the real nature of this relation is amorphous. Over time, the analogy evolves to develop a stable mapping to the source.

3) Visualizing counterfactuals: The visual structure of the model is not based on an existing structure recognized as the target; it is based on an imagined structure, a counterfactual scenario that is implemented by the building process.

4) Visualization as generator: The visualization of a model serves more as a way of generating many possible solutions, than providing a solution itself.

5) Manifest analogy: A built model can be exploited in communal processes. It can be

manipulated and tweaked by team members at will, and the input and output shared in real-time and compared with the target. This facilitates consensus on the generation and adoption of new ideas and thus focuses the creative work of a team.

2. Analogy in conceptual innovation: Creating the 'center of activity trajectory' (CAT)

Our research group has been carrying out a 6-year study of modeling practices in three interdisciplinary research laboratories in bio-sciences and bio-engineering.[4] A central investigational practice of the researchers in these laboratories is building and experimenting with physical *in vitro* models, through which they study *in vivo* phenomena that are either too difficult or for which it would be unethical to examine experimentally themselves. Thus the research in these laboratories is inherently model-based and analogical. These physical simulation devices are hybrid biological and engineered systems that are referred to as "model-systems." The episode of conceptual innovation summarized here occurred over a two-year period in a neural engineering laboratory, and involved constructing a computational model of an *in vitro* physical model of cortical neural network activity. The model was developed in order to understand certain thought-to-be undesired phenomena - spontaneous "bursts" - taking place with high frequency in the *in vitro* model, but not in properly functioning *in vivo*

[4] We gratefully acknowledge the support of the National Science Foundation ROLE Grants REC0106773 and DRL0411825 in conducting the research on the laboratories. Our analysis of this episode benefited from discussions with members of the research group, especially Wendy Newstetter, Lisa Osbeck, and Christopher Patton. We thank the research lab members for allowing us into their work environment, letting us observe them, and granting us numerous interviews.

animal brains. Significantly, as we will outline below, the computational model led to a reconceptualization of “burst” phenomena in the *in vitro* model, and the development of “programmable” neurons.

We have taken a mixed-method approach to this long-term investigation. Our approach involves conducting both ethnographic investigations of the day-to-day practices, and cognitive-historical analysis of the problems, artifacts, and models used in the research. The ethnographic part of the study (field observations and interviews) seeks to uncover the activities, tools, and interpretive frameworks that support the research as it is situated in the on-going practices of the community. Data from these sources are analyzed using qualitative methods of analytic induction based on grounded coding, broadly construed. The cognitive-historical part of the study, which includes collecting data from the customary range of historical sources, aims to capture the diachronic dimension of the research, by tracing the co-evolving trajectories of the human and technological components of the laboratory. We examine findings from both aspects in light of current cognitive science research, and, reflexively, use these to evaluate, elaborate, and extend cognitive interpretations and theories.

The neural engineering laboratory’s overarching problem is to understand the nature of learning in the brain at the neuronal level. At the time they began, nearly all learning research was carried out on single neurons. The lab director’s central research problem in the period of our investigation was to develop an account of learning and plasticity in networks of cultured neurons, which were thought to more closely model learning in the brain than single neurons. To address this problem experimentally, the lab director previously had developed a method for constructing and investigating plasticity in a model-system: an *in vitro* network of cultured neurons locally referred to as “the dish.” Building this *in vitro* model-system involves extracting neurons from embryonic rats, dissociating them (breaking the connections between neurons) and plating them on a dish with embedded electrodes known as an MEA (multi-electrode array), where the neurons regenerate connections and become a network. The researchers “talk to the dish” by stimulating the neuronal network with different electrical signals (electrophysiology) and developing “embodiments” (physical or virtual “bodies”) that support closed-loop feedback. The embodiments (using a translator program that maps the dish signal to motor commands) include both robotic devices and visualized “animats” that move around in simulated computational worlds.

The stated goal of the research during the period we discuss was to understand the dynamics of learning in the neuronal network in such a way that it would lead to the development of a *control structure,* which would allow the dish to be *trained* to control the embodiment systematically, using feedback. When we began our study, there existed little work on learning and plasticity in such cultured neuronal networks, thus the problem space was wide open. The only theoretical model guiding the lab’s work was the Hebbian rule (roughly stated as: neurons that fire together, wire together). The rule was originally developed for single neurons, but lab researchers assumed it extended to neuronal networks as well. The design of the dish incorporated constraints from the current understanding of neuro-biology and chemistry, as well as those relating to electrical engineering and other dimensions of the technologies used in the lab. As a model-system, “you can map it on to real life in a number of different ways” (Lab Director). To reduce the number of variables in the system, the researchers chose to simplify the model to a single layer of neurons that they believed provide a sufficiently close approximation to yield valid information. Note that this technique of building an *in vitro* model was developed and optimized over years of research by the lab director. Also, creating every instance of a dish involves weeks of 'bench' work. The dish by itself is thus an instance of diachronic analogy.

The dish is studied as a model of basic neurological processes in brain development and learning. In the episode we describe here the dish itself was the target of the problem solving. The researchers had been "playing with the dish," which is their term for exploring the problem space through stimulating the neuronal network using different electrical signals and tracking the output. The work began by trying to replicate a plasticity result reported by another group. This experiment was initiated by researcher D4, but she was not successful in replicating the reported results. One of the problems she faced was "bursting" - or what she termed "barrages" - a form of network wide electrical activity spontaneously exhibited by the *in vitro* neuronal networks. Figure 1 shows burst activity in each channel of the MEA dish as recorded on an oscilloscope.

[Insert Figure 1 here]

Bursting created a problem in understanding plasticity, because it prevented the detection of any systematic change that arose due to controlled stimulation of the network. According to D4, whenever she conducted a plasticity experiment, the network produced bursts. The lab interpreted those as "noise in the data…noise interference in the way…so it is clouding the effects of the learning that we want to induce." The group hypothesized that bursting arose because of deafferentation; that is, because the neurons lacked the sensory inputs they would ordinarily get if they were in a live animal's brain. Given this view, and the problem of the noise generated by bursts, D4 decided that she needed to get rid of bursts, and began working on "quieting" bursts in the dish. She hypothesized that it would be possible to lower bursting by providing the network with artificial sensory input, that is, some form of electrical stimulation. Interestingly, this revision illustrates the ongoing development and the tentative nature of the dish model. The revision is based on a hypothesis about the nature of the burst, namely that it arises from deafferentation, and is thus similar to noise.

D4 tried a range of stimulation patterns to lower the bursting activity in the networks, and sometime in late 2003, achieved a breakthrough, managing to "quiet" bursts entirely in a neuronal network using a sequence of background electrical stimulation. In subsequent work, it was shown that bursting could be quieted using three types of stimulation: a rapid stimulation in one electrode, a pattern of stimulation spread over several electrodes, and a fine-tuned multi-electrode stimulation using closed-loop feedback. The last was found to provide the best results. As per the group's deafferentation hypothesis, this result was presented as a way of using electrical stimulation to "substitute" for natural sensory input. However, despite the quieted network, for the next six months of research D4 was unsuccessful in inducing plasticity in the network. This was mostly because of a "drift" phenomenon, where the activity pattern evoked by a stimulus did not stay constant across trials, but drifted away to another pattern. This drift prevented her from tracking the effect of a stimulus, because the network never responded the same to a constant stimulus.

During this period another researcher, D2, was working on the embodiment module – the translation between the neuronal network and the motor commands controlling the animats and robots. This was to be part of the closed-loop model-systems to be used in training the dish once they had the sought-after control structure. The control structure, however, was ultimately derived through interaction with the computational modeling simulation we now discuss.

Early in the period when D4 was trying to quiet the network, D11 decided to branch away from working with the *in vitro* model-system entirely, and develop a computational model that mimicked it. As he put it, "*the advantage of modeling [computational] is that you can measure everything, every detail of the network......I felt that modeling could give us some information about the problem [bursting and control] we could not solve at the time [using the in vitro dish model-system]*." D11 felt that to understand the phenomena of bursting he needed to be able to "see" the dish activity at the level of individual neurons, make precise measurements of variables such as synaptic strength, and to run more controlled experiments than could be conducted with the physical dish. This different perspective on the dish illustrates a fundamental property of built analogies – the construction process supports a plurality of designs and views, and also provides a multifaceted approach to the problem. There is no requirement that everyone work with a standard analogical structure, each team member can start from a broad definition of the problem, and build up her own mapping between source and target.

Interestingly, and following from the above, D11 built the initial *in silico* model not on the experimental data from their dish, but drawing from intra-domain sources in neuroscience; in particular, from studies involving single neurons, brain slices, and other computationally simulated networks. His model was tested and optimized with data from other MEA dishes, as well as their own. The model, as developed, is thus a second-order *in vitro* system - an *in silico* model of the activity of a *generic* dish. The computational model is a hybrid construction using constraints deriving from the target (dish) and source (neuroscience), as well as constraints stemming from a neuroscience modeling platform (CSIM), using a standard simple model of neurons, known as “leaky-integrate-fire.” (This name derives from the way the artificial neurons respond to stimulus.)

Importantly, the constraints the model adapted from the target dish were not based on their experimental outcomes (i.e. the behavior of their dish), but had to do with the construction of the dish. These included the area of the artificial neurons, the placement grid of electrodes, the number of electrodes used for recording and stimulation, and the random location of the neurons. The other parameters (constraints) of the model, such as type of synapses, synaptic connection distance, percentage of excitatory and inhibitory neurons, conduction delay, conduction velocity, noise levels, action potential effects, and spontaneous activity were based on results reported in the source literature mentioned above.

After several months of building tentative versions of the computational simulation and probing these, D11started to get what he called a “feel” for how the computational network behaves under different conditions. He then developed a visualization that captured the activity of the network as it ran, which figured centrally in the development of the novel notion of “center of activity trajectory”. As he stated: “*I need to really need to look at the figure to see what is going on. So after I constructed* *this network – I just let it run and we visualized everything in it.*" He was at that point able to successfully replicate, first some of the results reported in the literature, and then results from their own dish, with the computational model.

A major contribution of the visualization is that it enabled D11 to notice - literally see - interesting patterns in the way his model responded to different stimuli. These patterns were novel and distinct from what was known about cultured dishes. The visualization of the network's activity shows the movement of an activity pattern across the entire network, in real time. In the *in vitro* model-system the individual neuronal activity is hidden. In a visual display of the dish, one can see activity across a channel as in Figure 1, but this display can only track activity at each electrode of the *in vitro* system, using graphs they had developed for this purpose. The graphs of activity in the electrodes capture which electrode is activated and by how much, but it does not have a representation of the entire network itself – it does not capture *burst movement across the network*. Thus, it was not possible to see from the dish display whether there were patterns moving across the network, and this meant that they had no specific hypotheses about such patterns. Such patterns, however, did show up in the visualization of the *in silico* model. As expressed by D11, "*I can visualize these fifty thousand synapses…so you can see…after you deliver a certain stimulation, you can see those distributions of synaptic weights change… or the synaptic state change.*" So, the image of changes across discrete channels, as in Figure 1, was replaced by an image of changes of the activity of the whole network as in Figure 2, for instance, from 2a to 2b. (Figure 2c is a statistical distribution of the center of activity trajectory we discuss below.)

[Insert Figure 2 here]

Note that there is nothing that *requires* the visualization to have a network structure, a point easily illustrated by the non-network nature of the older visualization in Figure 1. D11 constructed the network visualization by imagining what the network would look like when it is activated. The visualization he used derives from the way he thinks about the network, the way he imagined the network as behaving. One has to imagine in this case, since the dish is opaque. So he did not design the visualization to see what patterns emerge, he just designed it based on how he imagined networks are activated. He then went on to play around with the visualization to see what patterns emerge. The imagined visualization, then, is a counterfactual scenario, a pattern D11 wished to see and then chose to implement.

The computational model offers other advantages in exploring burst phenomena. The simulated network could be stopped at any point and started again from there. Further, it is possible to provide detailed measures of significant variables, such as synaptic strength, which are not accessible using the *in vitro* model. Finally, a large number of experiments could be run at no cost, since the computational model could be changed easily and does not require the careful and laborious processes involved in setting up and maintaining a living dish. When coupled with the visualization that enables visually tracking the activity of the network as it was happening, these features proved to be a powerful combination. They gave D11 immediate access to a range of configurations and data that the living dish could not provide, and the data could be examined and reexamined, and comparison experiments run instantly. In the process of running numerous simulations, he noticed that there were repeated *spatial* patterns in activity, as the activity propagated across the network. The spatial patterns were seen both when spontaneous bursts arose in the model network, and when the model network responded to stimuli. Basically, he found that there were “similar looking bursts” that propagated across the network, and a limited number of what he called “burst types.” He discussed this finding with D4, who was also having second thoughts about bursts, as she was not getting any results from her plasticity experiments on the burst-quieted network. The group then decided to investigate bursts further, as a possibly interesting pattern, i.e. *a signal*. Note the radical change in perspective here, where “burst” moves from being something akin to noise that needed to be quieted, to the status of a pattern, a signal that could possibly lead to a control structure for training the network. Note also that this change arises from *the group* running many simulations involving different situations and parameters over time, and this group process supported the group to arrive at a consensus on the conceptual shift. This group process was possible because of the manifest nature of the built analogy.

Further work proceeded as a joint effort of the group in large part by mapping potential problem solutions from the *in silico* model to the *in vitro* model. In particular, the various features of the computational model afforded a means of getting past the problem of the drift behavior of the dish. The group came up with a range of ways to quantify the spatial properties of moving bursts, using clustering algorithms and statistical techniques, and these measures proved to be immune to the drift problem. Of particular interest to us here is the fact that they were first developed for the computational model and then equivalents (by analogy) were developed for the *in vitro* model. These included the conceptual innovations of "burst types," "spatial extent" (an estimate of the size and location of the burst in the dish), "center of activity trajectory" (CAT, a vector capturing the spatial location of the electrode along with the firing rate), and "occurrence" of burst types. These spatial measures of bursts were then shown to be better indicators of plasticity than responses to probe stimuli. So, in the matter of a year, based to a significant extent on the patterns generated from the computational model, the group's theoretical position had shifted from bursts as noise to bursts as signal and a possible control structure.

All these measures were driven by the spatial patterns of activity noticed in the visualization of the computational model of dish activity. But of these, CAT is probably the most noteworthy, for two reasons. First, because it is a novel concept, articulated also on analogy with the physics notion of center of mass, which was applied to the neuronal activity pattern. Second, because it emerged entirely from the computational model, and would be almost impossible to conceptualize and properly formalize without it. Figure 2c is a visualization of the CAT for a changing network burst activity in an *in silico* dish. CAT is an averaging notion, similar to the notion of a population vector, which captures how the firing rates of a group of neurons that are only broadly tuned to an action or stimulus (say an arm movement), when taken together, provide an accurate representation of the action/stimulus. However, CAT is more complex than the population vector because it tracks the *spatial* properties of activity as it *moves* through the network. For instance, if the network is firing homogenously, the CAT will be at the center of the dish, but if the network fires mainly at the left corner, then the CAT will *move* in that direction. CAT thus tracks the "flow of activity" (not just activity) at the population scale, and on a much quicker time scale than population vectors.

When applied to the *in vitro* network, CAT provides a novel representation of neuronal activity - a new concept. Figure 3 provides a side-by-side comparison of a CAT visualization that corresponds to an MEAscope representation for an *in vitro* dish. The MEAscope representation (3a) shows activity across each recording channel over time, but the CAT (3b) is an integrated representation of activity across the entire network.

[Insert Figure 3 here]

The CAT is like a “signature” for a burst type in that each burst type has a corresponding range of similar-looking CATs specific to that type. Thus, while the CAT concept arose from the visualization, the relation between the CAT concept and the visualization is not a direct mapping link. The visualization worked as a generator of many types of activity, which when put together, led the new concept.

Although our account ends here, in recent work, the researchers have combined CAT and techniques developed for burst quieting to develop a set of stimulation patterns (a control structure) for the dish that has led to supervised learning by the living neuronal network, in effect making the *in vitro* dish neuron network “programmable.” Using these stimulation patterns, the living network in the dish was trained to control a computationally simulated animat, and extended it to a mechanical drawing arm. Further, as an example of nested analogy, recent work by new researchers in the lab is being guided by a hypothesis that certain neurological disorders, in particular epilepsy, involve burst phenomena. This derives by analogy from the understanding of bursts in the *in vitro* dish, which, as we have shown, derives in part from the analogy with the *in silico* model.

3. Discussion

Our case study provides an instance of creative problem solving where source and target constraints are used to create versions of a hybrid computational model over time, standing in between the source and target domains. At least three analogical source domains (dish, existing literature, CSIM modeling platform) serve as sources of constraints for building different levels of the model. Only in its completed form is it possible to map and transfer solutions from the computational model as potential solutions for the target problems in the dish model. Figure 4 provides a schematic representation of the iterative interactions of these domains with the model.

[Insert Figure 4 here]

The specific problems under investigation required conceptual innovations as part of their solutions. These were derived though several iterations of model construction, evaluation, and adaptation, and were then coupled with visualization and dynamical simulation of the processes represented. Visual representation was an indispensable component of the analogical reasoning processes; most notably it facilitated imagining three dimensional phenomena and processes taking place in time. Predictions were made on the basis of simulative enactments of model behaviors, facilitated by imagistic representations. For instance, D11's visualization of how the simulated burst activity across a network unfolds in time enabled him to predict that there were a limited number of such patterns.

Once completed, the computational model served as an analogy for the dish, but there are several novel insights it contributes to thinking about analogy. These were noted in context in the discussion above, but we enumerate these once more here for emphasis:

1) Role of building: Instead of making a mapping from the dish to another existing problem structure as in the cases traditionally studied in the analogy literature, D11 set out to *build* a structure that would *eventually mimic* the properties of the dish as closely as possible. The parameters of the model come from the source domain literature, but these were not mapped and transferred directly to the target dish. Rather, together with constraints stemming from the way in which the dish is constructed, they were used to build the model. Once the behaviors of the model were determined, a mapping of these to the dish allowed the concepts developed in the model to be transferred to the dish. Since the mapping to the dish happened only at the end of the process (after CAT is developed), a significant part of the conceptual shift arose from the building and running of the visualized simulation.

2) Creeping mapping: The model was tested by seeing whether it exhibits behaviors similar to established results in the literature. Only when the model replicated a series of established studies did D11 turn to replicating results from the dish. This means the model incrementally built up towards being an analogy for the dish, where each increment is evaluated using a success criterion, namely the ability to replicate an established result. It finally became an analogy for the *in vitro* dish after almost a year of constructing, tweaking parameters, and replicating existing results from the literature

first, and then from the dish. At the point of starting to build the model, there is the intention to make an analogy with the dish, but the nature of it is amorphous, and is determined in the future.

3) Visualizing counterfactuals: A significant reason for the computational model's success is its real-time visualization of neuronal activity. D11 built this visualization based on imagined activity in the dish. Each element of the visualization was built by imagining what it might look like, and how the element would capture the activity in the dish. The visualization arises from a counterfactual scenario: "if-only-we-could-see-into-the-dish." Although the visualization eventually supports the analogical transfer of activity patterns to the dish, the visualization is not built based on an existing structure recognized as the target; it is based on wanting to see a wished-for structure.

4) Visualization as generator: As pointed out above, the notion of CAT could not be developed if the model had not revealed the movement pattern of the neuronal activity, using visualization. Note that there is no reason why this movement pattern should map on to the actual dish, since the visualization is just one way of representing the activity of the model. Although there is an analogical relationship between the visualized network and the dish network, this analogy is not productive by itself. The novel notion of CAT arose out of multiple runs of the visualized network, and the mapping of CAT to the dish was not a straightforward process. The visualization thus serves more as a way of generating possible solutions, than providing a solution itself.

5) Manifest analogy: Unlike a the kinds of analogy traditionally studied, the computational model can be exploited in communal processes. It could be manipulated and tweaked by team members at will, and the input and output shared in real-time and

compared with the target. This facilitates consensus on the generation and adoption of new concepts such as the CAT, and helps in forming and refining new extensions and in rejecting false leads. The model thus focuses the creative work of the team.

This kind of problem solving through incremental and public model-building fits the metaphor of a *bootstrapping* method, and the constraint satisfaction processes provide explication of that metaphor: *constraints from target and sources each provide one strap, the intermediary hybrid models are strap crossings, and each crossing supports or contributes to further model building and enhanced target understanding by the entire group*. The problem solution does not arise from a single well-articulated model, but is an incremental process where a constructed model leads to partial target insights and new constraints, which in turn lead to further construction, until an adequate conceptualization is achieved. Looked at this way, it is possible to understand how analogy is a mechanism of conceptual innovation. Constructing the hybrid model representations integrates information from multiple sources specific to the problem-solving situation, which allows truly novel combinations to emerge as candidate solutions to problems. Constraints stemming from targets, sources, and the models themselves can interact in such a way that a model with heretofore unrepresented structures or behaviors can emerge, as in our exemplar.

Our exemplar illustrates that scientific innovation based on analogy can be much more complex and distributed across time, external structures, and people than instances commonly studied by analogy researchers. There is a core set of processes that are shared by everyday analogies and such scientific analogies, a major one being transfer through alignment and mapping of relational structures. In these processes, our exemplars fit the "structure mapping" criteria - first proposed by Dedre Gentner (1983), but now widely agreed upon - for what makes an analogy productive: (1) structural focus (preserving relational structures); (2) structural consistency (making isomorphic mappings between systems); and (3) systematicity (mapping systems of higher-order, interconnected relational structures). There is significant consensus on the criteria for what makes for productive analogies across the literature in cognitive science. In particular, Gentner's systematicity criterion is a novel and valuable insight. It is not mapping of relations alone, but mapping of interconnected structures of relations - that is, maintaining higher-order relations - that tend to make analogies most productive (Gentner et al. 1993). In our case, the computational model led to understanding plasticity (learning) in the *in vitro* "dish" model of cortical neurons in terms of a set of interconnected spatial measures derived from it.

How our scientists arrived at what to map and transfer, though, diverges from a purely syntactic theory of structural alignment and projection, and fits better a "multiconstraint" theory along the lines developed by Keith Holyoak and Paul Thagard (summarized in Holyoak & Thagard, 1996). They have argued that transfer involves evaluating the plausibility of an inference in the context, and thus semantic and pragmatic information help to determine the candidate mappings and what to transfer, and, indeed, they are operative in all of the processes of analogy. The researchers in our case did not make an *a priori* assessment of the source for relational structures that could serve as possible candidate mappings for transfer, as Gentner's structure mapping theory posits. In each phase of building the model, goals and interpretation guided selection throughout the iterative processes of model building, problem solution, mapping, and transfer. In general, salient target constraints, generated by the interpretation and goals of the target problem, guided the selection of the constraints and candidates for transfer within the source domains. These, in turn, provided constraints on model construction. The models served as interpretations, through which the problem solver reasoned, and transferred candidate solutions to the target phenomena.

However, what differentiates our case from current accounts of analogy is this: understanding how the problem of conceptualizing plasticity and learning in the *in vivo* neural network was solved requires factoring in how the model was constructed. That is, solving the problem of best representing the burst activity seen in the computational model, led to the notion of CAT. D11 used constraints drawn from the target, analogical sources, the modeling platform, and the models themselves to construct novel representations that became the objects of reasoning. The dynamical relations resulting from this meld were explored through visualizing and simulating the model in its physical realization, imaginatively, and through expression in other representational formats, including mathematics, language, and real-world experiments. Only once the constructed model was established to be an adequate representation could the inferences that flow from it be transferred to the target problem – in the form of hypotheses to be explored by means of the *in vitro* physical models.

Current research on analogy does not examine this creative work of *building the representations*, which is central in understanding how such scientific analogies function to solve their respective problems. Similarly, the iterative development of the hybrid model to a point where it can act as an analogy is not considered by current theories. So while our exemplar concurs with the current understanding that the core of an analogy lies in a relational comparison, it raises at least two theoretically interesting questions. One, what relations are to be compared and how are they selected - especially in the absence of a ready-to-hand problem solution? Two, how do external analogies – hybrid models built by incorporating a range of target, source, and model constraints – evolve over time to produce novel mappings to the source? Addressing these two questions is central to developing a model of the role of analogies in scientific cognition.

4. Conclusion

Our case study adds to a small literature on analogical problem-solving across time, which differs from the kind examined by current cognitive theories of analogy, most notably in that analogical sources in our case do not provide ready-to-hand problem analogs, from which a solution can be mapped and transferred.[5] The striking feature about the use of analogical sources in these kinds of exemplars is that information from the source domain is not mapped directly to the target problem; rather the source domain provides information in the form of partial constraints, which, together with constraints from the target domain, is used to create intermediary hybrid models. These models are themselves partial and possess their own model constraints, through which the problem solvers think and reason. The constructive and reasoning processes involve analogical, imagistic, and simulative processes. It is only after the problem is solved in the model that it serves as an analogical source from which to map a solution to the original (target) problem. Since the objective is to focus selectively on the relevant pieces, the models work well enough for reasoning purposes, even though they might not fully be feasible as

[5] See Nersessian 2008 for other examples and further elaboration of this point.

real-world entities or processes.

Although our case might be considered extraordinarily creative, our intuition is that if analogy use "in the wild" were to be studied systematically, the construction of such intermediary hybrid representations, making use of visualization and mental simulation, would be seen to be significant dimensions of mundane usage as well. From the viewpoint of such scientific practices, analogy, as customarily understood, stands at one end of a creative continuum that ranges from purely syntactic transformations that can be studied synchronically, to extremely complex instances spread over time and external structures, similar to the case we outline. To understand how analogy functions in creative problem solving in science, cognitive theories need to take into account such model-constructive, imagistic, and simulative processes.

References

Andersen, H., Barker, P., and Chen, X. (2006). *The Cognitive Structure of Scientific Revolutions*. Cambridge, Cambridge University Press.

Clement, J. (1989). Learning via model construction and criticism. In G. Glover, R. Ronning and C. Reynolds (Eds.), *Handbook of Creativity: Assessment, Theory, and Research*. (pp. 341-381). NY: Plenum.

Christensen, B. T. & C. D. Schunn. (2007). The Relationship between Analogical Distance to Analogical Function and Pre-Inventive Structure: The Case of Engineering Design. *Memory and Cognition,* 35, 29-38.

Darden, L. (1991). *Theory Change in Science: Strategies from Mendelian Genetics*. NY: Oxford University.

Dunbar, K. (1995). How Scientists Really Reason: Scientific Reasoning in Real-World Laboratories. In R. J. Sternberg and J. E. Davidson (Eds.), *The Nature of Insight*, (pp. 365-395). Cambridge, MA: MIT Press.

Dunbar, K., & Blanchette, I. (2001). The in vivo/in vitro approach to cognition: The case of analogy. *Trends in Cognitive Science,* 5, 334-339.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.

Gentner, D., S. Brem, R. W. Ferguson, A. B. Markman, B. B. Levidow, P. Wolff, & K. D. Forbus (1997). Analogical reasoning and conceptual change: A case study of Johannes Kepler. *The Journal of the Learning Sciences,* 6(1), 3-40.

Gentner, D., & K. J. Holyoak (1997). Reasoning and Learning by Analogy. *American Psychologist*, 52, 32-66.

Gentner, D., M. J. Ratterman & K. D. Forbus (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology*, 25, 524-575.

Giere, R. N. (1988). *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

Gooding, D. (1990). *Experiment and the Making of Meaning: Human Agency in Scientific Observation and Experiment*. Dordrecht: Kluwer.

Griesemer, J. R. (1991). Material models in biology. *Proceedings of PSA 1990*. East Lansing, MI: PSA. 2: 79-94.

Griesemer, J. R. & W. Wimsatt (1989). Picturing Weismannism: A case study of conceptual evolution. In M. Ruse (Ed.), *What the Philosophy of Biology is*, *Essays for David Hull*, (pp. 75-137). Dordrecht: Kluwer.

Holyoak, K. J. & P. Thagard (1996). *Mental Leaps: Analogy in Creative Thought*. Cambridge, MA, MIT Press.

Nersessian, N. J. (1984). *Faraday to Einstein: Constructing Meaning in Scientific Theories*. Dordrecht: Martinus Nijhoff/Kluwer Academic Publishers.

Nersessian, N. J. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In R. Giere (Ed.), *Minnesota Studies in the Philosophy of Science,* (pp. 3-45). Minneapolis: University of Minnesota Press.

Nersessian, N. J. (2005). Interpreting Scientific and Engineering Practices: Integrating the Cognitive, Social, and Cultural Dimensions. In M. Gorman, R. D. Tweney, D. Gooding and A. Kincannon (Eds.), *Scientific and Technological Thinking*, (pp. 17-56). Hillsdale, N. J.: Lawrence Erlbaum.

Nersessian, N. J. (2008). *Creating Scientific Concepts*. Cambridge, MA, MIT Press.

Nersessian, N. J., E. Kurz-Milcke, W. Newstetter, & J. Davies (2003). Research Laboratories as Evolving Distributed Cognitive Systems. In D. Alterman and D. Kirsh (Eds.) *Proceedings of the Cognitive Science Society 25,* (pp. 857-62). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shelley, C. (1996). Visual abductive reasoning in archeology. *Philosophy of Science,* 63, 278-301.

Spranzi, M. (2004). Galileo and the mountains of the moon: Analogical reasoning, models and metaphors in scientific discovery. *Cognition and Culture*, 4(3), 4451-4485.

Trafton, J. G., S. B. Trickett, & F. E. Mintz. (2005). Connecting Internal and External Representations: Spatial Transformations of Scientific Visualizations. *Foundations of Science*, 10, 89-106.

Thagard, P. (1992). *Conceptual Revolutions*. Princeton: Princeton University Press.

Trickett, S. B. & J. G. Trafton. (2007). "What If....": The Use of Conceptual Simulations in Scientific Reasoning. *Cognitive Science*, 31, 843-76.

Tweney, R. D. (1992). Stopping time: Faraday and the scientific creation of perceptual order. *Physis* 29, 149-164.

Tweney, R. D. (2004) Replication and the Experimental Ethnography of Science. *Cognition and Culture*, 4, 731-58.

Figure Legends

Figure 1: MEAscope display of bursting phenomena

Figure 2: Visual representations of changes in activity across the in silico network. (a) Time 1, (b) Time 2, (c) CAT T1 to T2

Figure 3: Visualization of an *in vitro* CAT (b) and corresponding dish burst activity (a).

Figure 4: Iterative modeling processes

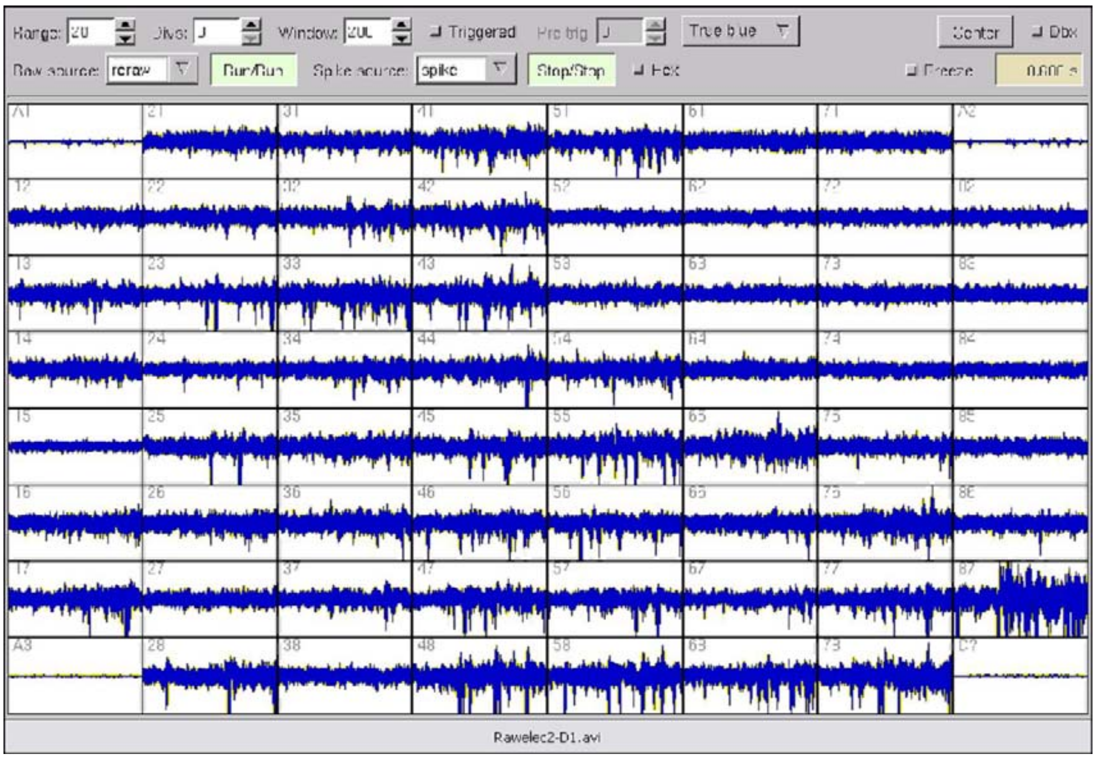
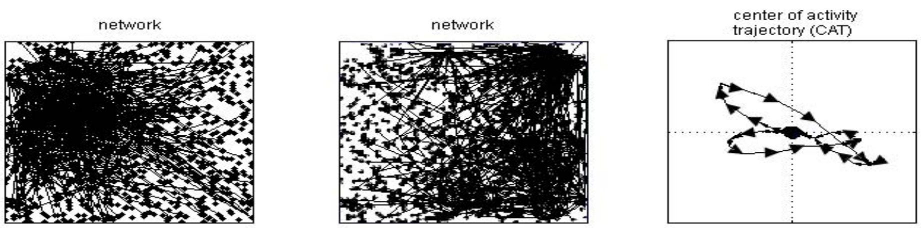
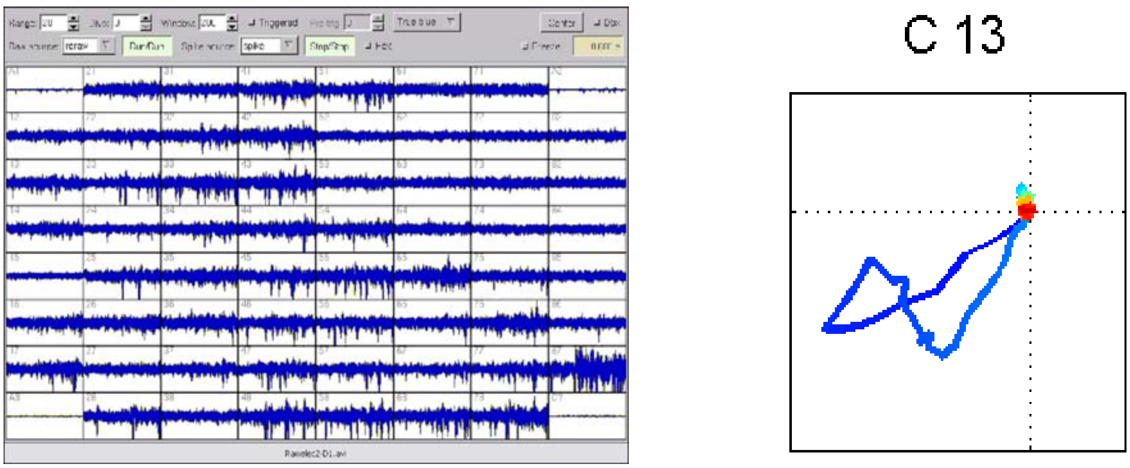

Range: 20
Divs: 0
Window: 200
Triggered
Pre trig
True blue
Center
Dbx
Raw source: rcraw
Run/Run
Spike source: spike
Stop/Stop
Hex
Freeze
Rawelec2-D1.avi

Figure 1



(a) (b) (c)

Figure 2



(a) (b)

Figure 3

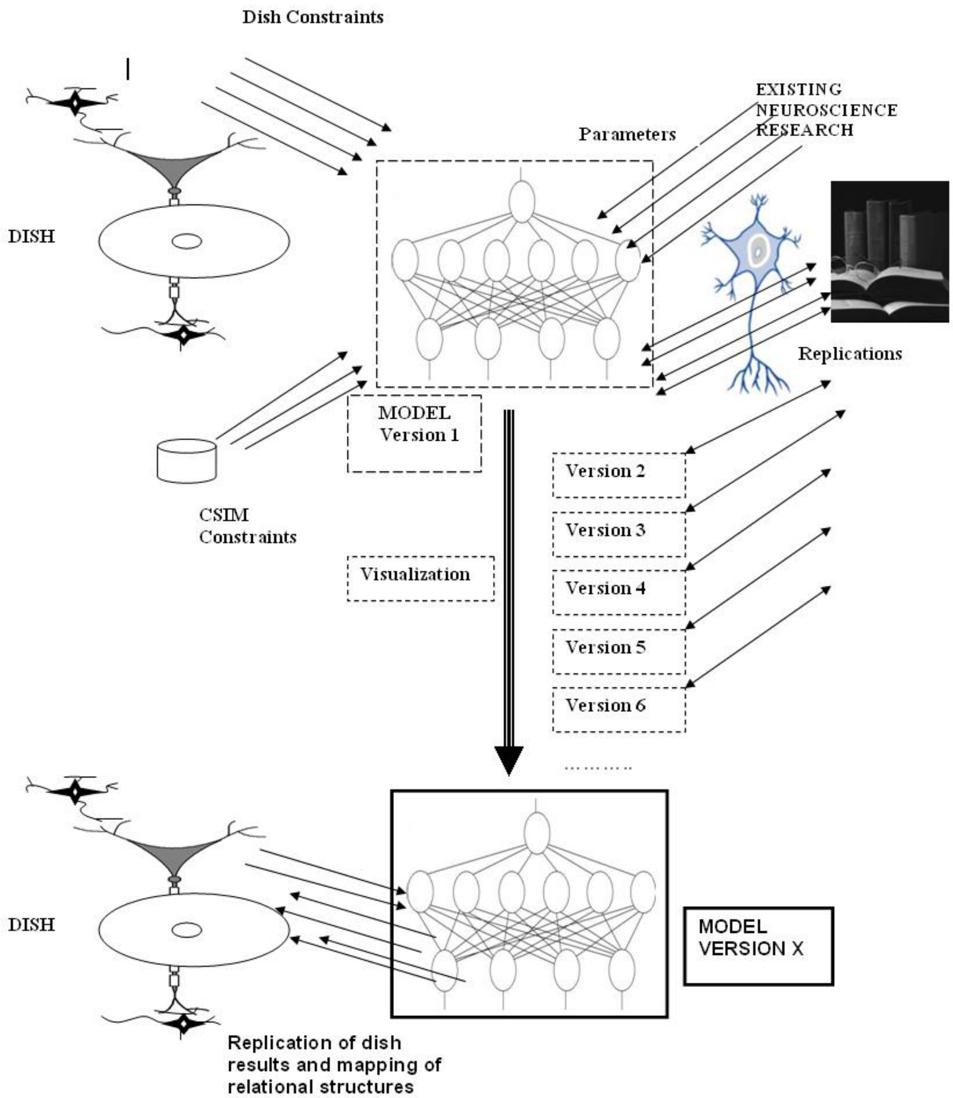

Dish Constraints
EXISTING
NEUROSCIENCE
RESEARCH
Parameters
DISH
Replications
MODEL
Version 1
CSIM
Constraints
Version 2
Version 3
Visualization
Version 4
Version 5
Version 6
DISH
MODEL
VERSION X
Replication of dish
results and mapping of
relational structures

Figure 4