in Scientific Concepts and Investigative Practice, U. Feest & F. Steinle, eds. (Berlin: DeGruyter, 2012). 245-269

Modeling Practices in Conceptual Innovation: An ethnographic study of a neural engineering research laboratory

Nancy J. Nersessian
Georgia Institute of Technology

> A concept is not an isolated, ossified changeless formation, but an active part of the intellectual process, constantly engaged in serving communication, understanding, and problem solving.
>
> Lev Vygotsky
> *Language and Thought*

1. Introduction

Vygotsky's statement captures the idea that mundane concepts are dynamic and socio-cultural in nature. As such, they are neither completely fixed units of representation nor solely mental representations, but arise, develop and live in the interactions among the people that create and use them. This idea is quite compatible with the notion of concepts as participants in the investigative practices of scientists.[1] As much research has demonstrated, concepts do not arise fully formed in the head of a scientist but are created in historical processes, which can extend for considerable periods and even span generations of scientists. As I have argued previously (Nersessian 1984, 2008), novel scientific concepts arise from the interplay of attempts to solve specific problems, use of conceptual, material and analytical resources provided by the problem situation, and often through *model-based reasoning* processes. In such reasoning processes, models are dynamical constructions through which scientists make inferences and

[1] It is also compatible with the view of "concepts in use" articulated by Kindi in contrast with the standard

solve problems that sometimes require conceptual innovation and change. In the conceptual modeling practices I have studied, analogical, visual, and simulative processes are used incrementally to construct models that embody various constraints drawn from the domain under investigation (target), analogical source domain(s), and, importantly, those that arise in the constructed model itself can lead the reasoner towards a problem solution. Nersessian (2008) details how novel concepts can arise from this kind of "bootstrapping process" in which hybrid models that abstract and integrate constraints from both the domain of the target problem and selected analogical source domains are constructed, analyzed, and evaluated incrementally towards the solution of the target problem. One of the most interesting aspects of this process is that in abstracting and integrating constraints from diverse domains (including constraints that arise from the models themselves), here-to-fore unrepresented structures or behaviors can emerge and lead to the formation of novel concepts.

Although it has long been known that analogy plays an important role in creating novel concepts, all the cases I have examined from several data sources – historical, think-aloud protocols, ethnographic studies – point to a significant facet of analogy in the modeling practices of scientists that is neglected in both the philosophical and cognitive science literatures: often in cutting-edge research, there is no ready-to-hand problem solution that can be retrieved and adapted analogically from a source domain. Rather, analogical domains only provide some constraints that are incorporated into models which are constructed in accord with the epistemic goals of the scientist explicitly to serve as analogical sources for the target domain. That is, the constructed model is built explicitly to provide a comparison to the target phenomena based on analogy. Thus, the core of the problem-solving process consists in building models that embody

---

philosophical notion of concepts as solely mental entities or abstract objects (Kindi, this volume).

constraints from the target phenomena and possible analogical source domain(s), solving problems in the constructed models, and then transferring the solution as a hypothesis to the target phenomena. This point will be elaborated in a fascinating case of conceptual innovation that emerged during an ethnographic study my research group was conducting.

Until recently, research into scientific concepts has drawn exclusively from historical records. Now observational and ethnographic studies of concepts in use and development are being added to the mix and can increase significantly our understanding of their role in investigative practices. For several years, I have been conducting ethnographic research, which in part aims to investigate how concepts are used, created, and articulated in research laboratories bio-engineering sciences. These frontier areas are interesting for investigating the role of concepts in practice because the nature of the research requires some measure of interdisciplinary synthesis, and thus such areas are likely to provide a good source for cases of concept transfer and adaptation and, possibly, the formation of novel concepts. Further, during the period of our investigation, practices in these labs did actually demonstrate and confirm the centrality of modeling to conceptual innovation. However, their modeling practices include not only conceptual models, but physical and computational models as well. Thus the ethnographic studies serve to extend historical accounts of conceptual innovation.

Extending the account of model-based reasoning to encompass these kinds of models was the primary reason for venturing into a program of ethnographic research. We know that physical models have been used throughout the history of science (de Chadarevian & Hopwood 2004) and many sciences now make extensive use of computational models. Although historical records might note that such models were developed, the archival records of these artifacts are scant, as

are detailed records of how they were made, the various versions and considerations that went into their making, and understanding of what they afford as embodied practices. As for computational models, records of the processes through which these were developed are even more scant, and consequently most of the philosophical literature is focused on the representational relations between the completed model and the world. Ethnographic research can be a valuable means for investigating model-based research in action. In this paper I will develop a case drawn from ethnographic studies of bio-engineering research labs which can help to understand how concepts are both generated by investigative practices of simulation modeling - physical and computational - and generative of such practices.

2. Methods: Cognitive-historical ethnography

Science studies researchers are most familiar with ethnography as a means of investigating the social and material practices of scientists. Over the last 20 years, researchers in cognitive science have adapted ethnographic methods from anthropology to study cognitive processes, such as reasoning and problem solving, in naturally situated practices (Hutchins, 1995, Hollan et al., Goodwin 1995, Lave, 1988). In line with this research, we conducted "cognitive ethnographies" (Hutchins 1995) of bio-engineering sciences research laboratories. Since 2001, I have been leading an interdisciplinary research group that has been investigating cognitive and learning practices in five research labs in bio-engineering fields: tissue engineering, neural engineering, bio-robotics, and, in on-going data collection, two in integrative systems biology, one that does only computational modeling and one that conducts experiments as well as modeling. Engineering sciences are emerging interdisciplinary fields where basic scientific research is conducted in the context of complex application problems. A major objective of our

research is to develop integrative accounts of problem-solving practices as embedded in social, cultural, and material contexts. Physical or computational modeling are the principal means through which research problems are addressed.

The tissue and neural engineering labs we studied between 2001 and 2008 both construct, primarily, physical models that serve as a means of conceptual and experimental exploration. Issues of control and, often, ethics, make it not possible to experiment on target in vivo phenomena, and so research in these labs is conducted by means of what they call simulation "devices" – in vitro physical models that are designed and built to serve as structural, behavioral, or functional analogs to selected aspects of in vivo phenomena. These devices participate in experimental research in various configurations called "model-systems". As one researcher put it: "I think you would be safe to use that [notion] as the integrated nature, the biological aspect coming together with the engineering aspect, so it's a multifaceted model system." Research is conducted with these in vitro devices and outcomes are transferred as candidate understandings and hypotheses to the in vivo phenomena. That is, a simulation device is designed "to predict what is going to happen in a system [in vivo]. Like people use mathematical models... to predict what is going to happen in a mechanical system? Well, this [model-system she was designing] is an experimental model that predicts – or at least you hope it predicts – what will happen in real life."

Intensive data collection was conducted in each laboratory for 2 years with follow-up of the participants, their research, and questions pertaining to our research for an additional 2 years. Several members of our research group became participant observers of the day-to-day practices in each lab. The ethnographic part of the study (observations and open (unstructured) interviews

sought to uncover the activities, tools, and meaning-making that support research as these are situated in the on-going practices of the community. We took field notes on our observations, audio taped interviews, and video and audio taped research meetings (full transcriptions are completed for 148 interviews and 40 research meetings). As a group we estimate our ethnographers (6) made over 800 hours of field observations. Early observations directed our attention to the custom-built simulation models as "hubs" for interlocking the cognitive and cultural dimensions of research. Because of this, the research meetings, though useful, assumed a lesser importance then they have in other research on cognitive practices in laboratories (see, esp., Dunbar 1995). We needed more to elicit from researchers their understanding and perceived relation to simulation artifacts, and see how they functioned within the life of the labs, which was better addressed through interviewing and extensive field observation.

Significantly, these laboratories are evolving systems, where the custom-built technologies are designed and redesigned in the context of an experiment or from one research project to another. Researchers (who are mostly students) and simulation artifacts have intersecting historical trajectories. To capture this and other historical dimensions of research in these laboratories we used also interpretive methods of cognitive-historical analysis (Nersessian 1992, 1995, 2008). In this investigation, cognitive-historical analysis examines historical records to advance an understanding of the integrated cognitive - social - cultural nature of research practices. Data collection for this part of our study included publications, grant proposals, dissertation proposals, power point presentations, laboratory notebooks, emails, technological artifacts, and interviews on lab history.

3. A case study of conceptual innovation in neural engineering

In this paper I develop a case from the neural engineering "Lab D" in which we collected data over 4 years. It involves both investigative practices of physical and of computational modeling as a means of creating basic scientific understanding and the interaction of these in a rather spectacular case of conceptual innovation. This case involves the research projects of three graduate student researchers that were brought together through the development of a computational model of an in vitro model – what might be considered a 2nd order model – constructed initially to understand what they were calling "burst" phenomena in a physical simulation model – a cultured network of living neurons, locally called "the dish," which is the focal model-system designed by the lab. Novel insights about the in silico (computational) dish were mapped and transferred to the in vitro dish and led to the development of what could prove to be a significant conceptual innovations in neuroscience. This case demonstrates the interactions of concepts and modeling practices that can lead to conceptual innovation through transfer of concepts as well as novel concept formation. Although there is overlap in the research being conducted in the lab in the period of interest, I divide the research into phases for clarity of exposition.

3a. Phase I: "playing with the dish"

Lab D had just begun its existence when we started collecting data. The Lab director (D6) was a new assistant professor who had spent an extended postdoctoral period developing the main technologies he would need to conduct research on living neural networks. He had been interested in computational neural network modeling as an undergraduate and during graduate school in biochemistry he continued "moonlighting as a cognitive scientist" reading, attending conferences, and doing neural network modeling for fun, plus taking courses on the

psychobiology of learning and memory. The current neuroscience paradigm for studying the fundamental properties of living neurons used single-cell recordings. D6 believed that to study learning there needed to be a way to study the properties of networks of neurons, since it is networks that learn in the brain. He recounted to us that somewhere around 1990 (in the middle of graduate school where he was engaged in what he felt was uninteresting research), he had the idea that "perhaps you could make a cell culture system that learns." Such a culture would more closely model learning in the brain, which is a network phenomenon, and also enable emergent properties to arise. Learning requires feedback, so the in vitro system would need to have sensory input or "embodiment" of some kind. Having read the proceedings of a conference about the simulation of adaptive behavior in computational animals or in robots using the computer as the "brain," which were called "animats" by that community, he thought "hey, you could do this in vitro – have an animat that is controlled by neurons and somehow embody it." Lab D was founded (twelve years later) to pursue his insight and the general hypothesis that advances could be made in the overarching problem of understanding the mechanisms of learning in the brain by investigating the network properties of living neurons. Figuring out the control structure for supervised learning in the living network, that is, how the dish could be trained to control the embodiment systematically using feedback, posed a significant and multifaceted problem, the solution to which would involve conceptual innovation.

The case developed here is sketched in Figure 1, which indicates what each of the three researchers was doing during the 3-year period of interest. At the time we were collecting data, we had no foreknowledge that we would be capturing what could prove to be highly significant conceptual innovations for the field. Fortunately, we had collected sufficient and relevant data as

the process was unfolding and then conducted follow-up interviews with the Lab D members as their major publications and dissertations on this research were being written. The three graduate students involved in the case were all recruited within a few months of one another. Much of the first year was directed towards constructing the dish model-systems and developing technology and software to interface with the neuron culture and record dish activity – what researchers called "talking to the dish." Building the in vitro dish model involves extracting cortical neurons from embryonic rats, dissociating their connections, and plating them (15-60K, depending on the desired kind of network) on a specially designed set of 64 electrodes called a multi-electrode array (MEA) where the neurons regenerate connections to become a network. The design of the dish model-system incorporates constraints from neurobiology, chemistry, and electrical engineering. Given the technical challenges of creating the dish, keeping it alive for extended periods (as long as 2 years), and controlling it, the group decided to start with a simple model of a single layer of dissociated cortical neurons.

[insert figure 1 here]

The dish model-system was constructed to provide a means of exploring whether learning could be induced in system of neurons with just network properties of the brain, abstracted from other brain structures. What the dish models was a subject of on-going discussion among the lab members. Some maintained that the dish is "a model of cortical neurons" while others claimed it "is a model of development [of the brain]," and, when pressed, some retreated to "it may just be a model of itself." However, all agreed with D6's belief that studying the dish will yield

understanding of the basic workings of network-level cortical neurons:

> First of all, it is a simplified model, I say that because the model is not – it's artificial, it's not how it is in the brain. But I think that the model would answer some basic questions, because the way the neurons interact should be the same whether it's inside or outside the brain.... I think the same rules apply.

The dish model-system is designed to provide a basic understanding of how neurons communicate and process information such that learning takes place. The intention of the director continues to be that after developing this understanding, the lab will move on to building an investigating more complex models such as "studying cultures with different brain parts mixed together, or specific 3-dimensional pieces that are put together." What is of interest for the case I develop is that in this early period, the dish was an object of interest in its own right.

When this research started, there were no established models of neuron communication. This pioneering research began with importing some concepts from single neuron studies to start to develop an understanding of the dish and work towards the goal of getting it to learn. As an indicator of learning, they assumed the standard psychology concept, which the director (D6) stated is "a lasting change in behavior resulting from experience." Various interfaces were developed to be able to record and stimulate (provide experiences to) the dish, including a suite of software programs they called MEAbench and two forms of embodiment: computationally simulated "animat" worlds (extending the computational concept of animat to include simulation worlds in which the behavior of the computational creature is determined by being connected to

living neural networks), and robotic devices ("hybrots," which stands for "hybrid robots") that could be connected to the dish. Both embodiments enabled closed-loop feedback experiments. They operationalized learning in terms of what is known in neuroscience as the "Hebbian" notion of learning as *plasticity* (basically, changing the brain by adding or removing neural connections or adding new cells in response to experience) and the mathematical formulation known as the *Hebbian rule* (basically, "neurons that fire together wire together") as a guide. As D4 recounted later,

> from Hebb's postulate – which talks about learning between two neurons – we thought our data will show that something has to be added to the known equation in order for it to manifest in a population of neurons. Our idea was to figure out that manifestation.... So it has gone from what Hebb said for two neurons to what would translate into a network.

For recording and displaying the dish activity, they transferred and modified the notion of a *spike* from single cell electrophysiology. There it refers to the electrical trace left behind when an individual neuron fires and it part of a well-established understanding that in neural firing there is a steep jump in voltage potential as the neuron de-polarizes and a proportional drop in potential as it recovers. In the multicelluar case of the dish, the researchers estimate that the electrical activity on a single electrode comes from an average of 3 to 5 neurons, so what records as a spike can come from several neurons firing together and it is impossible to differentiate among the firing of one or of many. The group has developed a software model of a spike that identifies MEA dish spikes according to specified criteria and keeps a record of the electrical activity immediately around the spike. Open-loop electrophysiology research begins with these

filtered spike data, which are represented visually in the format of an oscilloscope by the MEAscope program as output per-channel as in Figure 2.

[insert figure 2 here]

We are now ready to focus on the two borrowed concepts at the center of the significant developments in their research, the notion of *burst*, transferred from single neuron studies, and the engineering notion of *noise*. D4 (electrical engineering), who was the first of the graduate students, helped to construct protocols for the dish model-system before moving to open-loop electrophysiology research. D11 (life sciences and chemistry) entered in a few months later and worked on the spike sorting software and then moved to closed-loop research on animats and hybrots. D2 (mechanical engineering and cognitive science) started about a month later and worked on developing some of the MEAbench software and then started closed-loop research on animats and a specific hybrot: a robotic drawing arm. At the start of our case, they were all involved in "playing with the dish," which is their term for exploring the problem space by stimulating the dish with various electrical signals and tracking the outputs. D4 then began trying to replicate a plasticity result reported by another research group, but was unable to do so largely because the dish was exhibiting spontaneous synchronous network-wide electrical activity - she called this phenomenon "bursting," extending the meaning from single neuron studies where it means the spontaneous activity of one neuron. This dish-wide phenomenon was visualized in Figure 2 as the spike activity for each electrode per recording channel, across all channels. She first attempted to introduce the term "barrage" into the community to focus attention on the

network-wide nature of the phenomenon, but soon reverted to “burst” when her term did not catch on.

Bursting created a problem because it prevented the detection of any systematic change that might rise due to controlled stimulation; that is, it prevented detection of learning. The whole group believed that “bursts are bad” and thought of them in terms of the engineering concept of noise – as a disturbance in the signal that needs to be eliminated, “it’s noise in the data – noise, interference....so it’s clouding the effects of learning that we want to induce.” The group hypothesized that the cause of bursting was lack of sensory stimulation – “deafferentation” – and D4 decided to develop different patterns of electrical stimulation to see if she could “quiet” the bursts. After about a year, she managed to quiet the bursts entirely and initiated plasticity experiments, but for nearly a year she was unable to make any progress. A new problem arose with the quieted dish: activity patterns provoked by a constant stimulation did not stay constant across trials, but “drifted” to another pattern.

During the same period, D2 was trying to use various probe responses to produce learning through closed-loop embodiments. He also spent considerable time traveling the world with an art exhibit featuring the mechanical drawing arm, controlled via satellite by a dish living in Lab D. As a research project, he was trying to get the dish to learn to control the arm systematically, but as a mechanical art exhibit, its creativity required only that it draw, not that it draw within the lines! Early in the burst-quieting period, D11 decided to build a computational model of the dish model-system and physically moved out of the physical lab space to a cubicle with a computer. He felt the “dish is opaque” and what was needed to make progress was more control and measurement capabilities: “I feel that [computational] modeling can tell us something because

the advantage of modeling is that you can measure everything, every detail of the network." The Lab director doubted the computational modeling would lead to anything interesting, but gave his consent to work on it.

3b. Phase 2: computationally modeling the dish model-system

Similar to the kinds of bootstrapping processes detailed in my earlier research, this 2nd order model - the computational simulation of a generic dish model was developed and optimized through a bootstrapping process (see Figure 3) comprising many cycles of abstraction, construction, evaluation, and adaptation that included integrating constraints from the target (their dish model) and analogical sources domains (a wide range of neuroscience literature), as well as constraints of the computational model itself (the modeling platform CSIM and those that arose as the model gained in complexity). This computational dish model was built to serve as an analogical source for the physical dish model-system. That is, D11 hoped that insights derived from it could eventually be mapped and transferred over to the target problem: creating supervised learning in the dish model system. As it turned, out, the computational model proved a source for both novel concepts for understanding dish phenomena and a control structure for supervised learning that were successfully transferred to the physical dish, solving the original problem.

[insert figure 3 here]

I highlight only the most significant constraints. As a modeling platform D11 chose what

he called “the simplest neuron model out there – leaky-integrate-fire,” (CSIM modeling platform) to see if they could replicate network phenomena without going into too much detail, such as including synaptic models. The only constraints taken from their in vitro dish were structural: 8X8 grid, 60 electrodes, and random location of neurons (“I don’t know whether this is true though, looking under the microscope they look pretty random locations.”). In the model he only used 1K neurons, which he believed would produce sufficiently complex behaviors. To connect the neurons he took parameters from a paper about the distribution of neurons. All the parameters of the model – such as types of synapses, synaptic connections, synaptic connection distance, percentage of excitatory and inhibitory neurons, conduction velocity and delay, noise levels, action potential effects – were taken from the neuroscience literature on single neuron studies, brain slices, and other in vitro dish experiments. He then just let it run for a while to see what would emerge. For validating the model he first followed the same experimental protocol used with other in vitro dishes to see if he could replicate those data, and used the data (including the burst data) from their own dish only after he had succeeded with the literature replications dishes (a outcome he called “striking” given the simplicity of the model). By early year 3, he had developed the model network sufficiently to begin “playing with the [computational] dish” (seeing how the computational network behaves under different conditions) and had started getting what he called “some feeling about what happens actually in the [simulated] network.” Sometime in during this period, he moved back into the physical space of Lab D and all three researchers began working together.

Part of getting a feeling for the model involved developing a visualization of the dish activity that proved to be highly significant in solving the supervised learning problem by means

of articulating a cluster of conceptual innovations. As he noted, "I am sort of like a visual guy – I really need to look at the figure to see what is going on." It is important to realize that computational visualizations are largely arbitrary; he could have visualized the simulated dish in any number of ways, including the one the group was accustomed to: a per-channel spike representation from MEAscope (Figure 2). However, he imagined the dish as a network and visualized it that way: "I can visualize these 50K synapses and so you can see – after you deliver a certain stimulation you can see those distributions of synaptic weight change – or synaptic state change."

[insert figure 4 here]

This network visualization (Figure 4) is possible because the in silico dish affords control and visualization at the level of individual neurons, whereas the in vitro dish affords control and visualization only at the electrode level (clusters of neurons). D11 made movies of the dish visualization as it ran (that he showed the others and us, so we too could "come away with the same thing"), which showed the movement of activity patterns across the network over time. He began to notice something interesting: there were structurally similar looking bursts and there seemed to be only a small number of "patterns of propagation" of these. This led him to conclude "you get some feeling about what happens in the network – and what I feel is that... the spontaneous activity or spontaneous bursts are very stable." The next step was to attempt to develop a means of tracking the activity of the possibly "stable" bursts across the network.

3c. Phase III: controlling the in vitro dish model-system

From this point, things developed rapidly in the lab as the group worked together on statistical analyses, experimentation to see whether the "drift immune" measures developed for the computational network could be transferred to the in vitro dish, and whether the "burst feedback" in the in vitro dish could be used for supervised learning with the embodiments. This phase of research began with the idea that "bursts don't seem as evil as they once did" (D4). Most importantly, they began to develop the concept of bursts as signals (rather than only noise) that might be used to control the embodiments. Articulating the notion that bursts can be signals took the form of several interconnected novel concepts: *burst type*: one of limited number of burst patterns (10); *burst occurrence*: when a type appears; *spatial extent*: an estimation of burst size and specific channel location; and *CAT* ('center of activity trajectory'): a vector capturing the flow of activity at the population scale. With the exception of 'spatial extent' all of these concepts were developed for the simulated network first and then mapped to the in vitro dish and modified as required. Although each of these concepts is important, they are quite complex conceptually and mathematically, and so I will only provide some details of the development of 'CAT', which is an entirely novel concept for understanding neural activity and could prove to be of major importance to neuroscience.

D2 recounted during the final stages of analysis:

> ...the whole reason we began looking at the center of activity and the center of activity trajectory is because we are completely overwhelmed by all this data being recorded on the 60 electrodes – and we just can't comprehend it all. The big motivation to develop this is to actually have something – a visualization we can understand.

The mathematical representation of the CAT concept was articulated by making an analogy to the physics notion of center of mass and by drawing from three resources within the group: 1) D11's deep knowledge of statistical analyses from the earlier period in which he had tried to create sensory-motor mappings between the dish and the embodiments; 2) an earlier idea of another graduate student at D6's old institution (who had worked remotely with the group) that it might be possible to capture "the overall activity shift" in the in vivo dish by dividing the MEA grid into four quadrants and using some subtraction method; and 3) the idea that bursts seem to be initiated at specific sites as shown in a new graphical representation for the in vivo dish (spatial extent of a burst) developed by D4 after the computational model had replicated her in vitro dish results. D4 had been trying to see if she could get at some of the information the computational visualization was providing by graphing more specific information about bursts, in particular, their location and frequency over time. This became what they conceptualized as 'spatial extent': "the number of times any neuron near an electrode near an electrode initiated a burst in 30 minute segments of a 1.5 hour spontaneous recording." Spatial extent of bursts is represented by the color and size of the circles in Figure 5, which clearly represents different information than the MEAscope representation of bursts as spike per channel across the channels (Figure 2). However, it does not represent activity as it *propagates* across the network, which is what CAT does.

[insert figure 5 here]

The mathematical representation of CAT includes a temporal as well as spatial

dimension: “I not only care about how the channel’s involved in the burst, I also care about the spatial information in there and the temporal information in there – *how they propagate*” (D11). CAT tracks the spatial properties of activity as it moves through the network; that is, *the flow of activity at the population scale*, as displayed in the visualization screen shot (Figure 6c). It is an averaging notion similar to the notion of population vectors, which capture how the firing rates of a group of neurons that are only broadly tuned to a stimulus, when taken together, provide an accurate representation of the action/stimulus. CAT differs from a population vector and is more complex because it tracks the spatial properties of activity as it moves through the network. That is, if the network is firing homogeneously or is quiet, the CAT will stay at the center of the dish, but if the network fires mainly in a corner, the CAT will move in that direction (see Figure 6, 6c is the CAT representation). Thus, CAT tracks *the flow of activity* (not just activity) at the population scale, and much faster than population vectors. It is a novel conceptualization of neuronal activity. What the CAT analysis shows is that in letting the simulation run for a long time, only a limited number of burst types (classified by shape, size, and propagation pattern) occurs – approximately 10. Further, if a probe stimulus is given in the same channel, “the patterns are pretty similar.” Thus the CAT *provides a “signature” for burst types*.

[insert figure 6 here]

D11 was unsure whether it would be possible to transfer the concept to the in vitro dish because his “feeling” for what the simulated dish was doing, “but the problem is that I don’t think it is exactly the same as in the living network – when our experiment worked in the living

network I am surprised – I was surprised." But the group decided to try a mapping by replacing individual neurons by individual electrodes where there are clusters of neurons for in vitro CAT and began a range of experiments with the in vitro dish (open loop), the dish connected to an animat version of a robotic drawing arm, and then the dish connected to the real robotic drawing arm. D4 summed up the difference between CAT and the way they had been thinking of the in vitro dish prior to D11's simulation: "he [D11] describes a trajectory... we weren't thinking of vectors with direction... so think of it as a wave of activity that proceeds through the network. *So he was thinking like a wave, while we were thinking of a pattern*." The spatial extent analysis tracks a pattern of bursting across the channels (Figure 5), but the CAT analysis tracks a wave of busting activity across the network (Figure 6). Notably, she continued, "*we had the information always... the information was always there*." What CAT enabled them to do is to tap into the *behavior* of the system and, eventually, to control its learning.

D4 kept working with open loop experiments, did some preliminary research with someone in a medical school that transferred her new understanding of bursts as signals (derived from the computational and in vitro models) to epilepsy, and graduated. D2 and D11 stayed for an additional year and combined CAT and techniques D4 had developed for burst quieting to develop a range of stimulation patterns for the in vitro dish that led to supervised learning for the embodied dish. They got the mechanical arm to draw within the lines, and wrote and successfully defended dissertations on different aspects of this work. Interestingly, the control structure is unlike the customary structure for reinforcement learning, where the same stimulation is continually repeated. Their control structure consists of providing the network with a patterned stimulation, inducing plasticity, followed by providing a random background stimulation to

stabilize its response to the patterned stimulation. Again, this method is counterintuitive to existing notions of reinforcement learning, and emerged only in the context of the group's building and playing with the two different kinds of simulation models.

4. Discussion

There are several features of this case that are important for thinking about concepts in investigative practices. As Steinle (this volume) reminds us, there are a wide range of epistemic aims that participate in the dynamics of concept formation in science (see also Brigandt, this volume). In frontier biomedical engineering sciences labs, chief among the aims is developing an understanding of novel in vivo phenomena sufficient to enable some degree of intervention and control. The primary investigative practice in many areas is constructing physical models that adequately exemplify the phenomena of interest so as to be able to conduct controlled experiments with the models and transfer outcomes to in vivo phenomena. The case examined here also included developing a computational simulation of the physical model. Such physical and computational models are built towards becoming analogical sources/bases. From the outset, the intention is to build an analogy but the nature of that analogy is determined incrementally, over time, with only certain features of it selected at the time of building. Often to build an analogy requires configurations of more than one model, comprising both engineered artifacts and living matter. These "model-systems" are dynamical entities that perform as structural, functional, or behavioral analogs of selected features of the in vivo systems. Through experimenting with them, researchers develop hypotheses that they "hope [will] predict.... what will happen in real life." Such physical and computational simulation model-systems are *enactive systems* with emergent behavioral possibilities and the need to represent novel behaviors can

promote conceptual innovation (Chandrasekharan et al. in press; Chandrasekharan & Nersessian 2011).

Since these areas are conducting ground-breaking, frontier research, there is little understanding of the phenomena. So, initial ways of thinking about processes taking place in the in vitro model-systems are often provided by transferring concepts from what are thought to be related domains. In the case of Lab D, we saw, for instance, that several concepts were transferred from the well-developed area of single neuron studies to provide an initial understanding of "the dish" - a living neural network. The physical dish model is a hybrid construction, merging constraints, methods, and materials from biology, neuroscience, and engineering. The initial understanding of phenomena exhibited by the in vitro model was in terms of concepts borrowed from single neuron studies (spike, burst) and engineering (noise). They understood that the emergent properties of the network might require modification of these concepts. In practice, these transferred concepts both facilitated and impeded the research. The concept of spike facilitated developing stimulation and recording methods and interpretations of the output of clusters of neurons surrounding an electrode. The concept of burst, when extended to spontaneous dish-wide electrical activity, and categorized as noise (in the engineering sense), and thus something to be "quieted," impeded the research for an extended period.

To deal with the impasse D11 introduced a new modeling method into the lab research, computational simulation of a physical model, which lead to the formation of a cluster of novel concepts ('CAT', 'spatial extent', 'burst type', 'burst occurrence') which together enabled them to understand that bursts could be signals (as well as noise). This second-order model was constructed to eventually provide an analogy to the physical model; that is, once it had

sufficiently replicated in vitro dish behavior, inferences made about the phenomena taking place in it were to be transferred to the in vitro model, and potentially from there to the in vivo phenomena. The computational model merges modeling constraints, intra-domain constraints from other areas of neuroscience (brain slices, single neuron studies), and dish constraints. Constructing the simulation model facilitated D11's thinking of the dish at the level of individual neurons and in terms of populations of neurons and how these interact dynamically to produce behavior. The network visualization reinforced this idea and provided a dynamical simulation (captured in movies that could be examined more carefully) of the real-time propagation of the activity across the network that allowed D11 to literally see that there were similar looking burst patterns – and these seemed to be limited in number. Further, he could show these to the others who could also see these phenomena. It enabled them, as they said, "to look inside the dish." The visualization could have taken numerous forms. However, the network visualization and simulation exemplify (Goodman 1968; Elgin 2009) the network features of the phenomena, whereas the per-channel representation of the MEAscope graph does not. This enabled the group to develop a better grasp of the *behavior* of the network. The spatial extent graphs only capture structural information, whereas the CAT captures behavioral information as it unfolds over time. This notion could not have been formalized without the interaction of all members with the simulated dish. The information might have been "always there," but the computational simulation and visualizations make it accessible. These physical and computational simulation models are *enactive systems* with emergent behavioral possibilities that can lead to novel insights, which can promote conceptual innovation (Chandrasekharan et al. in press; Chandrasekharan & Nersessian 2011).

This interaction underscores that explaining how the conceptual innovations arise requires an analysis of the interacting components within an evolving system of researchers, artifact models, and practices. Physical and computational simulation models embody researchers' current understandings and suppositions and thus serve as simulations of these as well as of the phenomena they are constructed to model. The case provides an exemplar of what cognitive scientists would call *distributed cognition*. Concepts and models exist not only as representations "in the head" of the researchers (mental models), but also as representations in the form of physical and computational simulation models and in inferential processes there is a coupling of human and artifact representations (Nersessian 2009). Further, the manifest nature of the in silico dish network through its visualization enabled the group to exploit it communally. In particular, it facilitated making joint inferences about what might map to the in vitro dish, rejecting false leads, developing extensions, and coming to consensus. Once they had the new concepts, they could think about a range of new investigations, such as how to control the embodied dish model-systems (hybrot and animat) and how to control epilepsy in patients.

In concluding, what do we learn about modeling practices around concepts by studying on-going research beyond what historical analysis can provide? For one thing, the investigations of on-going research establish that model-based reasoning has been contributing to conceptual innovation and change across a wide range of sciences and historical periods and on into present-day science. Of course, the specific kinds of modeling possibilities have enlarged over the history of science bringing with them new affordances, for instance, those of dynamical simulation and visualization of the sort afforded by computational modeling. Still, as I noted at the outset, the model-based bootstrapping to conceptual innovation in this and other cases from our

ethnographic studies exhibit same kinds of processes in the abstract as in the historical cases:

- analogical domains: sources of constraints for building models
- imagistic representation: facilitate perceptual inference and simulation
- simulation: inference to new states via model manipulation
- cycles of: construction simulation/manipulation, evaluation, adaptation
- emergent analogical relation between the model and the target

I did not go into the ethnographic research with the intent to apply the analysis of modeling derived from cognitive-historical research. However, the features that emerged of physical and computational modeling processes were parallel to my earlier analysis of model-based reasoning in the respects noted above. To use a notion drawn from ethnographic analysis, this kind of conceptual innovation process *transfers* robustly across different time periods and also across several sources of data and methods of analysis. Thus, the science-in-action studies also lend support to the interpretations of the less rich historical records. Further, as has been established in historical cases, the ethnographic cases of conceptual innovation and articulation underscore that model-based reasoning (in general) is closely connected with analogical reasoning. This has implications for both the analogy and models literatures. The built models are designed to share certain relations with the in vivo phenomena so it does not matter that in many respects they are "false models," which the philosophical literature has puzzled about. What matters is if these relations are of the same kind as those they are meant to exemplify. For instance, the per-channel visualization does not capture the network features possessed by the in vitro dish and by the in vivo phenomena and the spatial extent visualization captures only a pattern or structure, but not behavior. The network visualization captures the network structure

and behavior of both dish models.

Perhaps most importantly, though, collecting field observations and interviews surrounding cognitive-social-cultural practices during research processes provides a wealth of insight into creative processes and helps those who study science fathom aspects of such practices that would never make it into the historical records. Most prominent are the evolving dynamics of the interactions among the members of research group, between them and the modeling artifacts, and the evolution of those artifacts. Even for physical records there are many pieces that are unlikely to be archived, since researchers rarely keep detailed records of process. The computational visualization that enabled them (and me) literally to see the burst patterns provides an example: a sentence in a publication remarking that "burst patterns were noted" does not convey its cognitive impact or the change it sparked in group dynamics which led to integration across the three research projects. My research group was, of course, not able to make all the observations and collect all the records that are pertinent to these conceptual innovations since ethnographic data collection is complex and time consuming, and of necessity selective. Once it because apparent that major scientific developments were coming out of the Lab D research (nearly 2 years into our research), however, we did have sufficient data to mine to and could go back and collect additional data to develop the most salient aspects of the innovation processes, which are analyzed in outline form in this paper.

Acknowledgments

I thank the members of the Cognition and Learning in Interdisciplinary Cultures research group (www.clic.gatech.edu) who worked on this project, Wendy Newstetter (co-PI), Lisa Osbeck, Ellie Harmon, Chris Patton, and Sanjay Chandrasekharan. The Lab D case analysis, specifically, was developed together with Chris and Sanjay. I thank also the Lab D researchers for being generous with their time in letting us observe and interview them about their research over an

extended period. The research was conducted in accordance with Institute Review Board criteria on human subjects research and the identities of the researchers cannot be revealed. This research was made possible with the support of the US National Science Foundation grants DRL0411825 and DRL097394084.

References

Brigandt, I. (this volume), "The dynamics of scientific concepts: The relevance of scientific aims and values."

deChadarevian, S./Hopwood, N. (Eds.) (2004), *Models: The Third Dimension of Science*. Stanford: CA: Stanford University Press.

Chandrasekharan, S./Nersessian, N. J./Subramaninan, V. (in press), "Computational modeling: Is this the end of thought experiments in science." In J. Brown, M. Frappier & L. Meynell (Eds.), *Thought Experiments in Philosophy, Science and the Arts*. London: Routledge.

Chandrasekharan, S./Nersessian, N. J. (2011), "Building cognition: The construction of external representations for discovery." In L. Carlson, C. Hölscher & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp 267-272).. Austin, TX: Cognitive Science Society.

Dunbar, K. (1995), "How scientists really reason: Scientific reasoning in real-world laboratories." In R. J. Sternberg & J. E. Davidson (Eds.), *The Nature of Insight* (pp. 365-395). Cambridge, MA: MIT Press.

Elgin, C. Z. (2009), "Exemplification, idealization, and understanding." In M. Suárez (Ed.), *Fictions in Science: Essays on Idealization and Modeling* (pp. 77-90). London: Routledge.

Goodman, N. (1968), *Languages of Art*. Indianapolis: Hackett.

Goodwin, C. (1995), "Seeing in depth." *Social Studies of Science, 25*, 237-274.

Hollan, J./Hutchins, E./Kirsch, D. (2000), "Distributed cognition: Toward a new foundation for human-computer interaction research." *ACM Transactions on Computer-Human Interaction, 7*(2), 174-196.

Hutchins, E. (1995), *Cognition in the Wild*. Cambridge, MA: MIT Press.

Kindi, Vasso (this volume), "Concept as vessel and concept as use."

Lave, J. (1988), *Cognition in Practice: Mind, Mathematics, and Culture in Everyday Life*. New York: Cambridge University Press.

Nersessian, N. J. (1984), *Faraday to Einstein: Creating Meaning in Scientific Theories*. Dordrecht: 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. (1995). "Opening the black box: Cognitive science and the history of science." *Osiris, 10* (Constructing Knowledge in the History of Science, A. Thackray, ed.), 194-211.

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

Nersessian, N. J. (2009), "How do engineering scientists think? Model-based simulation in biomedical engineering research laboratories." *Topics in Cognitive Science, 1*, 730-757.

Steinle, F. (this volume), "Goals and fates of concepts: The case of magnetic poles."

Figure Captions

Figure 1

Our representation of the approximate time line of the research leading to the conceptual innovations and development of the control structure for supervised learning with the robotic and computational "embodiments" of the in vitro dish (years 2-4 of the existence of Lab D). The dashed line represents the period after D11 moved back into the main part of physical space of the lab and all three researchers began to actively collaborate on exploiting the findings stemming from the in silico dish.

Figure 2

The MEAscope per channel visualization of in vivo dish activity showing spontaneous bursting across the channels of the dish. Spontaneous bursting activity is represented by the spikes appearing in the channels. A relatively "quiet" dish would have no spikes in the channels, with all channels looking closer to channel 15.

Figure 3

Our representation of the bootstrapping processes involved in constructing, evaluating, and adapting the computational dish through numerous iterations. Once the in silico dish was able to replicate the in vitro dish behavior and the novel concepts were developed for it, the analysis was mapped (adapted to the specifics of its design) and transferred to the in vitro dish, and evaluated

for it.

Figure 4

A screen shot of the network visualization of bursting in the in silico dish.

Figure 5

Spatial extent captures the location and frequency of bursts over time in the in vivo dish per channel. The figures show the burst initiation sites and the number of times (count) any neuron near an electrode initiated a burst in 30 minute segments of a 1.5 hour of spontaneous recording. The color and size of the circles represents the number of times any electrode initiated a burst in the 30 minute of recording. The per channel representation of the MEAscope visualization (Figure 2) is kept, but much different information is analyzed and displayed.

Figure 6

The two screen shots of the computational visualization of the network show the flow of burst activity in simulated dish at (a) burst time 1, (b) burst time 2 and (c) shows a corresponding CAT from T1 to T2. The CAT tracks the spatial properties of activity of the population of neurons as the activity moves across the network. That is, if the network is firing homogeneously or is quiet, the CAT will stay at the center of the dish (on analogy with a center of mass), but if the network fires from one direction to another as in (a) to (b), the CAT will move in that direction.

Figure 1

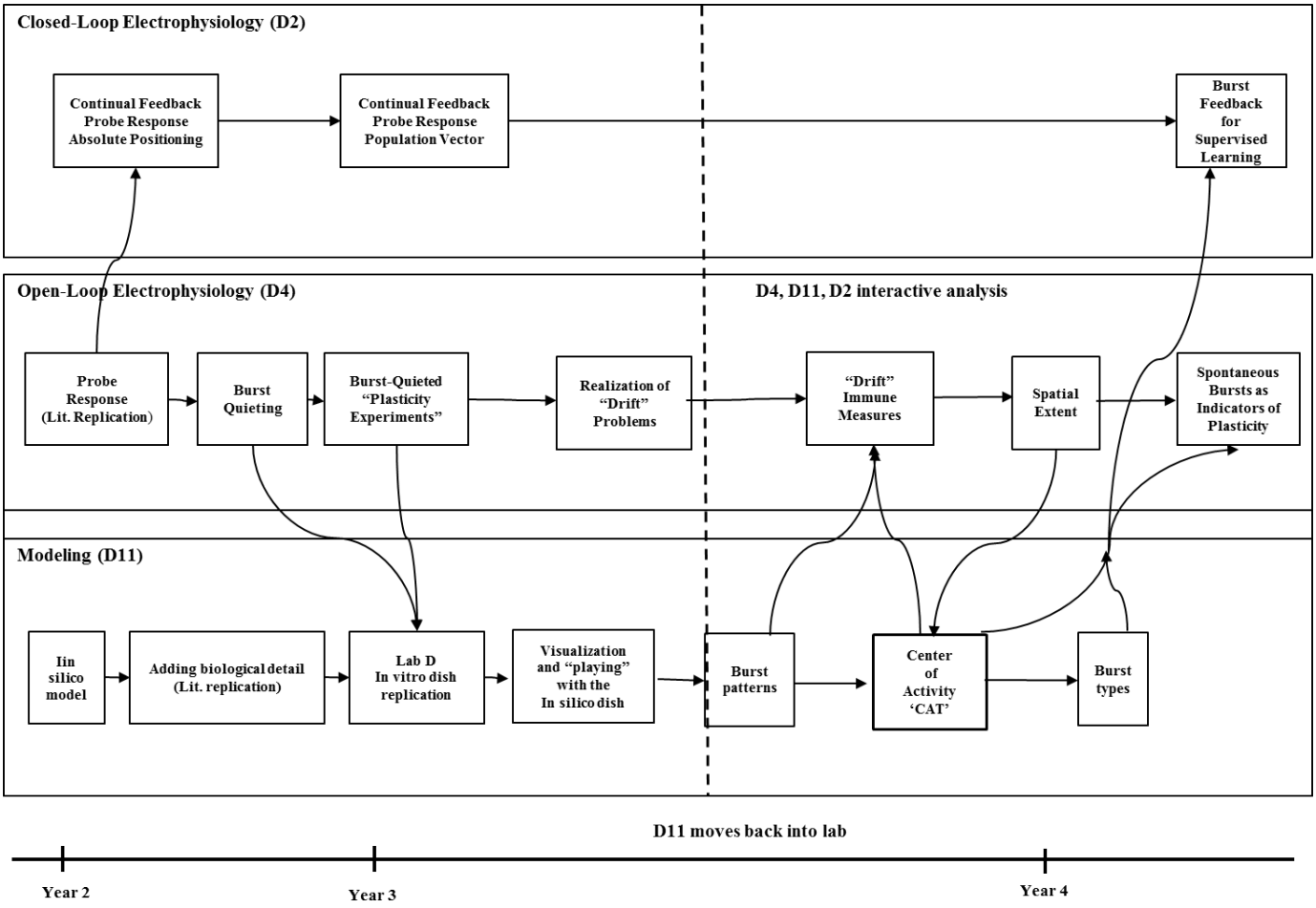

Closed-Loop Electrophysiology (D2)
Continual Feedback Probe Response Absolute Positioning
Continual Feedback Probe Response Population Vector
Burst Feedback for Supervised Learning
Open-Loop Electrophysiology (D4)
D4, D11, D2 interactive analysis
Probe Response (Lit. Replication)
Burst Quieting
Burst-Quieted "Plasticity Experiments"
Realization of "Drift" Problems
"Drift" Immune Measures
Spatial Extent
Spontaneous Bursts as Indicators of Plasticity
Modeling (D11)
Iin silico model
Adding biological detail (Lit. replication)
Lab D In vitro dish replication
Visualization and "playing" with the In silico dish
Burst patterns
Center of Activity 'CAT'
Burst types
D11 moves back into lab
Year 2
Year 3
Year 4

Figure 2

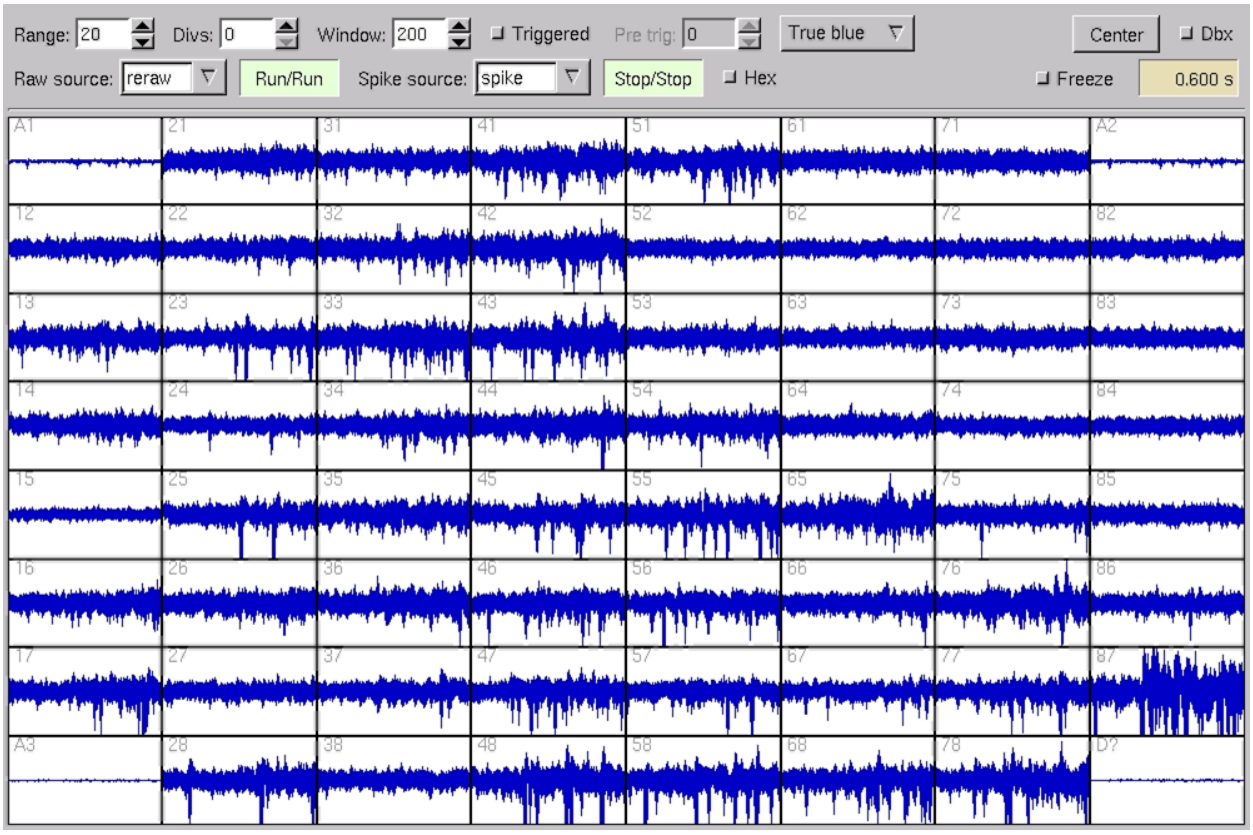

Range: 20
Divs: 0
Window: 200
Triggered
Pre trig: 0
True blue
Center
Dbx
Raw source: reraw
Run/Run
Spike source: spike
Stop/Stop
Hex
Freeze
0.600 s
A1
21
31
41
51
61
71
A2
12
22
32
42
52
62
72
82
13
23
33
43
53
63
73
83
14
24
34
44
54
64
74
84
15
25
35
45
55
65
75
85
16
26
36
46
56
66
76
86
17
27
37
47
57
67
77
87
A3
28
38
48
58
68
78
D?

Figure 3

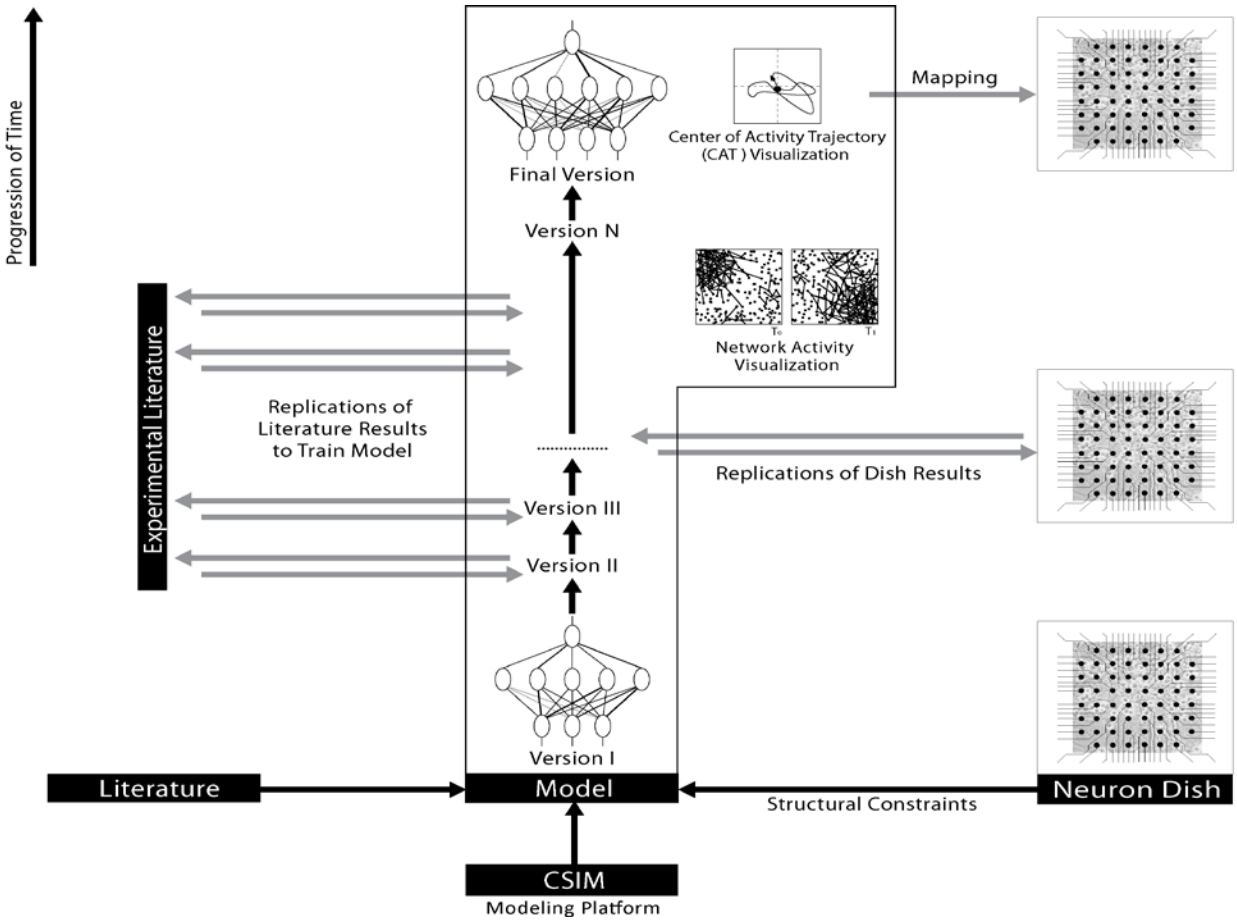

Progression of Time
Final Version
Version N
Center of Activity Trajectory
(CAT) Visualization
Mapping
Network Activity
Visualization
Experimental Literature
Replications of
Literature Results
to Train Model
Replications of Dish Results
Version III
Version II
Version I
Literature
Model
Neuron Dish
Structural Constraints
CSIM
Modeling Platform

Figure 4

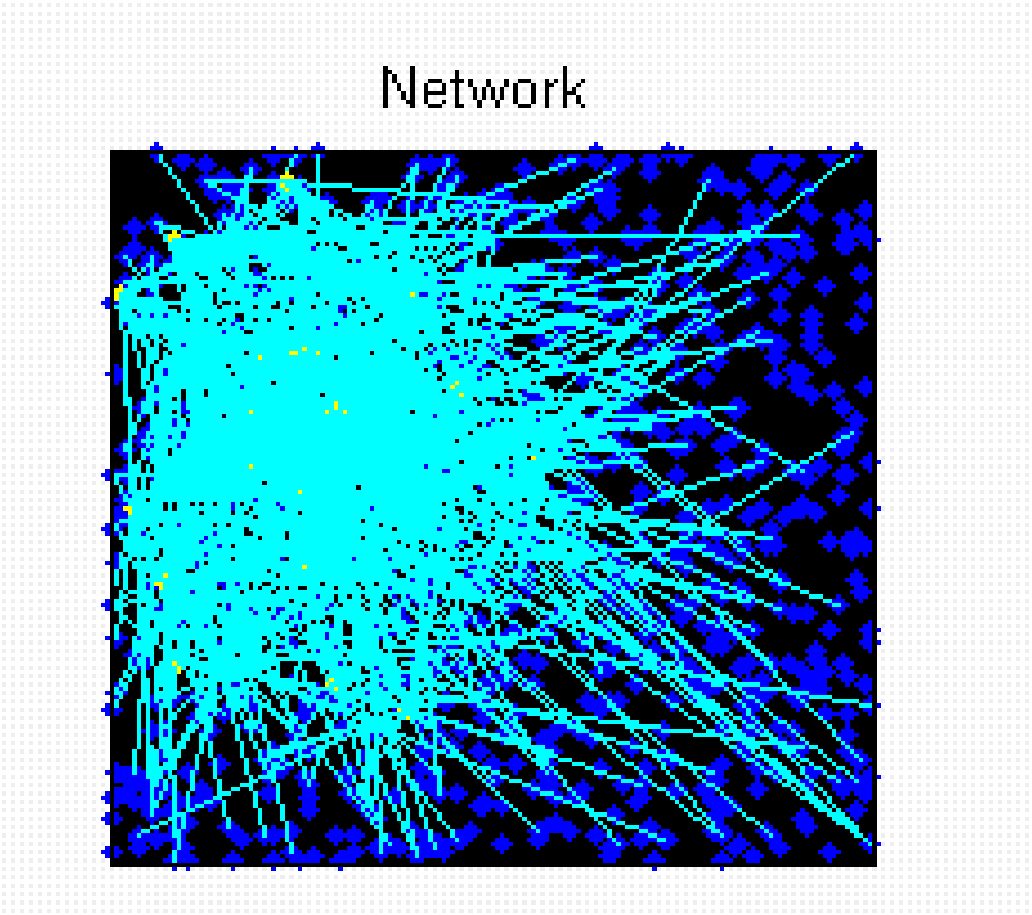


Figure 5

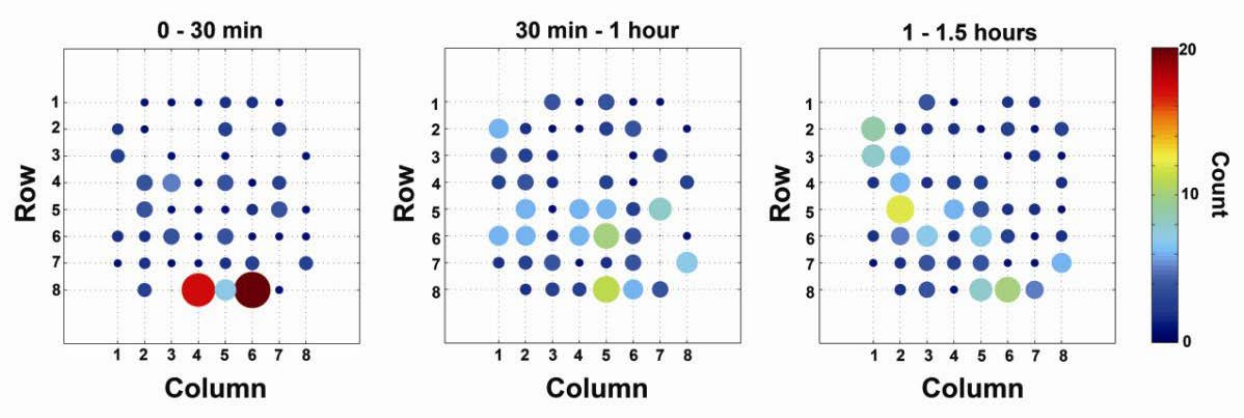
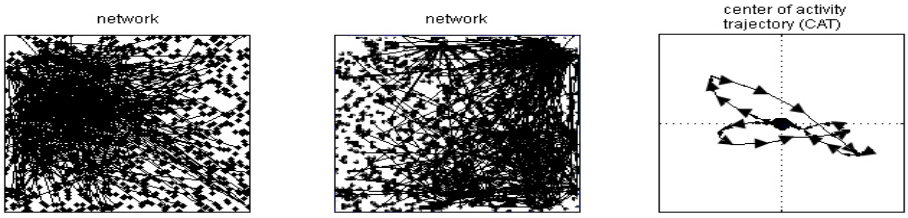

0 - 30 min
30 min - 1 hour
1 - 1.5 hours
Row
Column
Count
1
2
3
4
5
6
7
8
0
10
20

Figure 6



(a) (b) (c)