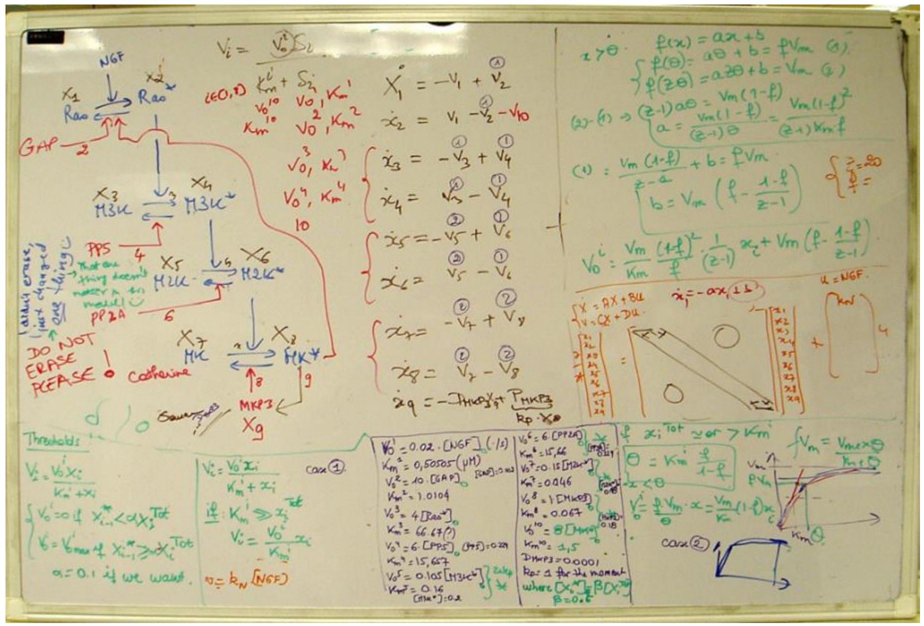
Cognitive constraints, complexity and model-building



Miles MacLeod
(University of Helsinki)

Nancy J Nersessian
(Harvard University)

The relevance of cognitive science to methodological choice

Finnish Centre of Excellence in the Philosophy of the Social Sciences

Background: The Limits of PoS

- Philosophy of science for the most part has ignored cognitive science (or the role of the human agent in scientific practice) by….
- Demarcating discovery from justification or by pursuing only normative theories of evidence and confirmation
- Filing away cognitive factors as ”pragmatic” factors of low or unimportant explanatory or normative value ….
- Developing cognitively loaded concepts like ”intelligibility” and ”visualization” without the input of cogsci.

- PoS has ignored cogsci (or the human agent) for instance in the context of methodological choice.
- Choice usually conceived of as an issue of rational principles alone!
- But there may be rational reasons based on cognitive limitations that favor certain methodologies over others….especially where complexity is concerned.

This Talk

- Aim: illustrate cases from modern computational systems biology in which cognitive constraints are clearly factored into decisions over how and what to model, even entering into explicit methodological strategies reseachers advocate.
- In these cases cognitive factors play a considerable role determining the nature of representation in the field, and standards of explanation and understanding… (i.e. philosophical characteristics of model-building in the field).

Ethnographic Approach

- An ethnographic study of model-building practices in two systems biology labs.

1. Lab G – computational lab: contains only modelers (unimodal researchers). Works by collaboration with experimental labs. Studies a variety of topics concerning metabolic and cell signaling systems.
2. Lab C – a fully equipped wet lab: contains experimenters, modelers and bimodal researchers who do both. Studies particularly Reductive Oxidation Signaling systems.

Method: grounded analysis/coding + longitudinal studies (grad. reseachers)

- 44 interviews lab G
- 62 interviews lab C
- 7 lab G group meetings
- 16 lab C group meetings
- Many hours of Lab C observations
- Lab output: grant proposals, papers etc

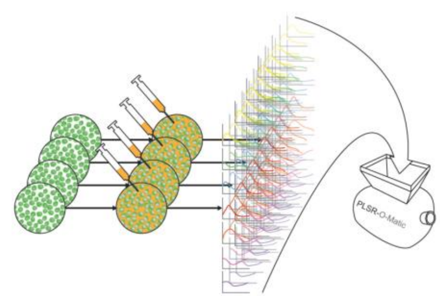
Two Parts

1. Show at least one way sophisticated cognition is depended upon in model-building practices in sys-bio - and the resulting constraints this places on what these modelers can do.
2. Show how methodological responses are shaped with respect to these constraints – particularly the case of Mesoscopic Modeling.

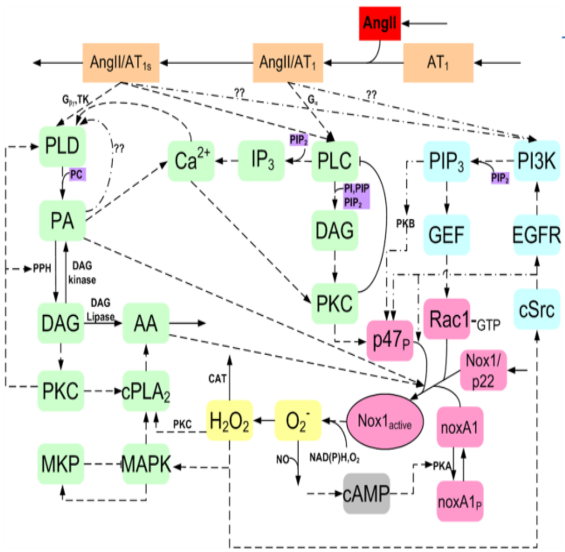
Part 1: Cognitive Dimension of Model-Building in ISB

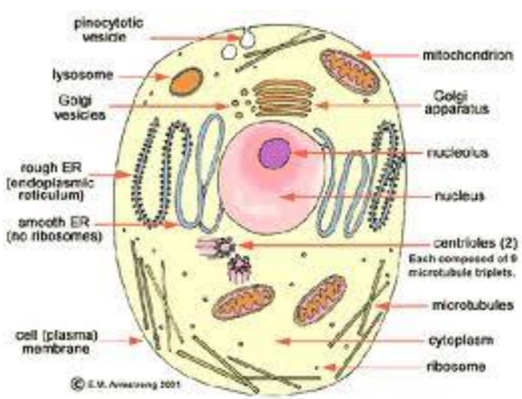
<u>Systems Biology</u>

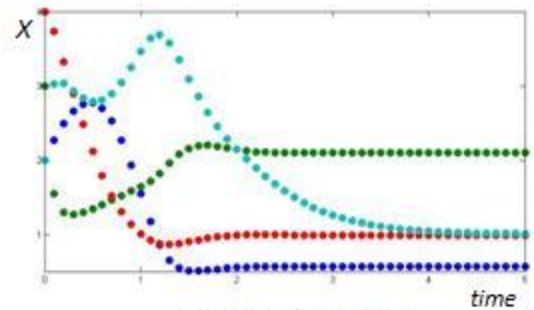
- Field 15 years old: aims to model large-scale biological systems (mostly cell signaling, metabolic or gene regulatory networks) by computational means.
- Facilitated by high-throughput time-series data
- Systems biologists come from engineering. Most work is collaborative (as in Lab G)



Building the simulation model: the canonical version





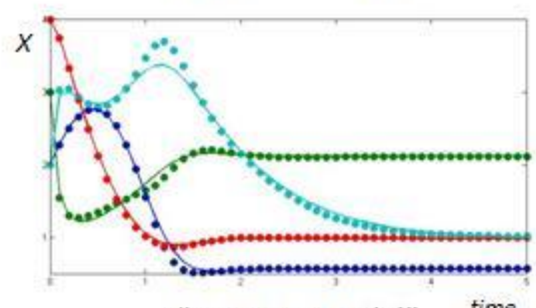


Time Series Data

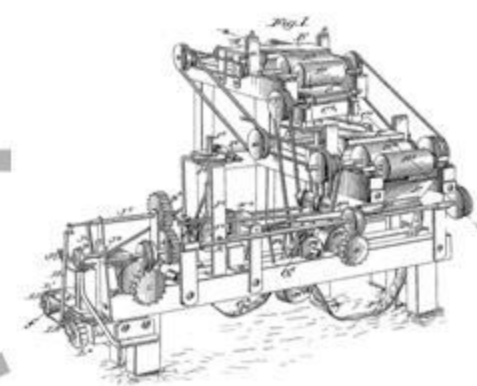
Model of Choice

$$\frac{dX_i}{dt} = \dot{X}_i = \sum_{p=1}^{P_i}\left(\pm\gamma_{ip}\prod_{j=1}^{n} X_j^{f_{ipj}}\right), \quad i = 1,2,\ldots,n$$

"Symbolic Model"



"Dynamic Model"



Parameter Estimation

Unable to produce acceptable fits

Complex problem solving tasks

- Model-building in ISB essentially aims at building understanding of dynamical relationships between variables in a system. Such understanding is itself essential to progressing the model-building process (as we'll see)

- Model-building characterized by complex problem-solving tasks.

1. Complex nonlinear biological networks
2. Particular constraints
   - Data constraints
   - Collaboration constraints
   - Computational constraints
3. Lack of theory for model-building that applies generally

$\Rightarrow$ cognitively difficult search tasks

Researchers have to develop methodological strategies for their particular problems to overcome these constraints.

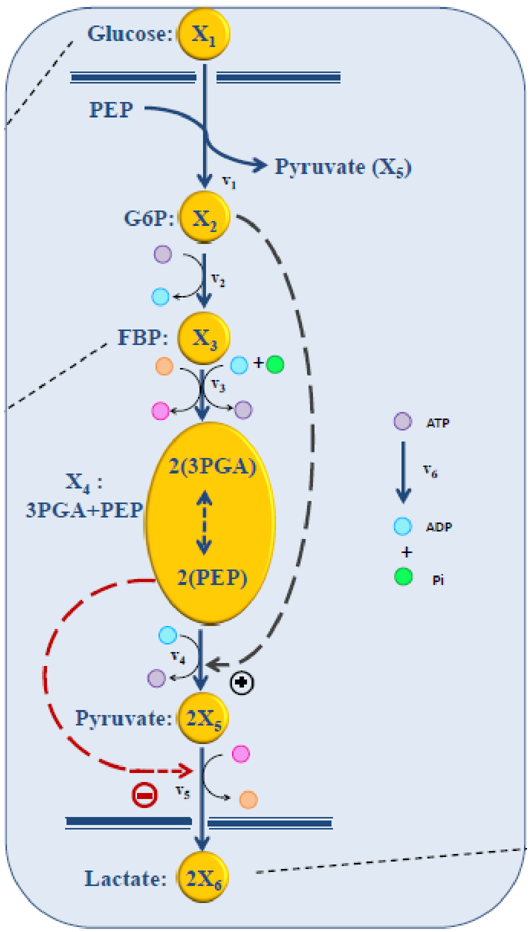
Model-Based Inference

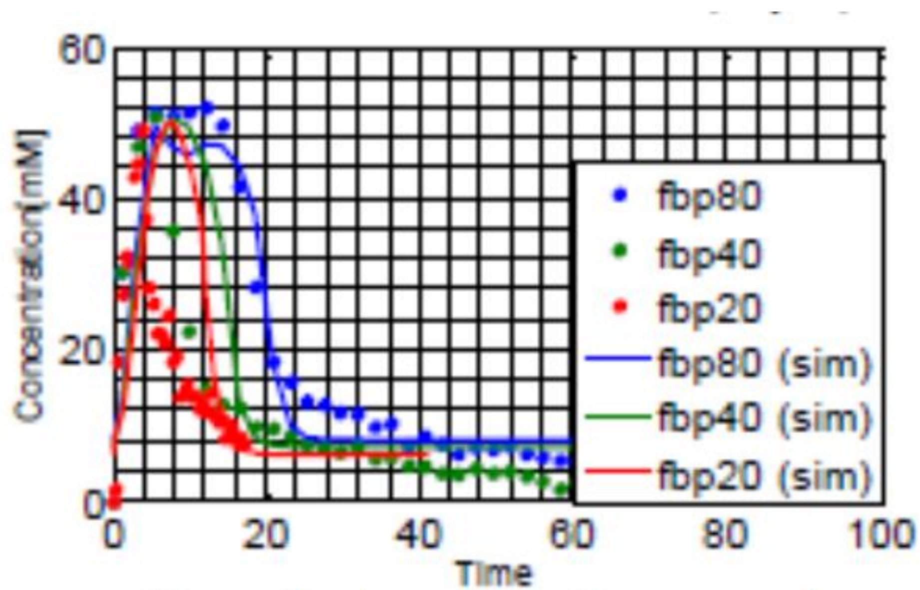
- To build effective models of dynamical relationships modelers need to be able to….

1. Infer missing network structure or inaccurate parameters (at the right places in the network)
2. Infer dominant dynamical dependencies (to reduce complexity)
3. Infer likely subsets of parameter spaces in which to search for global best fits

1.Inferring New Network Structure

<u>G16: Modeling Glycolysis in *Lactococcus Lactis*</u>





- She inferred that an interaction was being damped, hypothesizing that a step function could capture the appropriate relationships.

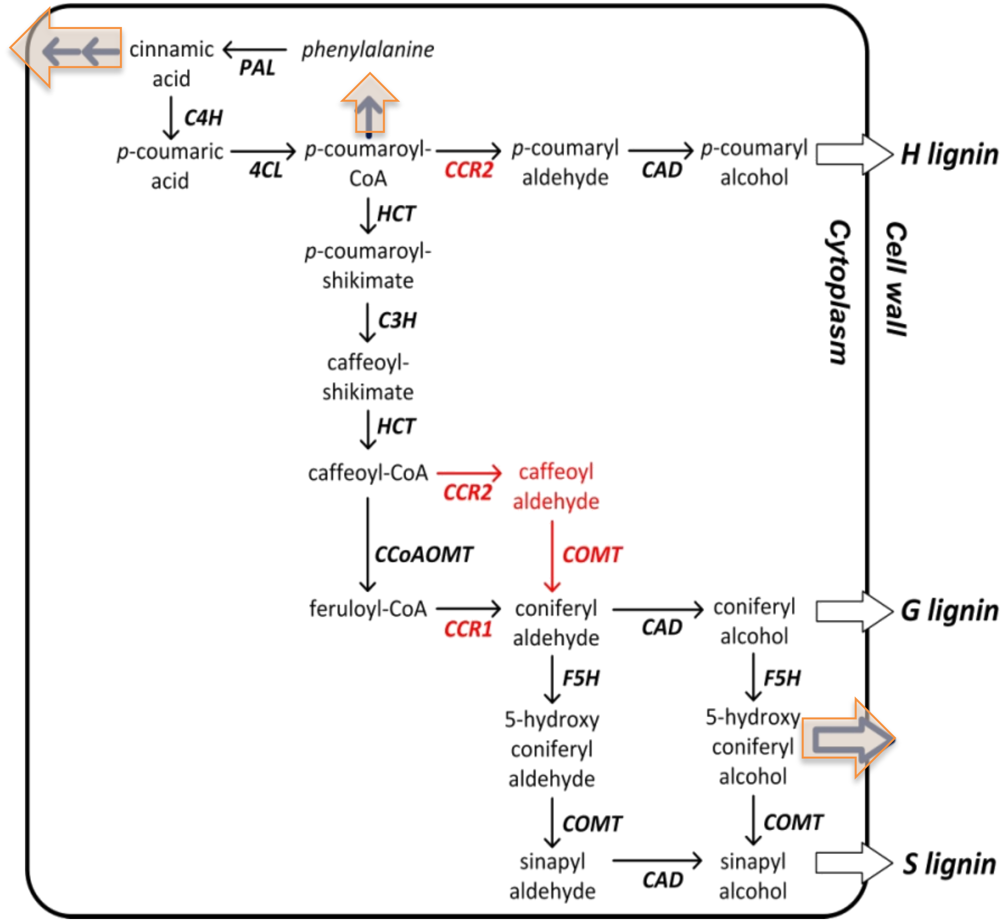
- Her strategy was to try out different step-function based interaction relationships upstream that would propogate through and affect the peak appropriately and fix parameters to see if she could get a model that fit.

- “So mathematically with a step function I can get the results .So I should make some variations like what if this term is affected. What if only this term? And by all those variations I will try to understand what exactly happens.”

G10: Modeling Lignin Synthesis

One of G10's tasks was to model the lignin synthesis pathway in order to better optimize current transgenic biomass producing species to break down lignin.

- He built a pathway from available results….it only worked at steady-state (wild-type equilbrium)
- G10 studied his model structure closely to hypothesize where blockages were happening in the network.
- Thinking about how extra flux might be modulated to give the right outputs, G10 hypothesized particular additional fluxes, which he translated to more precise mathematical modifications, that would relieve the system.

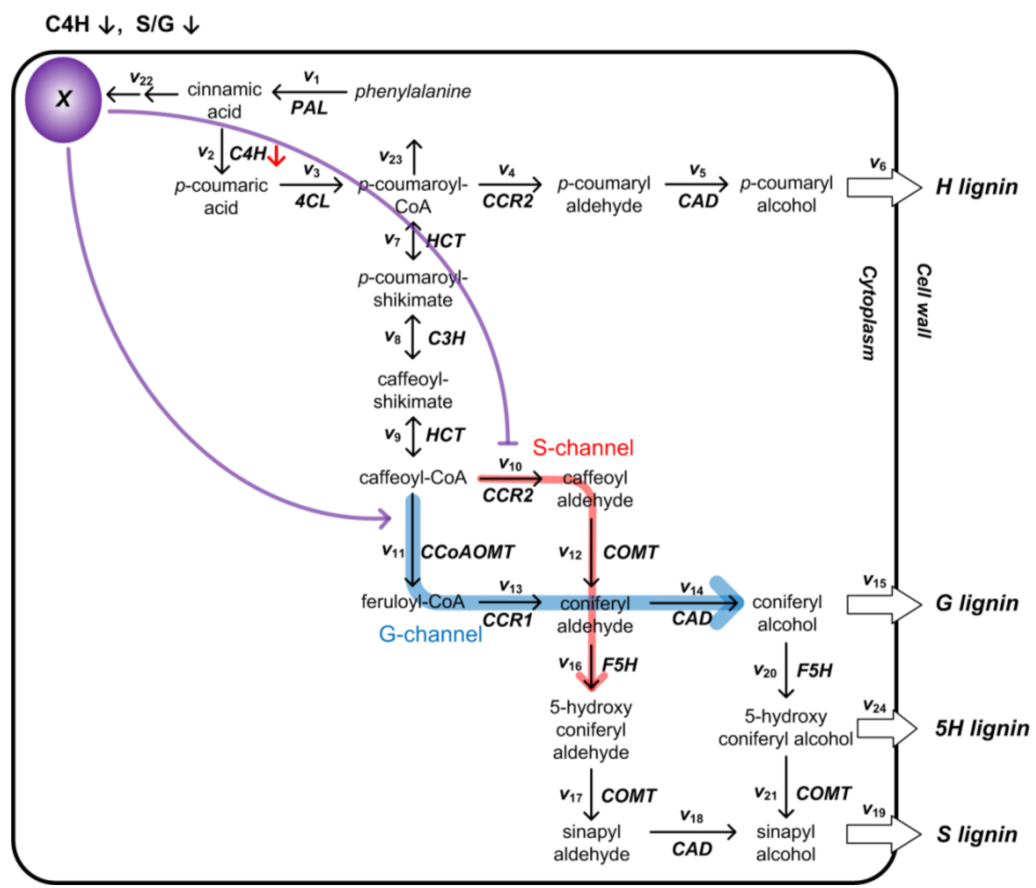


"this is an important piece of knowledge that comes from the model"

Element X:

Using information he had on down-regulation and up-regulation of particular variables and their effects on G and S lignin production, G10 reasoned that G and S lignin production was happening in ways outside of what was mathematically possible within the model.

“So this is actually the biggest finding from our model. So by adding this reaction you can see that we hypothesize that there is another compound that can give a regulation….give a feed forward regulation to other parts of the pathway.”



2.Inferring Dominant Dependencies (sensitivity analysis)

- Inferring which variables have limited effect (under reasonable parameter ranges).

- Mathematical arguments: comparing individual terms in the mathematical formulation
- Using computation or other methods to visually track flux the network and observe dependencies through manipulation.
- Developing statistical techniques that sample parameters (supported by mathematical argumentation) (G10 - Pearson coefficients)

”feel for the model”

- Such inferential processes depend on the ability of modelers to understand how their mathematical models operate…
- “I find glitches in the model, and why is it that, for example…And (in this case) when you look at it there’s no way it can get better because it depends on two things, and those two other things, for example, are increasing. So you can never get it decreasing for a period of time from those two. Maybe something else has a role that I haven’t taking into account.” (G16)

- To develop such a “feeling” mathematical equations need to be interpreted.

“So the thing is…..when you want to solve a mathematical problems, you gotta,…sometimes you use numbers and try numbers, something give you a feel of…like intuitively how this, for example, equation works and all. So I’m trying out numbers and then trying to make the steps kind of discrete. ….like sort of a state machine,…kind of thinking like we’re in this state. And then now this much is going to this other metabolite pool and then at the same time we have less of that. So I’m trying to see what the constraints are by actually like doing step by step sort of thing.” (G16)

This should get faster in the beginni

	0	1	2	3	4
Glc	2000	1997	1995.2	1994.12	
G6P	0	3	1.8	1.08	
F6P	0	0	3	1.8	1.08
FBP	0	0	0	3	1.8
GAP 2(BPG)	0	0	0	0	6
3PGA	9	5.4	3.24	1.944	
PEP	6	3.6	2.16	4.296	
Pyr	0	6	6.8	7.88	8.528
Lac	0	0	1	1	1
NAD^+	5	5	6	6	0
NADH	1	1	0	0	6
ATP	3	6	7.8	5.78	4.728
ADP	5	2	0.2	2.2	3.272

ATP: $\dot{ATP} = -v_0 - v_2 + v_4 + v_5$

$\dot{ADP} = v_6 + v_2 - v_4 - v_3$

$P_i = v_6 +$

It takes no time for 3PGA & PEP to get to steady state together.

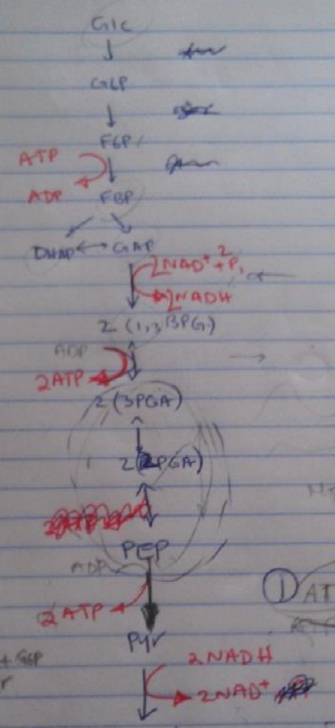
* why do we consume so much P_i?

AMP $\xrightarrow{\text{start of}}$ ADP ?

I wonder if being ... or Moore make a difference. I mean dep... on current time opposed to previ... time point.

Talk to elect her?

Glc + ATP → G6P



state machine

state space model

agent based model

① ATP → ADP + P_i

② FBP doesn't → initial value

P_i buffer

ATP

P_i

Glc + PEP → Pyr + G6P

PEP → Pyr

	⓪	①	② lactate
Glc	2000	1997	1997
G6P	0	3	0
F6P	0	0	3
FBP	0	0	0
2BPG	0	0	0
3PGA	9	9	5.4
PEP	6	0	3.6
Pyr	0	6	3
Lac	0	0	3
NAD^+	5	5	8
NADH	[illegible]	[illegible]	0
ATP	[illegible]	[illegible]	
ADP	1	0	

9 5.4

6 3.6

Simulative Mental Models (Nersessian 2008)

- This understanding (‘feel for the model’) is encapsulated within qualitative and piecemeal mental models that simulate network dynamics.
- These are built on constraints derived from the mathematical model, the pathway diagram, and computational simulation of the model which facilitate qualitative interpretations of dynamic behaviors and effects -like “increasing” and “decreasing”, “inhibiting” and “promoting”.

Bigger Picture: Building cognition

- In general computational simulation plays the central role! Problem-solving is dístributed between computational and mental simulation.
- Simulation helps build cognitive capacities (network understanding) necessary for making inferences.

Computational simulation provides,

1. visual representation of the model‘s dynamics that can be translated into qualitative relationships.
2. Piecemeal and selective representation that can be mentally represented within cognitive capacities
3. Mental model calibration by allowing modelers to check results of mental simulation and correct mental models.
4. testing of mental inferences about network structure and behavior

This “ coupled system“ extends cognitive capabilities to resolve these complex dynamical problems

A coupled system

Inference &
Calibration
(mental simulation)

Mental Model → Computational Simulation

Mental Model ← Computational Simulation

Visualization
(envisioning)

Cognitive Constraints

- Computational simulation can extend cognitive capacities but is still limited by them...
- People in general aren't very good at building and using causal mental models even less so with nonlinear systems (Doyle, Radzicki and Trees 2008).
- Envisioning and mental simulation tasks constrained by working memory and the qualitative requirements of these models.
- Complex features of networks introduce more factors that need to be simultaneously processed and which are often quantitatively sensitive. e.g.
    - feedback relations
    - competitive reactions and multiple network functions of elements
    - more interactions at any node
    - longer chains between relevant variables

The more of these the harder it is to conceive relationships and to mentally simulate the dynamics.

- Researchers cite the concentration and intensity required for this kind of research. ” It’s not like rocket science, but it is not easy as well. Like, you really need to be very much focused when you’re working on it. So, cause you…you focus on something and then something else goes wrong and then how you go about…umm…checking everything and putting things together …umm…you gotta be very…umm…like, present minded if that works.” (G16)

Part 2: Factoring in Constraints

- Larger scale or more realistic representations of systems have more of these features.
- Cognitive constraints place their own limits on the scale and complexity of models that can be constructed (and thus on the complexity of the systems that can be represented).
- Modelers are aware of this. ”But I think he’s (G4) been in the real world long enough doing this systems stuff long enough that he knows to start small… so when I first came to him, I had the proteomics systems …. we’ve seen about 10% changes in about all the changes in all the systems of the CF cell…versus a non-CF cell. Now when you think about the number of systems that are in cells, 10% changes in all of those systems or changes in 10% … is a considerable amount, I mean that is a lot of information. So when I first went to G4 I’m like let’s just.. here it is… ya know…He’s like you are diluting yourself. So then we decided … to narrow it down to energetic pathways that are very well modeled.” (G70: experimental collaborator).

Epistemic Consequences

- We’ve noted that modelers adapt their model-building targets to fit their various constraints over the course of model-buidling (*adaptive problem solving*).
- In general the scale of systems or system representations they target are limited to ”mid-size” models (to keep complexity manageable), while abstracting out external influences on their networks. They remove or simplify nonlinearities.
- Goals limited to capturing specific input and output relations of networks (rather than overall system dynamics).
- Concepts of understanding at play in these practices diverge from mechanistic understanding of whole systems favoring causal understandings of just slices of them.

Representations take on particular forms and have particular epistemic justifications that meet problem-solving capabilities.

Mesoscopic Modeling

- Rhetoric of systems biology is that control is distributed over large scales, so large scale models required – but current practices don't achieve those scales (how then to rationalize them?)
- "If one would survey all computational systems models in biology, published during the past decade, one would find that the vast majority are neither small enough to permit elegant mathematical analyses of organizing principles nor large enough to approach the reality of cell or disease processes with high fidelity." (Voit et al. 2013)
- Voit et al (2013) and Voit (2014) advocate formally a mesoscopic approach:

- "models of an intermediate size that are neither purely empirical, nor contain complete mechanistic detail." (tractable but abstract mathematical representations of interactions that can be fit to the data)
- These models can be manageably developed, middle-out!
- They are cognitively and computationally tractable.

Learning through mesoscopic modeling

- The role of cognition in model-building justify the need for restricting modeling to mid-size models. These models provide a,

”coarse structure that allows us to investigate high-level functioning of the system at one hand – and to test to what degree we understand, at least in broad strokes, how key components of a biological system interact to generate responses” (Voit et al.)

- A mesoscopic model can then help build further understanding of systems through processes of hierachical learning.

”Like a flight simulator that is used in pilot training, a disease simulator could mimic simple, frequently encountered situations, as well as very rare and complex emergencies, and thereby hone skill and intuition.” (Voit et al.)

- Ability to learn through processes of mesoscopic modeling serves as a cognitive rationalization of why systems biologists build mid-size models. Mesoscopic models are cognitively tractable starting points.

Conclusions

- Mesoscopic modeling describes much modeling in systems biology.
- We've tried to show at least one of the ways in which model-building is cognitively dependent and cognitively constrained; which shape representations and methodology in systems biology.
- The role of cognition in method choice has been made more explicit by researchers themselves in this case and given broader justification within the field.
- This case serves as an illustration of the role cognitive constraints can play in methodological choice and decision making, with interesting questions to be asked about how this field might differ in this respect from others.
- The justification for limiting model scale is not on the basis of the organization of nature nor is it purely epistemic. But it still represents a rational choice……(i.e. On the theory of bounded rationality)

If the goal of PoS is to understand scientific decision making and justification then in this case the context of discovery and cogsci intrudes…

Simulative Mental Modeling

(Nesessian 2008)

- Modelers build simulative mental models of causal relationships in their networks; interpret flux flows through pathway diagrams using the mathematical equations (the constraints)– in order to identify errors and screen plausible modifications.
- These models have the following characteristics (which reflect earlier results mainly on causal (mechanical) models in physics and engineerng):

i. Modal (maps pathway structure) – Analogical (using various analogical interpretations to understand the equations)
ii. Qualitative (Roschelle and Greeno 1987, de Kleer and Brown 1981)
iii. Piecemeal (Hergarty 1992, Schwartz and Black 1998) and selective (elements needn't be contiguous to be represented, modelers black box and bracket)
iv. Externally coupled with visual representations
v. *Envisioning* process and inference coupled with mathematical knowledge (a relational framework, Roschelle and Greeno 1987)

Expertise helps: "It's not something that's precise,.... you need a lot of intuition and experience, just to figure out what components you want to include and what you want to exclude from your model (when inferring relevant structure). Because everything is so intricately linked in biological systems". (C7)

Building Mental Models

- Building mental models (using de Kleer and Brown, 1983):

1. First step is to be able to replicate the dynamics mentally (learning through *envisioning*)
    - *propogating* values/changes through the network
    - relating mathematical/theoretical knowledge
2. Checking models (*Running)*
    - standards: consistency, correspondence, robustness

- A working mental model can be perturbed and experimented with and in turn help generate inferences.