Conceptual and methodological biases in network models

Ehud Lamm
The Cohn Institute For the History and Philosophy of Science and Ideas, Tel Aviv University, Tel Aviv 69978, Israel
ehudlamm@post.tau.ac.il

Keywords: variation, weak linkage, epigenetic inheritance, evolutionary transitions, evo-devo

Abstract: Many natural and biological phenomena can be depicted as networks. Theoretical and empirical analyses of networks have become prevalent. I discuss theoretical biases involved in the delineation of biological networks. The network perspective is shown to dissolve the distinction between regulatory architecture and regulatory state, consistent with the theoretical impossibility of distinguishing a priori between "program" and "data". The evolutionary significance of the dynamics of trans-generational and inter-organism regulatory networks is explored and implications are presented for understanding the evolution of the biological categories development-heredity; plasticity-evolvability; and epigenetic-genetic.

Many natural and biological phenomena can be depicted as "networks", that is as a set of "nodes" (or vertices) connected by "arcs" (or edges). Many of these networks seem to fall into the formal (theoretical) category of scale free networks**.** The network metaphor, as well as the mathematical language for studying them (a branch of graph theory) and the empirical attempts to elucidate Gene Regulatory Networks for Development have become prevalent. In this paper I argue that the modeling of biological networks should stress the significance of networks that cross organism boundaries and employ the distinction between network state and architecture cautiously. I emphasize the conceptual biases involved in delineating biological regulatory networks, and suggest how understanding the evolution of boundaries is related to the study of evolutionary transitions.

Before discussing networks it is helpful to reconsider one of the fundamental building blocks of gene regulatory networks, cis-regulatory elements. The *lac operon* remains a classic and still very telling example of the fundamental aspects of the constituents of regulatory networks. James Shapiro (2002) highlighted general conclusions from the lac operon regarding genomic regulation: "(1) Information transfer occurs by the use of chemical symbols to represent empirical data about the physiological

environment; cAMP, allolactose, and protein phosphorylation levels represent the availability of glucose and lactose. (2) The regulatory network integrates many different aspects of cell activity (transport, cytoplasmic enzymology, and energy metabolism) into the transcriptional decision. *In other words, it is literally impossible to separate physiology from genomic regulation in E. coli - and, indeed, in any living cells*" [1] (my italics). Some of the implications of this observation will become apparent in the discussion below which deals with regulatory networks more generally.

The data-program distinction, probably originating from the application of computational metaphors to the study of biological networks, underlies a lot of the conceptual complications analyzed below, specifically the distinction between architecture and state of networks. In addition, as I will show, the computational perspective on networks can be a slippery slope leading to conceptually problematic questions such as what are the constituents of the network as opposed to external "*inputs*", what constitutes a change *to* the network, rather than a change of the network's "state", and how to delineate the boundaries of networks. I will demonstrate how some of the complications arising from the computational perspective on biological networks stem from misuse of computational concepts studied by Computer Science. The notion of "genome informatics" introduced by Shapiro[2] should thus not be understood as a description of how computation is performed, or the only level of analysis needed in order to explain it, but rather as a framework for elucidating the task requiring computation (cf. the discussion of Marr by Bechtel [3]).

In a statement echoing Shapiro, Wilkins observed that "it is intrinsically impossible for theoretical treatments to fill the gap" between generic accounts of networks and understanding the patterns of network evolution in the real world of living things.[4] The following list illustrates several properties of biological networks that seem to be of particular interest from an evolutionary perspective, and that perhaps do not receive enough attention when concentrating on the properties of generic scale-free networks: (1) cells (and tissues) belong to different (stabilized) regulatory states (i.e., cell differentiation); (2) duplication events (e.g., whole genome duplication), can duplicate networks and subnetworks, not only individual genes, leading to redundancy[5] and

opening up evolutionary possibilities;[6] (3) networks exhibit overlayered circuit design with subcircuits that may have evolved independently and at different rates;[7] (4) regulatory networks may cross organism boundaries; (5) epigenetic as well as genetic factors are involved in regulation and in establishing the hereditary properties of networks.

I concentrate here on the conceptual implications of emphasizing phenomena of the last two types, cross-organism regulatory interactions, and the role of epigenetic network changes. Networks are dynamic entities, changing during development and evolution, and the study of their shifting boundaries is one important way of studying these dynamics.

Generalizing Networks – Biases in Defining Boundaries

Biological networks of a variety of kinds (e.g., gene regulatory networks, transcription networks, protein interaction networks, metabolic networks, developmental genetic networks as well as neural networks, immunological and hormonal interactions, and inter-organism interactions) are interwoven, both inside and outside the organism**.** When discussing particular networks we are thus necessarily looking at *sampled networks*, based on non-random sampling along several scales (cf. the discussion of sampling in Ref. 8). The networks that surround the focal one are hierarchical (e.g., cellular organisation, cells, tissues, organisms etc.). The sampling is biased both because of obvious experimental biases and because the questions that are considered interesting and important are biased (e.g., by what is considered evolutionarily significant). As a way of highlighting the hierarchical view of biological phenomena I suggest extending the networks being studied beyond the traditional boundaries delineating to a large extent the types of interactions depicted in regulatory networks of various kinds.

Specifically, we should keep in mind that regulatory networks cross organism boundaries. There are many examples of such phenomena, several of which are discussed below, but the implications of this observation run deeper than might at first be supposed, and attention to cross-organism networks may help overcome some of the conceptual biases alluded to above. The fundamental importance of this

observation from a network modeling perspective becomes clear when it is noticed that deciding on the boundaries of the network (i.e., intra-cellular, intra-organism etc.) determines the network connections found, and as a result also determines network topology and other network-level properties. An indirect connection between two elements may only be visible when including a third, connecting element, in the description of the network.

The following examples illustrate properties that may be exhibited by cross-organism networks, and some of the evolutionary questions involved in their analysis.

Mitochondria and plastids are understood to be the result of endosymbiosis, followed by gene transfer between the symbionts. Essential components of both electron transport and mitochondrial protein synthesis are encoded *by nuclear genes* with mRNA that is translated by cytosolic ribosomes followed by transport of the products to the mitochondrion. The questions this raises about the reasons and evolutionary history of this division of labor are well known, and the study of the evolutions of plastid has made important advances in recent years.[9-11] Clearly, if mitochondria originated from a distinct organism, the constituents of the genetic network moved across organism boundaries. Regulation had to happen across these boundaries as well. A second and perhaps more convincing example of a cross organism network (since the network clearly operates across the traditional boundaries of the organisms involved) is the bacteria-legume symbiosis leading to nitrogen fixation. The signaling between Rhizobia and plants is *bi-directional and specific*. What makes this example even more remarkable is the fact that these interactions exhibit properties that help explain the evolution of intricate aspects of cooperation, namely *sanctions*.[12]

The evolution of the mitochondria illustrates that not only can network connections cross organism boundaries: the *boundaries and the locations of network constituents may shift during evolution*. These are the type of processes that the network perspective should embrace, explore and analyze theoretically as well as empirically (I am intentionally blurring the endosymbiont/organelle distinction, which maybe too restrictive. cf. Refs. 10,11).

It is striking that the mitochondria-nucleus network is formed not solely by virtue of components being transported, and because the locations of the constituents can vary evolutionarily, but also via *specific signaling*: some complex mitochondrial signal sequences include "stop transfer" signals, which when recognized during transport of the mitochondrial protein precursor halt translocation causing the protein to become an integral membrane protein. The present cross-boundary regulation depends on specific regulatory mechanisms not just universal regulatory apparatus (e.g., *cis*-regulatory elements).

The existence of specific mechanisms makes it more difficult to dismiss examples of cross-boundary networks such as these as simply a single network (e.g., the mitochondrial) whose "data" is distributed between two storage locations. The network spans two compartments with signals flowing from one to the other (and, furthermore, the model should reflect that there is routinely more than a single mitochondrion per cell). This perspective calls for a co-evolutionary model based on interacting networks, which are *also* modeled as a single network, by virtue of the interaction and signaling between the components. The more tightly coupled the organisms are (as in the case of endosymbiosis) the less need there is to model them as independent networks, but in general the probabilities and timing of interactions between organisms will affect the network dynamics of the integrated network and thus have to be reflected in models of networks encompassing more than one organism. These properties can obviously be influenced by inputs from other actors in the network, by the behavior and state of each organism, and so on. The less tightly coupled and obligatory the interactions, the more crucial it is to explicitly model the parameters affecting interactions, since evolutionary dynamics may lead participants to escape from the network, become parasitic etc., and these changes may lead to changes in the behavior of the integrated networks. Significantly, models that presuppose the degree of integration, or worse treat all interactions uniformly thereby hiding the fact that groups of elements (e.g., genes) – bundled in organisms – share developmental and evolutionary fate, are not suitable for studying how the degree of integration changes evolutionarily and how the integration is achieved and maintained (e.g., as a result of gene transfer from one participant in the network to another). There is no *a priori* reason to suppose that transitions similar to those between the mitochondrial and nuclear genomes can only occur when the two organisms are

endosymbiotic and has to lead to an organelle-like relationship with the host. For a discussion of the complexity involved even in the evolutionary dynamics of the regulatory integration in the endosymbiotic case, see Ref. 11.

The observation that nuclear genes can lead to *highly specific rearrangements* of the mitochondrial genome and affect the mutation rates of other plastids,[13] is of particular significance for integrating genomic variational mechanisms with the network perspective. Not all these cases are understood at the molecular level, but the fact that plastid DNA Polymerase is nuclear, seems highly significant. How did such a division of genetic labor occur, and why? These observations raise similar issues to those raised by the observed genomic interactions and re-patterning that result from hybridization and polyploidization and possibly other forms of genetic stress. Such interactions may lead to repeatable, wide-ranging yet specific, genomic and chromosomal changes that involve massive epigenetic changes involving DNA methylation and histones modifications, transcriptional and post transcriptional gene silencing through the RNA interference (RNAi) system, as well as targeted genetic changes.[14] In both cases interaction between organisms causes specific genomic rearrangements that may prove to be adaptive. Specifically, it is of great interest to unravel the genomic mechanisms involved, their regulatory control (e.g, excitatory or inhibitory) , their specificity and evolutionary history, and their possible effects on reticulate evolution, the importance of which is receiving growing appreciation.[15, 16]

According to this picture there are several levels of interaction that need to be considered: cross-organism interactions that can undergo adaptive evolution, stress related plastic gene responses and the evolution of stress-sensitive genes,[17] as well as genome-level heritable responses to hybridization and polyplolidization. The latter may depend on intimate details of genomic architecture and mechanisms, both inter- and intra-chromosomal, and may be adaptive responses.[14]

There are many other examples of cross-organism networks: There are species that allow certain bacteria to be vertically inherited through the host's oocytes as observed in sponges, clams and aphids (reviewed in Ref. 18). This can be considered *an epigenetic conservation mechanism of cross-organism (genetic) networks*. Coral

microbiology has led Rosenberg and colleagues to propose the hologenome theory of evolution according to which the hologenome, which refers to the genomes of a host and its symbiont population (the holobiont), should be considered the unit of natural selection.[19] Metabolic interactions in the gut illustrate some of the many types of interactions and issues involved in cross organism networks (between mammalian host and bacteria, as well as among the symbiotic bacteria). These include: highly coupled metabolic co-processing involving both the host and microbial systems, bacteria that manipulate host gene expression, and horizontal gene transfer between the bacteria in the gut.[20-23] Cross organism networks may have a crucial role in development. For example, bacteria have an important role in the *development* of mammalian intestines.[21, 22]

The term *superorganism* has been applied to the complex ecosystem encompassing a host organism such as a human and its symbiotic microbiota and parasites,[24] and indeed the close and interactive relationships in the cross-organism networks in the gut tend to the more cohesive end of the gradient between loosely and tightly coupled interactions. It should be emphasized that while sharing developmental fate, the participants in these networks do not necessarily share genetic fate, and may thus more easily escape the relationship.

The examples mentioned above show that cross-organism networks can involve a variety of types of interaction and regulation, be obligatory in varying degrees, and involve from little or no genetic exchange to almost complete genomic integration. Each type of relationship is the result of different evolutionary pressures, and constrains subsequent evolution of the constituent organisms and of the network as a whole as an adaptive unit. The dynamics of the cross-organism network depends not only on the ostensible network connections, but also on the evolvability and developmental plasticity of the organisms that constitute it, which are sources of variation for the evolution of the network. The organisms encounter the environment and selection as part of the network (though if the relationship is not wholly obligatory some may escape this context), but are also usually facing selection at the level of the individual organism. The network is both a superorganism and the selective and inducing environment that the constituent organisms occupy.

The discussion above illustrates the reasons to extend the boundaries of networks to include cross-organism interactions, and some of the implications of such a move. It should be emphasized, however, that the appeal to networks by biologists, is not only a descriptive strategy. Networks can be invoked to ameliorate well known difficulties with the "one gene-one trait" paradigm. The difficulties include non-additive genetic effects as exemplified by polygeny, pleiotropy, epistasis; gene shortage, and the small variation in DNA sequence between widely divergent species, as well as neutrality and the fact that single gene changes have on average a neutral effect (cf., Jablonka[25]). Networks can also be seen as helping counter the challenges from molecular biology for the one gene-one protein paradigm, while retaining the explanatory and casual priority of the genetic information. An appeal to networks for these purposes, however, without acknowledging the evolutionary factors discussed above, can be seen as an attempt to patch up the (gene centric) Modern Synthesis. The marginalization of the implications of developmental processes for understanding evolution is retained by retaining the explanatory priority given to genetic factors over genomic as well as developmental factors.

Extending the boundaries of the networks being studied can overcome this difficulty, and ensure that the role of genomic and developmental factors is reflected in more of the networks being analyzed. Put differently, by not including cross-organism interactions and environmental interactions in the networks we study, we are risking misunderstanding network topologies – this being an extreme form of *biased* sampling – due to prior theoretical commitments.

Cross Organism Networks and Multi-Level Selection

Networks provide a way to talk about selection operating beyond the level of the gene, while remaining "gene centric": selection does not operate directly on individual genes, but on genetic networks. However, this does not amount to accepting multiple levels of selection, or selection operating above the level of individuals. Here I only want to point out that the notion of networks used is (implicitly) relevant to this issue. In the preceding section I mentioned several factors that influence the selection experienced by the participants in generalized networks.

Multi-level selection may not be able to capture the types of interactions emphasized here, because interactions between compartments (e.g., organisms) lead to "levels" that are not hierarchical. Selection is not either within or between groups (i.e., collections of genes, or organisms). Rather, the cross-organism network perspective emphasizes the shared interests of elements inside a group with elements outside it, in addition to the shared interests that groups may have, and the interest of individual elements (be they genes, individuals or higher level entities).

Figure 1 illustrates some of the possible scenarios. A and B can be understood as individual organisms, and the inner circles as genes, but the discussion can be generalized to higher level entities (I am not making any empirical claim about the actual existence of interactions of the sort depicted in the diagram).

The within-group competition between a and b, is understood better when the network is expanded to include organism B, since a and b in fact interact and belong to one network. c which is not directly involved in the network involving a and b, none the less belongs to the same "coalition" since it influences A's behavior (e.g., choice of niche), thereby increasing the probability that A and B be adjacent.

[Figure 1 around here]

This scenario is meant to illustrate that neither examining the within-group selection between a, b, and c nor the between-group selection between A and B is sufficient. Moreover, *both* these aspects taken together miss the relationship between some of A's constituents and some of B's. It is tempting to argue that the groupings suggested by the diagrams are wrong, and the network (spanning both organisms) is the natural grouping. This perspective, however, is in danger of missing the point that A and B are cohesive and their constituents share evolutionary fate, so that selection operating on A and B must be taken into account. *Both* types of groupings have to be factored in. To complicate matters, changes in element c may lead to the activation of a different network (not shown) connecting other constituents of A and B. This suggests that, with or without multi-level selection, a co-evolutionary account might be needed, as alluded to in the discussion of the examples of cross-organism networks, above. In other words, adding higher level groupings and selection is not enough by itself to

fully capture the implications of networks that cross the boundaries of entities facing selection.

Generalizing Networks – Biases of the Architecture-State Distinction

Networks change over ontogenetic as well as phylogenetic time, both as part of development and physiology and in response to unexpected external stimuli. It is typical to distinguish between three different categories of network change: 1. Evolutionary; 2. Developmental (typically in the time-scale of hours to decades) 3. Physiological (typically sub-seconds to weeks). During evolution network interactions that are regulated can become fixed (e.g., canalization and genetic assimilation of various types), and vice versa.

There clearly is a distinction to be made between connections that *can* occur in the network at time *t* (what I will call potential network connections), and connections that *are* happening. This distinction highlights an important methodological issue: when describing a network we can display the (regulatory) connections between components (e.g., genes), which determine the structure of the network, or we can show the *active* connections (regulations). These capture the notions of potential and active connections, respectively. Both views are essential (and, of course, interrelated), but it is easy to confuse them, or be ambiguous about which is being described. Davidson and his colleagues refer to the "*view from the genome*" in which all relevant inputs into each cis-regulatory element that occur in all cells at all times are shown at once, and the "*view from the nucleus*" which highlights only those interactions occurring in given nuclei in a particular time frame.[26]

I focus here on the architecture of the regulatory network rather than description of active connections, and thus on what I called the potential connections at any given time. It is helpful for the following discussion to distinguish between three levels (or types) of description of networks. Consider a network consisting of elements A, B and C, where A regulates B, but only in the presence of C. According to what I term the *structural view* A is connected to B; in the *network state* view A is connected to B only when the network is in the state C; while in the *network behavior* view A is connect to B only in C, and when A is expressed. According to this classification the

network state view is a midway point between the structural view, ostensibly specifying the (relatively static) network architecture, and the network behavior view. It is on the ability to distinguish between network architecture and network state that I concentrate.

As an illustration of the difficulties involved consider epigenetic changes. Should methylations or chromatin remodeling and similar epigenetic modifications be considered changes *to* the network or as the same genetic network, at different *"states"* or with different *"parameters"* if you will (to use a Computer Science metaphor)? The latter perspective, of a "two tier" model distinguishing epigenetic state changes and genetic changes that are regarded as changes to the network architecture, may reflect the implicit assumption that genetic networks only change evolutionarily due to mutations and selection (i.e., some form of Neo-Darwinism). Epigenetic changes to the network may lead to altered network behaviour, for example to the silencing of genes, and hence to these genes' "disappearance" from actual network behavior. Moreover, such epigenetic changes can be inherited and influence evolutionary dynamics (e.g., Zuckerkandl & Cavalli[27] about evolution of complex traits; Lamm & Jablonka[14]). Disregarding these shared properties of epigenetic and genetic changes with regard to network dynamics in classifying network change is thus a reflection of prior theoretical commitments. This can arguably be seen as taking the computation metaphor too far, as well. Transcriptional memory in GAL (galactose-utilization) genes [28, 29] is a nice example of how the inheritance of epigenetic changes can be adaptive, and highlights that way in which network change can be part of the explanation of adaptation, while at the same time calling for an evolutionary explanation of the regulatory regions themselves as adaptations. A transcriptional self-sustaining loop is most likely responsible for white-opaque switching in *Candida albicans*, a change in phenotypic state that involves a change in cell appearance, mating behavior, and preferred host tissues which is heritable for many generations.[30] This example illustrates the phenotypic scope and long term heritability of epigenetic changes in transcriptional networks. The apparatus for controlling transcription, which establishes the network, opens up trans-generational possibilities, even if the transcriptional memory is directly effective only for a limited number of generations. A network evolutionary account of a phenomenon such as this should not presuppose a distinction between two types of

network change (i.e., genetic and epigenetic), but rather explore all relevant changes and their influence on one another.

While it may be tempting to look for criteria that might make the distinction between architecture and state clear-cut and formal, based on classes of biological entities, we should be cautious. Since epigenetic changes can be inherited and potentially have evolutionary significance, basing the distinction between architecture and state on the sequence/non-sequence distinction is unwarranted. Coincidently, the lac operon is a clear example that regulation cannot simply be understood as a sequence based phenomenon (as emphasized by Shapiro[1, 2]), both in terms of the molecules involved in the regulation and the physical changes that establish the different regulated states. The alternative, basing the distinction between architecture and state on generational boundaries, with state defined to consist only of change that is confined within one generation, is also not straightforward. Network states may be developmentally recreated in the next generation, even if not inherited.[25] Additionally, inheritance rather than being an instantaneous reset back to an underlying network architecture, is better viewed as a process of changes reproducing a new biological entity of the relevant kind.[31] Thus, deciding on a point in time separating state changes from changes to network architecture reflects theoretical commitments about inheritance and evolution and is not a theory-neutral description of network dynamics. Combined, these arguments show that distinguishing between state and architecture based either on the distinction between genetic and epigenetic or between intra-generation and cross-generational change relies on what might be called Neo-Darwinian theoretical commitments. If distinctions between changes of various kinds are to be integral to the network approach they should emerge from (multi-generational) network dynamics, not from problematic theoretical presuppositions, and will ultimately be context dependent rather than universal distinctions.

A second problematic issue related to the notion of network state is delineating, in advance of studying the network dynamics, between network constituents and mere "external" inputs. Consider, for example, bacteria quorum sensing. The signaling molecules that come from outside the bacteria (clearly "input", one might suggest) are the same as those produced by the bacteria itself (i.e., autoinducers), that may even react to its own signals. Another example is once again the *lac* operon. When both

glucose and lactose are available, glucose is preferred. This happens because glucose has a regulatory role, when pumped from outside the cell, while internally produced glucose (when lactose is metabolized into glucose and galactose) does not. These examples show that the same molecule can be seen as part of the *network regulatory mechanism* or as external to the network (*input).*

One immediate consequence of the above arguments is that deciding on the time-scale studied is not a theory-neutral decision. Specifically, it should not be implicitly assumed that changes in "state" are ontogenetic phenomena, and that changes of the network itself are, in contrast, necessarily congruent with generational boundaries. Biological networks are dynamic interacting networks. Attempts to *statically* define what constitutes (or can potentially constitute) a **"***part***"** of the network regulatory mechanisms and what constitutes the **"***state*" of the network is in at given time, can lead to arbitrary distinctions that are not sustainable.

The untenability of the Program-Data dichotomy

While the earlier discussion of epigenetic and chromatin changes problematized the notion of network *state*, the examples just discussed problematize the notion of network input.

When studying a molecule, or a specific cellular change (keeping in mind that network activity may be influenced by sugars, concentration changes, etc. as well as viral gene transfer, HGT etc.), it is impossible to know, before elucidating the network dynamics, to which category it belongs. The distinctions discussed above should emerge, when they are relevant, from the actual networks dynamics (since even focusing on localized interactions can be misleading, as we just saw), and cannot be arbitrarily imposed (unless as heuristic aids that demand further scrutiny) without the baggage of prior conceptual commitments about evolutionary change. Commitments which lead to unsustainable network boundaries of the type illustrated.

The distinction between program and data, that may underlie the distinction between the genetic network and input, and possibly the distinction between relatively static network architecture and dynamically changing state, is problematic for fundamental theoretical Computer Science reasons as well as the biological reasons just discussed.

The simplest manifestation of this fundamental result is the existence of Universal Turing Machines, capable of accepting as input the specification of any other Turing Machine and emulating it. The implication is that “data” (input) can be a specification of any “program” (Turing Machine). Accordingly, it would have been surprising if a simple distinction was found to hold for biological phenomena, not the least because their origin in contingent evolutionary history. An alternative way of putting this same point is that when assigning biological phenomena, such as genetic and epigenetic, roles analogous to program and data (or parameters), we should expect to observe transitions of information and function between the two categories.

As we saw, deciding on what is and what is not part of the constituents of the network is not a neutral decision, and reflects conceptual commitments about what constitutes evolutionary change (whether conscious and acknowledged or not). In other words, the evolutionary paradigm determines the network. Work on networks that integrates epigenetic modifications as part of the description of the network itself (rather than as "inputs" or "parameters") is a step in a extended-Synthesis direction, since it eliminates the presupposition that for changes to be evolutionarily significant they have to arise from random genetic mutations. Depending on the network dynamics (and evolutionary dynamics) these epigenetic changes may become genetically assimilated or lead to various canalization responses.[32] Gerhart and Kirschner argue that external stimuli that merely trigger elaborate responses that are self-inhibited, are more easily internalized as genetic.[33] The perspective presented here captures the fact that the self-inhibition may be epigenetic and that the processes in which genetic cues replace external stimuli may be involved in the evolutionary dynamics of cross-organism networks not only in the replacement of environmental cues by genetic ones.

The Evolution of Boundaries

For the most part, work in genetics, and indeed on evolution, operates according to a set of dualities which are at least tacitly assumed as given: development-heredity; plasticity-evolvability; epigenetic-genetic, and assumes a sharp distinction between “inside” and “outside” the spatial and temporal boundaries of the individuals and generations.

No one would deny that these distinctions reflect reality to some extent. But two types of questions are becoming both pressing and approachable: (1) How did these distinctions come to be (i.e., how did they evolve)? (2) How sharp are the boundaries between these phenomena *in present day organisms*?[16]

The examples discussed above showed that these distinctions may not be as sharp as one might assume (in many diverse taxa), and that this assumption distracts attention from frequently occurring phenomena. While sharpening or establishing of the boundaries may have been a factor in major transitions in evolution, *blurring* of boundaries may have also been adaptive (the role of hybridization in evolution can be understood as "blurring" species boundaries, for examples and discussion of the role of hybridization in evolution see Refs. 34-36).

Due to the centrality of the categories just listed for biological research and the apparent prevalence of the distinctions underlying them in the biological world, it is important to clarify the status of these distinctions and their evolutionary history. The prevalence of the boundaries implied by these distinctions in the biological world, coupled with their susceptibility to evolutionary change, and the prevalence of biological processes involved in enacting these distinctions or dependent on them (in many cases as sharp boundaries) suggest that in addition to having emerged early in the evolution of life these boundaries have adaptive advantage and are not merely frozen accidents.

Most of the evolutionary transitions identified by Maynard Smith and Szathmáry[37] involve changes in several of the dimensions (or boundaries) implied by the dualities listed above in addition to possible changes in mechanisms that are co-opted to establish and protect the nascent boundary.

Consider, as an example, the transition from independent genes to chromosomes. Jablonka and Lamb argue that epigenetic mechanisms such as chromatin marking were selected to enable existing states of gene activity to be rapidly re-established after cell division and allow continuous functioning.[38] New possibilities for networks become available (among regulatory regions on chromosomes, rather than gene-gene interaction), which can provide new capabilities, some of them cross-generational

(e.g., transcriptional memory, as discussed above).These networks, that are based on epigenetic mechanisms, may contribute to evolvability as well as plasticity.[14] A new type of "foreign" genetic material (i.e., extra-chromosomal) comes into existence, opening up the potential for the evolution of mechanisms for protecting the integrity of chromosomes. Thus, the genetic-epigenetic dimension was modified, as well as that of plasticity-evolvability.

The transition giving rise to eukaryotic cells exemplifies more types of boundary-related changes. The evolution of organelles from endosymbionts involves the internalization of cross-organism networks, as discussed above. Epigenetic inheritance mechanisms based on 3 dimensional templating were recruited to ensure the correct replication of the symbiont's membranes within the new environment of the host.[38] From the perspective of cross-compartment networks, membranes can be considered internal boundaries (see the discussion of mitochondria above). The mechanisms and regulatory control ensuring membrane inheritance, during development and cell division of the host, thus assume the function of the mechanisms that maintained the boundaries of the endosymbionts making up the cross-organism network. It is significant to note that mitochondria can replicate independently of the eukaryotic cell-cycle, in response to energy needs. In the context of the eukaryotic organism this replication is clearly a developmental response, and thus can be described as development assuming a function reproduction (and thus heredity) had prior to the transition. Prior to endosymbiosis, in the absence of a co-evolutionary cross-organism network perspective, the replication of the mitochondrion-precursor belonged to the realm of heredity and evolution rather than plasticity and development. The cross-organism network perspective emphasizes that the complex interaction underlying the developmental response could have existed prior to endosymbiosis, and allows us to study both in a unified framework. Additionally, as discussed earlier, the cross-organism network perspective refines the description of the extant eukaryotic network.

From the discussion above it should be apparent how an evolutionary perspective stressing cross-boundary interactions – paradigmatically cross-organism interactions – directs attention to significant changes that occurred in major evolutionary transitions. An evolutionary perspective on the boundaries enacted and penetrated problematizes

the theoretical categories listed above which embody a commitment to sharp biological boundaries. It is important to distinguish two complementary ways in which this occurs. First, the same, or evolutionarily continuous, biological phenomena are classified as belonging to different categories such as development and heredity, before and after transitions, and hence studied using different methodologies. Similarly, the notion of epigenetic inheritance changes as new forms of epigenetic inheritance emerge. These issues concern the concepts used in biological discourse. Second, the division of labor between biological mechanisms can change. The same biological mechanisms, it should be noted, may be involved in phenomena on both sides of a boundary, influencing how the boundary can evolve. RNA interference for example, which operates as an ontogenetic developmental epigenetic mechanism, silences transposable elements, which cause genomic (rather than epigenetic) changes. Furthermore, different yet possibly evolutionarily related RNA interference pathways establish the silencing in somatic cells and in the germline.[39] Similarly, epigenetic inheritance plays a role in both developmental plasticity and evolvability.[14] Thus, the very same mechanisms can be involved in phenomena comprising both sides of the duality, just as interaction between cooperating or competing mechanisms can establish both, as well as a continuum between them. Movement along these continuums can happen, as well as specialization (i.e., division of labor), and changes to the relative importance of mechanisms closer to each of the edge cases. Attempting to define the boundary between the two extremes is probably not going to be helpful for studying how this happened in most cases, since this explanation has to be on the level of mechanisms not of the emergent categories.

Conclusions

Delineating biological regulatory networks has conceptual ramifications, and may involve conceptual presuppositions that lead to biases in the sampling of the networks studied from the interwoven hierarchal mélange of networks inside and outside the organism. The assumption that it is possible to distinguish between network architecture and network state oblivious of network dynamics is undermined by the wide phenotypic range and heritability of epigenetic change. The biases discussed in this paper may obscure evolutionary dynamics, which may involve plasticity and evolvability on multiple levels, especially in cases in which boundaries shifted during evolution, or the location of network constituents changed. Moreover, it may lead to

biased generalizations regarding network topologies and other network level properties.

I argued that these biases may be remedied, or at least mitigated, by accepting as paradigmatic biological regulatory networks those networks that potentially cross organism boundaries. Furthermore, the evolutionary analysis of network dynamics should not be based on presuppositions about the types of changes that cross generational boundaries, since epigenetic inheritance and developmentally recreated effects can be functionally equivalent to genetic change and establish structurally equivalent regulatory structures. Changes of the various kinds I discussed can become developmentally canalized, as well as genetically assimilated, and the distinction between network architecture and network state should be used with great caution. The generalized perspective on networks argued for here seems to be required for understanding evolutionary changes that affect the demarcation of central biological categories such as development-heredity; plasticity-evolvability; epigenetic-genetic. These distinctions, as well as other elements of biological discourse rely on the existence of a sharp distinction between "inside" and "outside" the spatial and temporal boundaries of individuals and generations. The universality of these distinctions is explicitly rejected by the generalized network perspective I argued for.

Results indicating that hybridization and alloploidization can cause repeatable large scale genomic changes suggest that the analysis of cross-organism networks should not be restricted to the study of regulatory interaction and development, but should also be integrated into the analysis of the origins of hereditary change and variation, since cross-organism interactions can have direct hereditary effect which may have evolutionary significance and explain large evolutionary changes or "saltations".[14] These effects may complement or enhance the rapid evolution of network regulatory regions which has been observed to follow whole-genome duplication events.[40]

Gerhart and Kirschner define weak regulatory linkage as regulatory linkage in which interaction is mediated and hence does not require stereochemical complementarity and in which the output can be much more complicated than the regulatory input due to the fact that the output is independent of the nature of the signal.[33] Paradigmatically the response in weak regulatory linkage is self-inhibited, and the signal merely

actuates the response. Weak regulatory linkage may help to account for the generation of viable phenotypic variation in general, as well as the replacement of plastic developmental responses to environmental cues by genetic stimulus.[33] The interactions in cross-organism networks can establish weak regulatory linkage, since they are often mediated and may invoke previously selected capacities of the organisms. Interactions between genomes, such as those due to hybridization, which may lead to large scale genomic change as discussed above, probably involve heterochromatin changes and utilize the heterochromatin as platform for recruitment of variety of regulatory proteins,[42] a process which may also lead to weak regulatory linkage in cross-organism networks.

Responses in cross-organism networks may involve organism-level buffering as well as adaptive plastic behavior by the organisms (which may be a factor in determining the identity of the interacting partners and the timing of interaction, as in the case of movement, for example). As a result, cross-organism networks not only provide additional degrees of freedom in the evolution of regulatory networks, they can themselves take part in developmental plastic responses. The heritability of such adaptive variation may depend on developmentally recreated or persistent interactions, rather than genetic or genomic changes in the species.[25] The composition of the network can provide it with enhanced responsiveness, or adaptability, to environmental demands, for example due to the high evolvability of microorganisms.[41] The network response to plastic or hereditary change in the constituent organisms may require changes in other participant in the network. However, organisms may also buffer the rest of the network from changes they incur, for example, by masking the environmental changes that made them necessary. An additional level of buffering against changes incurred by the constituents of the network may be afforded by the organization of the network. Whether the network response to the initial cue is considered a developmental response or a hereditary change depends on further considerations.

To summarize, I argued that the evolutionary paradigm of researchers determines the boundaries of the networks studied, and showed that theoretical biases stemming from Neo-Darwinian geno-centrism and computational metaphors may be overcome by

paying attention to significant properties of empirical biological networks, specifically those of cross-organism and cross-generation networks.

Acknowldegment

I thank Eva Jablonka for comments on earlier drafts of this paper. The work presented in this paper is part of the author's PhD research, which was financially supported by The Cohn Institute For the History and Philosophy of Science and Ideas and The School of Philosophy at Tel Aviv University.

References

1. Shapiro, J. A. 2002. Genome Organization and Reorganization in Evolution: Formatting for Computation and Function. Ann NY Acad Sci. **981**: 111-134.
2. Shapiro, J. A. 2006. Genome Informatics: The Role of DNA in Cellular Computations. Biological Theory. **1**: 288-301.
3. Bechtel, W. 2007. Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience. Lawrence Erlbaum.
4. Wilkins, A. S. 2005. Recasting developmental evolution in terms of genetic pathway and network evolution? and the implications for comparative biology. Brain Research Bulletin. **66**: 495-509.
5. Kellis, M., B. W. Birren & E. S. Lander. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature. **428**: 617-24.
6. Olson, E. N. 2006. Gene Regulatory Networks in the Evolution and Development of the Heart. Science. **313**: 1922-1927.
7. Oliveri, P. & E. H. Davidson. 2007. DEVELOPMENT: Built to Run, Not Fail. Science. **315**: 1510-1511.
8. de Silva, E. & M. P. H. Stumpf. 2005. Complex networks and simple models in biology. Journal of The Royal Society Interface. **2**: 419-430.
9. Keeling, P. J. & J. M. Archibald. 2008. Organelle Evolution: What's in a Name? Current Biology. **18**: R345-R347.
10. Lane, C. E. & J. M. Archibald. 2008. The eukaryotic tree of life: endosymbiosis takes its TOL. Trends in Ecology & Evolution. **23**: 268-275.
11. Bhattacharya, D., *et al.* 2007. How do endosymbionts become organelles? Understanding early events in plastid evolution. BioEssays. **29**: 1239-1246.
12. Kiers, E. T., *et al.* 2003. Host sanctions and the legume-rhizobium mutualism. Nature. **425**: 78-81.
13. Martinez-Zapater, J. M., *et al.* 1992. Mutations at the Arabidopsis CHM Locus Promote Rearrangements of the Mitochondrial Genome. Plant Cell. **4**: 889-899.
14. Lamm, E. & E. Jablonka. 2008. The Nurture of Nature: Hereditary Plasticity in Evolution. Philosophical Psychology. **21**: 305-319.
15. Doolittle, W. F. & E. Bapteste. 2007. Inaugural Article: Pattern pluralism and the Tree of Life hypothesis. Proceedings of the National Academy of Sciences. **104**: 2043.

16. Goldenfeld, N. & C. Woese. 2007. Connections Biology's next revolution. Nature. **445**: 369.
17. Lopez-Maury, L., S. Marguerat & J. Bahler. 2008. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. Nat Rev Genet. **9**: 583-593.
18. Kitano, H. & K. Oda. 2006. Self-Extending Symbiosis: A Mechanism for Increasing Robustness Through Evolution. Biological Theory. **1**: 61-66.
19. Rosenberg, E., *et al.* 2007. The role of microorganisms in coral health, disease and evolution. Nat Rev Micro. **5**: 355-362.
20. Nicholson, J. K., E. Holmes & I. D. Wilson. 2005. Gut microorganisms, mammalian metabolism and personalized health care. Nat Rev Micro. **3**: 431-438.
21. Xu, J. & J. I. Gordon. 2003. Inaugural Article: Honor thy symbionts. Proceedings of the National Academy of Sciences. **100**: 10452-10459.
22. Hooper, L. V., *et al.* 2001. Molecular Analysis of Commensal Host-Microbial Relationships in the Intestine. Science. **291**: 881-884.
23. Nicholson, J. K. & I. D. Wilson. 2003. Understanding 'Global' Systems Biology: Metabonomics and the Continuum of Metabolism. Nat Rev Drug Discov. **2**: 668-676.
24. Nicholson, J. K., *et al.* 2004. The challenges of modeling mammalian biocomplexity. Nat Biotech. **22**: 1268-1274.
25. Jablonka, E. 2004. From Replicators to Heritably Varying Phenotypic Traits: The Extended Phenotype Revisited. Biology and Philosophy. **19**: 353-375.
26. Davidson, E. H., *et al.* 2002. A genomic regulatory network for development. Science (New York, N.Y.). **295**: 1669-78.
27. Zuckerkandl, E. & G. Cavalli. 2007. Combinatorial epigenetics,?junk DNA?, and the evolution of complex organisms. Gene. **390**: 232-242.
28. Ptashne, M. 2008. Transcription: A Mechanism for Short-Term Memory. Current Biology. **18**: R25-R27.
29. Zacharioudakis, I., T. Gligoris & D. Tzamarias. 2007. A yeast catabolic enzyme controls transcriptional memory. Current biology : CB. **17**: 2041-6.
30. Zordan, R. E., D. J. Galgoczy & A. D. Johnson. 2006. Epigenetic properties of white-opaque switching in Candida albicans are based on a self-sustaining transcriptional feedback loop. Proceedings of the National Academy of Sciences of the United States of America. **103**: 12807-12.
31. Griesemer, J. 2000. Development, Culture, and the Units of Inheritance. Philosophy of Science. **67**: S348-S368.
32. West-Eberhard, M. J. 2003. Developmental Plasticity and Evolution. Oxford University Press, USA.
33. Gerhart, J. & M. Kirschner. 2007. The theory of facilitated variation. Proceedings of the National Academy of Sciences of the United States of America. **104 Suppl 1**: 8582-9.
34. Arnold, M. L. 1997. Natural Hybridization and Evolution. Oxford University Press, USA.
35. Arnold, M. L. 2006. Evolution Through Genetic Exchange. Oxford University Press, USA.
36. Pfennig, K. S. 2007. Facultative Mate Choice Drives Adaptive Hybridization. Science. **318**: 965-967.
37. Maynard-Smith, J. & E. Szathmary. 1995. The major transitions in evolution. WH Freeman. Oxford.

38. Jablonka, E. & M. J. Lamb. 2006. The evolution of information in the major transitions. Journal of Theoretical Biology. **239**: 236-246.
39. Chung, W.-J., *et al.* 2008. Endogenous RNA Interference Provides a Somatic Defense against Drosophila Transposons. Current Biology. **18**: 795-802.
40. Gu, X., Z. Zhang & W. Huang. 2005. Rapid evolution of expression and regulatory divergences after yeast gene duplication. Proceedings of the National Academy of Sciences. **102**: 707-712.
41. Zilber-Rosenberg, I. & E. Rosenberg. 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiology Reviews.
42. Grewal, S. I. S. & S. Jia. 2007. Heterochromatin revisited. Nat Rev Genet. **8**: 35-46.

Figure legends:

Figure 1: Possible selection scenarios. All black components share evolutionary interests (so selection is not strictly between the groups A and B). a and b interact, but only if you include organism B in the network. c belongs to the coalition not by directly interacting, but by leading to behavior that makes interaction between a b and their partners in organism B more likely; in this context c is adaptive.