The Topology of the Possible

Walter Fontana

Santa Fe Institute
1399 Hyde Park Road
Santa Fe, NM 87501, USA
walter@santafe.edu

Draft 19th May 2003

1 The problem of change

The classical framework of evolutionary change is based on two aspects of biological organization: phenotype and genotype. The notion of phenotype refers to the morphological, organizational and behavioral expression of an organism during its lifetime. "Phenotype" emphasizes the systemic nature of biological organizations. The notion of genotype refers to a *heritable* repository of information that participates in the production of molecules whose interactions, in conjunction with the environment, generate and maintain the phenotype. This description is circuitous for a reason: A collection of molecular genes all by themselves (the genotype) is pretty useless. That collection must be queried, read and interpreted to become a causal agency. This is done by an endogenous and exogenous context that changes in complex ways as a consequence of its genetic readouts. Phenotype depends on genotype and genotype depends on phenotype – a circle of cause and effect known as *development.* The more biologists unravel the molecular and cellular control systems underlying development and organismal plasticity, the less neat appears the split into phenotype and genotype. These notions are useful for distinguishing the roles of adaptation (through phenotypes) and inheritance (through genotypes) in the evolutionary process. Whether those roles can always be cleanly and uniquely mapped onto separate material carriers is an open issue. I alert to this complexity up front, because I want to continue using the classical notions of phenotype and genotype for expository convenience in the present context.

At its simplest, evolution is driven by the selection of phenotypes, which causes the amplification of their genotypes, and by the production of novel phenotypes through genetic mutations (Figure 1 top). These two factors of evolutionary change are the focal points of distinct research agendas. I refer to them as *evolutionary dynamics* and *evolutionary kinematics*. Evolutionary kinematics needs to be distinguished from evolutionary dynamics in precisely the same way that physics distinguishes kinematics from dynamics. In physics, kinematics is about *how* things move, dynamics is about *why* things move. Dynamics is about forces causing motion. Kinematics is about constraints to motion – constraints that are endogenous to the system that is moving. Such constraints restrict the dimensionality of the configuration space available to a system as well as the geometry of the feasible paths between two configurations. For example, parallel wheels cannot move

sideways. This complicates maneuvering a car into a tight parking lot and generates path-dependency. Kinematics is not just about degrees of freedom, but about controllable degrees of freedom.

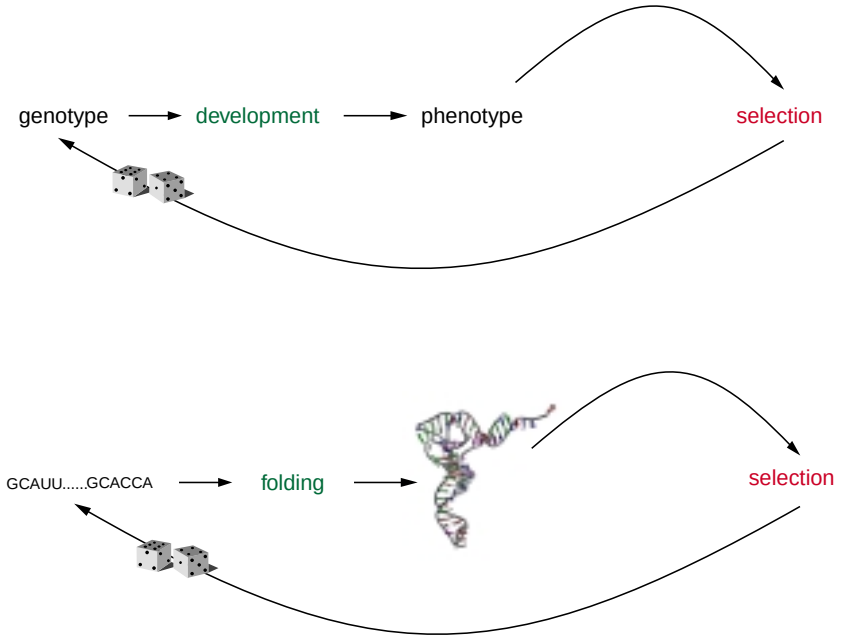

Figure 1: **The folding of RNA sequences into shapes as a proxy of a genotype-phenotype map.** Mutations occur at the genetic level. Their consequences at the phenotypic level are mediated by development, the suite of processes by which phenotype is constructed from genotype. RNA folding is a transparent and tractable model that captures this indirection of innovation within a single molecule. The RNA folding map is characterized by a number of remarkable statistical regularities with profound evolutionary consequences. These regularities may generalize to more complex forms of development.

When thinking about evolutionary motion, one is inevitably drawn to thinking in terms of "trajectories" or "paths" in some "space" in plain analogy to physical motion. In the evolutionary case, the "points" of this "space" are phenotypes, that is, biological organizations or their functional subsystems. The task of evolutionary kinematics is to formalize the vague notions of evolutionary path and evolutionary space by studying how function changes with structure.

Like its counterpart in physics, evolutionary dynamics deals with the causes or "forces" of evolutionary motion, such as selection, drift and Mendelian segregation. Selection is a dynamical phenomenon that arises spontaneously in constrained populations of autocatalytically reproducing entities. The research agenda focussed on selection aims at characterizing the conditions under which a phenotypic innovation can, once generated, invade an existing population. The classical fields of inquiry concerned with selection are population genetics and ecology. The dependent variables are frequencies of genes, or species representatives, whose spatio-temporal change is typically described by a nonlinear dynamical system. This line of research is not concerned with a deep understanding of how change arises in the first place nor with the origin of the internal architecture of biological entities and how that architecture influences its own *capacity* to vary. Rather, evolving entities are assumed as given and their innovation is some stochastic process with simple isotropic characteristics typically motivated by the need for mathematical tractability.

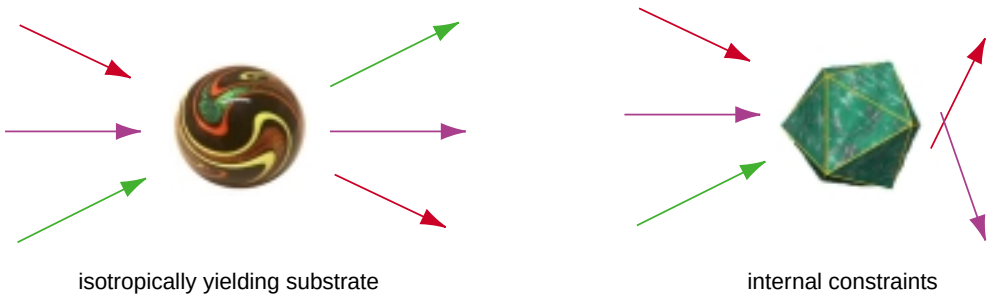

Figure 2: **Evolutionary kinematics and Galton's polyhedron.** In 1889, Sir Francis Galton used a polyhedron (bottom) as a metaphor of organization to suggest how internal structure channels the effects (arrows on the right) of perturbations (arrows on the left) along definite and difficult-to-predict axes. This is in contrast to a marble (top) whose changes in direction and momentum directly reflect the impacting agency. Galton's metaphor is discussed by Gould (2002); Gould and Lewontin (1979).

The kinematics question of how a phenotypic innovation arises in the first place has, so far, been the concern of a different research track. The *heritable* modification of a phenotype usually does not involve a direct intervention at the phenotypic level, but proceeds indirectly through change at the genetic

level (Figure 1). This forces the processes that link genotype to phenotype, aka development, into the picture. Evolutionary trajectories depend on development, because development mediates the phenotypic effects of genetic mutations (Figure 2) and therefore structures the *accessibilty*[1] of one phenotype from another. Aside from constraining and promoting variation, the mechanisms of development are themselves subject to evolution, creating feedback between evolution and development[2].

In the early days of the neo-Darwinian school of evolutionary thought, insufficient knowledge of developmental processes justified ignoring the relationship between genotype and phenotype. This pragmatic approach yielded a powerful conceptual and formal framework that has been exported – with mixed success – to fields outside biology, particularly "evolutionary" flavors of economics and the social sciences [[citations]]. The problem remains that important phenomena of phenotypic evolution do not result naturally from an adaptationist framework that assumes phenotypes to be completely malleable by natural selection. These phenomena comprise the punctuated mode (the partially discontinuous nature) of evolutionary change (Eldredge and Gould, 1972), constraints to variation (Maynard-Smith et al., 1985), the origin of novelties (Müller and Wagner, 1991), directionality in evolution, path-dependency or historicity in organismic structure at all levels and phenotypic stability or homology (for a discussion see Gould (2002)).

The arrival of a new phenotype must necessarily precede its survival or spread in a population through selection. While selection is clearly an important driving force of evolution, the dynamics of selection doesn't teach us much about how evolutionary innovations arise in the first place. To say that a mutation occurred and that it was selected is not very informative. A model of the genotype-phenotype relation (development) is necessary to illuminate how genetic change maps into phenotypic change. In the following I will informally discuss the simplest molecular stand-in of such a relation. It is based on the shapes of RNA sequences (Schuster et al., 1994; Fontana and Schuster, 1998a; Stadler et al., 2001), see Figure 1 bottom. RNA is an extreme case, because genotype (sequence) and phenotype (folded sequence or shape) are different aspects of the same molecule, as illustrated in Figure

[1] I understand accessibility in the sense of a collection of paths along which phenotype A can be transformed into phenotype B.

[2] Insiders refer to the field of biological studies concerned with this feedback as "evo-devo".

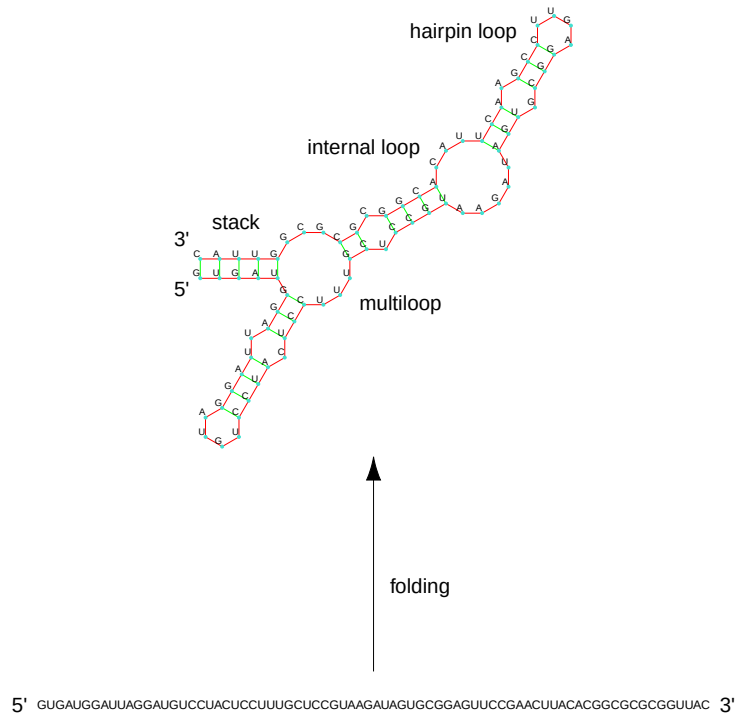

Figure 3: **RNA shape.** At the level of resolution considered here, RNA shape is a graph of contacts between building blocks (nucleotides) at positions $i = 1, \ldots, n$ along the sequence. Position 1 is the 5'-end. The graph has two types of edges: the backbone connecting nucleotide i with nucleotide $i+1$ (red) and hydrogen-bonded base pairings between positions (green).

3. The sequence of an RNA molecule functions as a genotype, since an RNA sequence can be directly copied (or replicated, or inheritably transmitted) by a molecular machinery. In contrast, the shape of an RNA molecule is never copied, but is generated through the folding process of the sequence. The shape conveys functional behavior (phenotype) to an RNA molecule and is a target of selection. The mapping from RNA sequences to shapes constitutes the perhaps simplest biophysically grounded (mini-)example of a mapping from genotypes to phenotypes that is tractable both theoretically and experimentally[3].

[3]It is important to keep in mind the limitations of the model. RNA is a meaningful

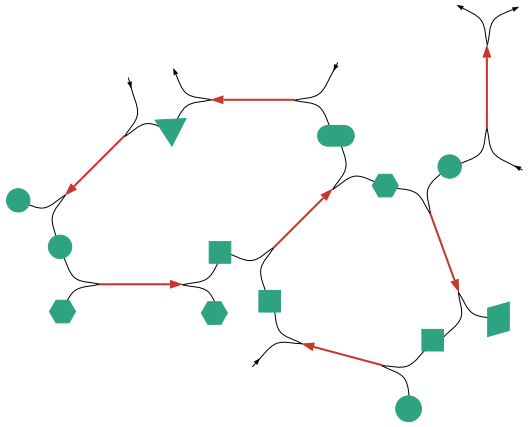

Figure 4: **Organization as a self-maintaining network of production pathways.** Molecules (green shapes) react (red arrows) to generate new molecules that react to generate further molecules. In this way, an initial collection of molecules generates a network of chemical transformations. Molecules are assumed to have a finite lifetime because of degradation or outflow. A reaction can also be viewed in analogy to a logical inference or derivation: the reactants are premises and the products are conclusions which can be premises for further inferences. "Organizations" are those networks that have become self-maintaining.

The RNA model allows us to study a simple but important question: given a "physics" (a mechanism) that generates phenotype from genotype, what are the statistical characteristics of the relation between gentotype and phenotype and how do they affect the dynamics of evolutionary change? In other

cartoon of a genotype-phenotype relation (for reasons that should become apparent in the next section), although it falls far short from even being a caricature of organismal development. The regulatory networks of gene expression and signal transduction that coordinate the production of phenotype (for a recent overview see (Carroll et al., 2001)) have no analogue in the RNA sequence-to-structure map. Furthermore, the molecular processes underlying organismal development are themselves genetically influenced and thus subject to evolution. Major innovations in the history of life are associated with the emergence of new developmental mechanisms and processes. In contrast, the folding of an RNA sequence into a shape is essentially governed by physico-chemical principles that are not controlled by any sequence. This said, I believe that the RNA model is relevant, because it offers perspectives that are potentially generalizable to other, far more complex situations.

words, what can we say statistically about how phenotype changes with genotype and what does this teach us about innovation? After sketching an answer to this question, I will suggest possible connections to the social and economic sciences in section 3.

I will shift gears in section 4 and briefly discuss a quite different model that touches more deeply on the notion of phenotype as "organization" (Figure 4). The model emphasizes organization as a functional closure, more specifically, as a system of component processes with a finite lifetime that act by transforming one another such as to ensure the regeneration of these same processes (Fontana and Buss, 1994a, 1996). The particular level of abstraction of this model has the advantage of providing a clear formalization of "organization" as a specification of those functional relationships among components that ensure self-maintenance of the system. This formalization may offer ways to understand how such relationships constrain and promote change.

2 RNA shape as a systemic phenotype

Figure 3 sketches all one needs to know about RNA folding for the present discussion. At one level, we are given a sequence of fixed length over an alphabet of four letters, {A,U,G,C}. The letters represent certain molecular building blocks. RNA is chemically a very close relative of DNA. As in DNA, the building blocks stick to one another according to specific rules dictated by their shapes: A pairs with U, G pairs with C, and G also pairs with U. We call building blocks that pair with one another "complementary". The difference to DNA is that RNA occurs single-stranded. Rather than pairing up with a second complementary strand, as in the DNA double-helix, a RNA sequence folds up on itself by matching complementary segments along the sequence, as shown in Figure 3. The notion of structure depicted in Figure 3 simply consists in the pattern of pairings between positions along the sequence. The structure formation is driven by changes in (free) energy upon pairing. Stretches of pairs (the "ladders" or stacks in Figure 3) stabilize a structure, while loops destabilize it. Notice that the pairing of two sequence segments necessarily creates a loop. Moreover, in this abstraction, pairings between positions located in different loops are not allowed. As a consequence, the formation of a paired stretch (which is energetically "good")

generates not only a loop (which is energetically "bad"), but it prevents the positions within that loop from pairing with any positions outside of it. A huge number of structures are possible for any given sequence. Yet, one or a few of these structures balance the mentioned trade-offs in an energetically optimal fashion. I shall call the optimal fold of a sequence simply its "shape". This shape can be computed by cleverly designed and fast algorithms (Waterman and Smith, 1978; Nussinov and Jacobson, 1980; Zuker and Stiegler, 1981; Zuker and Sankoff, 1984), based on empirically measured energy parameters (Turner et al., 1988; Jaeger et al., 1989; He et al., 1991; Walter et al., 1994; Mathews et al., 1999).

The real shape of a sequence is a three-dimensional structure (see, for example, the inset at the bottom of Figure 1) that cannot be computed at the time of this writing. Although the shape in Figure 3 is a very crude model, it is not entirely a fiction. The pairings can be established empirically by several methods and it turns out that the three-dimensional structure indeed contains many of the pairings predicted by this abstract notion of shape.

In sum, the pairing rules and their energetics establish a map that assigns a shape to each sequence. The relevant point for the present discussion is that a shape cannot be changed directly, but requires modification of the sequence and the operation of the folding process to obtain the new shape. This mirrors the indirection in transforming phenotypes through genetic mutations mentioned above (Figure 1). Consider that the set of possible sequences of small to moderate length is already hyper-astronomical. For example, there are 4^{100} or 10^{60} sequences of length 100. The issue is to characterize the global *statistical* properties of the mapping from sequences to shapes.

This notion of RNA shape is interesting because it constitutes a simple example of a systemic property. Changing the building block at a sequence position alters many pairing possibilities throughout the sequence, potentially tipping the optimal pairing pattern. The effect at the shape-level of changing a sequence position is highly dependent on the remaining sequence context. Actually, the most consequential fact is that several building block substitutions (mutations) at various sequence positions do *not* affect the shape. In a four letter alphabet, there are three possible substitutions, or mutations, at each sequence position. I call a position *"neutral"*, if it allows for at least one mutation that does not alter the shape of the sequence. Similarly, I call such a mutation a neutral mutation. Although a neutral mutation leaves the

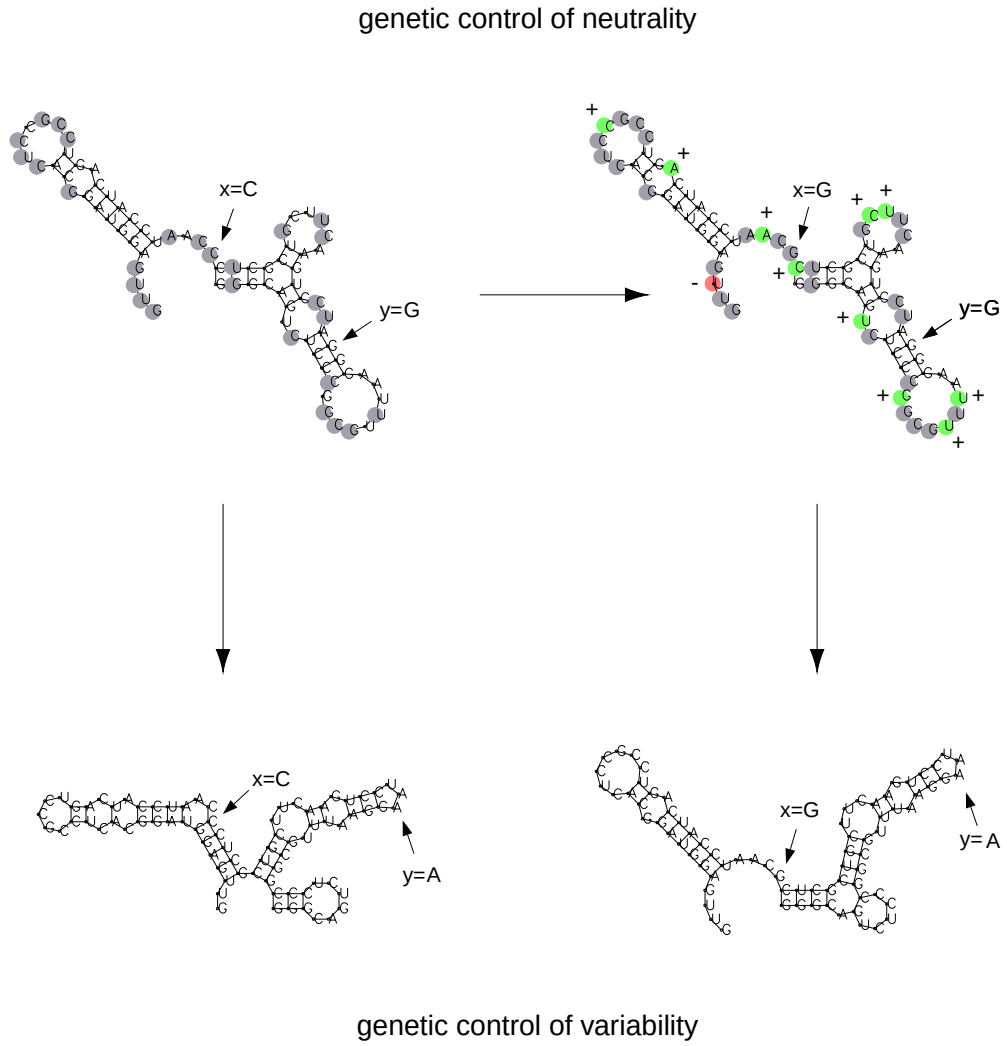

Figure 5: **Epistasis.** Bullets indicate a neutral position. In the top left sequence, position x is neutral because the substitution of G for C preserves the shape, as shown in the top right sequence. Yet, neutral positions *do* change as a consequence of a neutral mutation. The green and red bullets in the top right sequence indicate positions that have gained or lost neutrality, respectively. The lower part illustrates the context sensitivity of mutational effects. The neutral mutation from C to G at x affects the consequences of swapping A for G at the (non-neutral) position y.

shape of a sequence invariant, it has two important contextual effects illustrated in Figure 5. The shape shown at the top left of Figure 5 remains the same if C is substituted by G at the position labelled x. Yet, whether x is C or G determines the shape obtained as a result of mutating the non-neutral position y from G to A (Figure 5, lower half). This dependency of the effect of a mutation at position y on the building block at position x is called *"epistasis"*. Notice another, more subtle, effect: although the replacement of C by G at x does not alter the shape, it alters the number and identity of neutral positions (Figure 5, top right). Thus, while a neutral mutation does not change the phenotype, it does change the potential for change.

The tendency of a sequence to adopt a different shape upon mutation is called *variability*[4]. In the present context, variability is just the opposite of robustness to genetic perturbations. Being able to evolve means being able to access, through genetic mutation, alternative phenotypes with higher selective value. The capacity to evolve (*evolvability*) in response to selective pressures depends on phenotypic variability. Yet, a large amount of theoretical work has been devoted to the motion on fitness landscapes regardless of the constraints imposed by development on the motion through phenotype space. This has put an excessive explanatory burden on fitness considerations.

Figure 5 illustrates that variability (quantified as the number of non-neutral positions) is sequence dependent. Variability can therefore evolve (Wagner and Altenberg, 1996; Wagner, 1996). Note that the change in variability occurs here without the mechanisms of folding (development) themselves changing. This sequence-dependent change of variability is related to Waddington's concept of canalization (Waddington, 1942). [[Explain.]]

2.1 Neutral networks

I will focus on one statistical property of the RNA folding map that is particularly consequential for this discussion.

In what follows, it is useful to think of all possible sequences of a given length

[4]Variability is a capacity and should not be confused with variance, which is the diversity of individuals in a population (Wagner and Altenberg, 1996). Matephorically (*sic*), one could think of variability as the derivative of phenotype with respect to genotype.

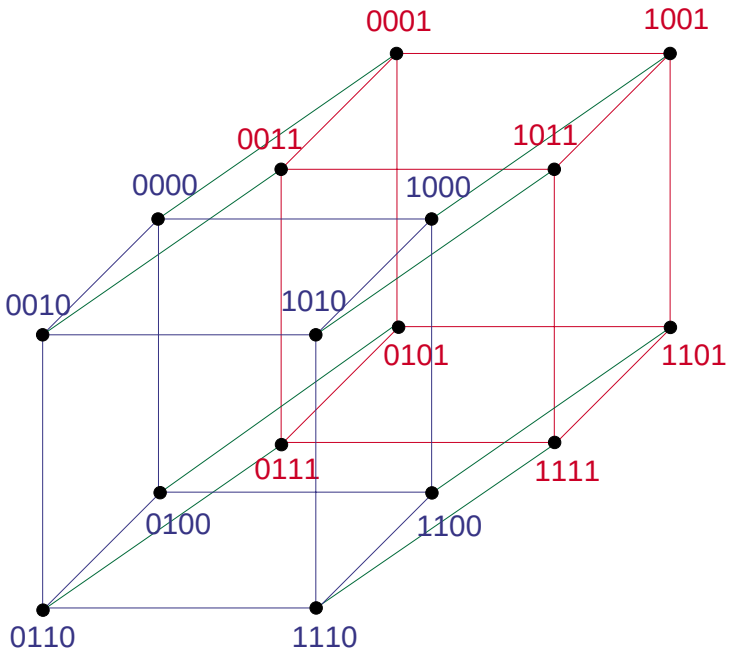

Figure 6: **Sequence space.** An illustration of sequence space for sequences of length 4 over the binary alphabet {0,1}. The connections correspond to single mutations that change one position only. The colors hint at how the 4-dimensional sequence space is constructed from the three-dimensional space (a cube).

as related to one another by a notion of distance. The distance between two sequences is the number of mutations required to convert one sequence into the other. The direct neighbors of a sequence are all sequences one mutation away. This generates a very high-dimensional metric space (Eigen, 1971), since every position constitutes a dimension (Figure 6). Despite its abstract nature, this so-called sequence space is "real" in the sense that mutations are actual physico-chemical events that interconvert sequences and occur with a certain probability. The probability of going from one sequence to another in a single step depends on their distance. The farther apart they are, the lower the probability for a direct jump.

Equipped with this image of a sequence space, consider now the sequence and its shape depicted at the top left of Figure 5. The positions marked by

grey bullets indicate neutral mutations. A sequence has typically a significant fraction of neutral neighbors one and two mutations away. The same, however, holds for these neighbors. (In fact, the sequence shown at the top right of Figure 5 also has several neutral positions marked by bullets.) In this way, we can jump in steps of one or two mutations from sequence to sequence, while preserving the phenotype. This results in an extensive, mutationally connected network of sequences, for which we coined the term *neutral network* (Schuster et al., 1994), see Figure 7A.

2.2 Robustness enables change

The possibility of changing the genotype while preserving the phenotype is both a manifestation of a certain phenotypic robustness to genetic mutations and, at the same time, a key factor underlying evolvability. It seems paradoxical that robustness enables change. To understand this, imagine (with the help of Figure 7A) a population with phenotype "Green" in a situation where phenotype "Blue" would be advantageous or desirable. Phenotype Blue, however, may not be accessible in the vicinity of the population's current location in genotype space, say, somewhere in the northern portion of the Green network of Figure 7A. In the popular image of a rugged fitness landscape, the population would be stuck at a local peak, forever waiting for an exceedingly unlikely event to deliver the right combination of several mutations that yield phenotype Blue. Yet, if phenotype Green has an associated neutral network in genotype space, the population is not stuck, but can drift on that network into far away regions, vastly improving its chances of encountering the neutral network associated with phenotype Blue (Huynen et al., 1996; Huynen, 1996; Fontana and Schuster, 1998a; van Nimwegen and Crutchfield, 2000), see Figure 6A. Neutral networks enable phenotypic innovation by permitting the gradual accumulation of neutral (phenotypically inconsequential) mutations. These mutations, however, alter the genetic context, enabling a subsequent mutation to become phenotypically consequential in an advantageous way. Of course, there is no guidance on a neutral network, but drift dramatically increases the chance of eventually encountering the network of Blue, compared to a direct jump from some isolated Green-location to some far-away Blue-location as would be required in the absence of neutral networks. What appears to be a sudden and abrupt change at the phenotypic level has been the result of neutral genetic drift. The existence

of neutral paths in RNA sequence space was impressively demonstrated in a recent experiment (Schultes and Bartel, 2000).

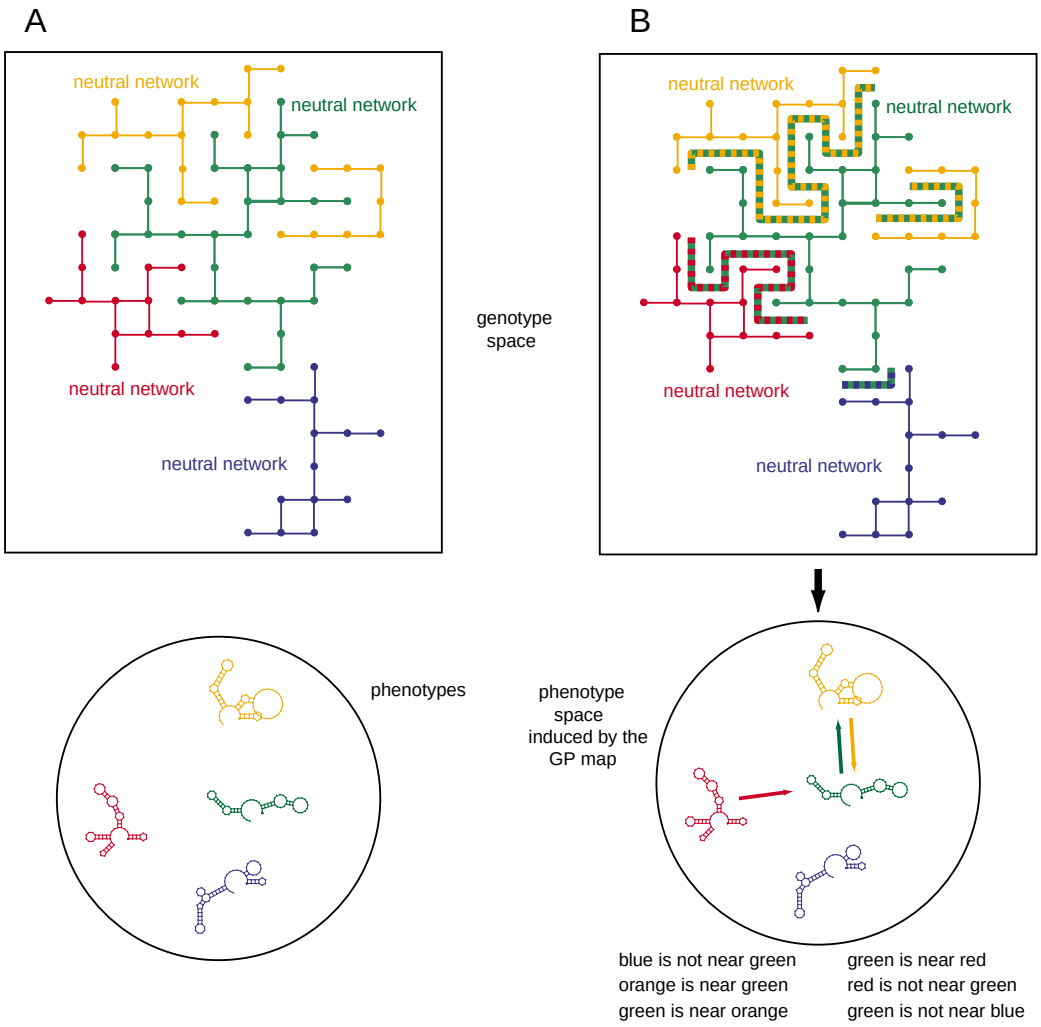

Figure 7: **Neutral networks and shape space topology. (A)** A schematic depiction of neutral networks in sequence space (upper part). The color of a sequence (and hence of a network) indicates its phenotype in the lower part of the figure. A population located in the northern portion of the "green" phenotype cannot access the "blue" phenotype, but it can diffuse on the green network until it encounters the blue network. **(B)** Phenotypic neighborhood as fraction of shared

boundary. The green network is near the red one, because a random step out of red has a high probability of yielding green. The red network, however, is not near the green one, because a random step out of green has a low probability of yielding red. This effect results from very differently sized networks. In another case, the green and blue networks have similar sizes, but they border one another only rarely; green and blue are not in each other's neighborhood. This is the case for shift transformations like the one shown in Figure 8A.

The importance of neutrality for evolution was long recognized by the Japanese population geneticist Motoo Kimura (Kimura, 1968, 1983). Population genetics is mostly concerned with the dynamics of gene (or allele) frequencies. In that context, neutrality is interesting because it leads to a diffusion dynamics of gene frequencies (recall the discussion in section 1). Here I wish to emphasize a different consequence of neutral networks: the organization of phenotype space.

2.3 The topology of phenotype space

A space is formally a set of elements with a structure that derives from relationships among its elements. A relation of distance gives rise to a metric space. As mentioned before, the distance between two RNA sequences is the number of positions in which they differ. For a metric space to be relevant, the distance must be defined in terms of naturally occurring operations that interconvert elements. This is indeed the case with sequences. But what is a space of phenotypes, if phenotypes cannot be modified directly? True, any number of distance (or similarity) measures between phenotypes could be defined and these measures are perfectly useful in establishing selection or sorting criteria. Yet, a phenotype space so-defined is of no help in understanding *evolutionary* histories, because it relates phenotypes without taking into account the indirection required to change them, an indirection that runs through the processes by which phenotypes arise from genotypes. Rather, what is needed is a criterion of *accessibility* of one phenotype from another by means of mutations on their underlying genetic representation. Such a notion of accessibility can then be used to define a concept of *neighborhood* which generates the structure of phenotype space in the absence of a distance notion (Fontana and Schuster, 1998a; Stadler et al., 2001).

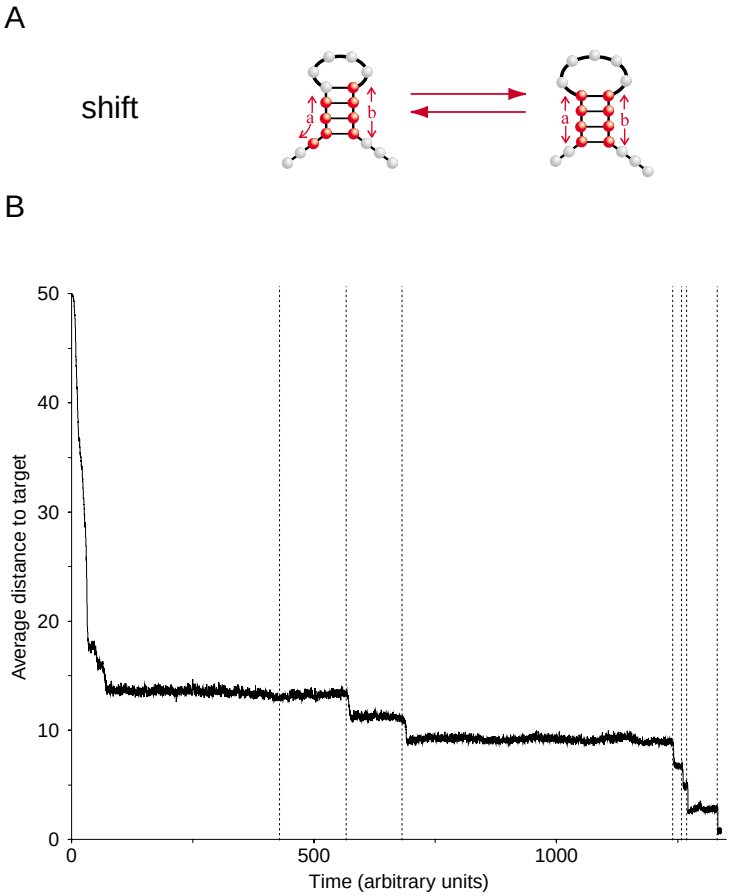

Figure 8: **Punctuated dynamics.** **A**: An example of a discontinuous shape transformation in RNA. In this case - a shift - one strand slides against the other in a paired segment. All pairings must slide in one single event, since any partial sliding would create bubbles, destabilizing the intermediate. Triggering such a shift by a single mutation requires special sequence contexts (Fontana and Schuster, 1998a,b). **B**: Punctuation in evolving RNA populations. A population of RNA sequences evolves under selection for a specific target shape. The homing in on the target shape shows periods of stasis punctuated by sudden improvements. (Fitness is maximal when the distance to the target shape has become zero.) Yet, the phenotypic discontinuities (marker lines), as revealed by an *ex post* reconstruction of the evolutionary trajectory, are not always congruent with the fitness picture. In this example, the first two jumps in fitness turn out to be continuous in the phenotype topology discussed here and a crucial discontinuous transition is fitness neutral (first marker line) (Fontana and Schuster, 1998b; Schuster and Fontana, 1999).

Recall that a neutral network is the set of all genotypes adopting a particular phenotype. In Figure 7B, I define a relation of accessibility between two shapes, say, Green and Red, in terms of the adjacency of their corresponding neutral networks in genotype space (Fontana and Schuster, 1998a,b). The *boundary* of a neutral network consists of all sequences that are one mutation off the network. The intersection of the neutral network of Red with the boundary of the neutral network of Green, relative to the total boundary of Green's network, is a measure of the probability that one step off a random point on the neutral network of Green there is a sequence folding into Red. A rough schematic of such intersections is depicted in Figure 7B as two-color thick lines that mark the apposition of the networks of the corresponding colors.

To fix ideas with a cartoon, think of Europe as sequence space and of a neutral network as a European state (a "phenotype"), say, France. Now scan the boundary of France and measure the fraction of boundary it has in common with some other state, say Germany. That fraction measures the likelihood of ending up in Germany when making a random step out of France. Doing this for all states bordering France yields a distribution of relative border lengths. This distribution has an abstract average. Now define the neighbors of France as those states whose relative boundary with France is longer than that average. This construction is needed in actual genotype space because a neutral network is a very high-dimensional object that borders to a huge number of other neutral networks (the example of geographic states fails miserably in this regard). Most of these boundaries, however, are tiny. (In fact, the actual distribution of relative boundary sizes is a generalized power-law (Fontana and Schuster, 1998b).) In the case of sequence space, this construction of neighborhood is mathematically clean, but I shall not pursue technical details here (for a discussion see Stadler et al. (2001)). Notice something funny, though. France is near Monaco, because a large fraction of Monaco's boundary is with France. Yet, Monaco is not near France, since France's boundary to Monaco is tiny. Returning to our original picture, a phenotype may be easily accessible from another phenotype by genetic mutations, but the reverse may not be true. I will not provide RNA examples here, since their discussion requires too much terminology that is of little import to the target readership of this paper. The important message is that accessibility is *not* a distance. A distance relation is always a symmetric relation, while accessibility is not as illustrated in Figure 7B.

The RNA phenotype space is not a metric space, but a much weaker and less intuitive space, known as a "pretopology" (Stadler et al., 2001).

2.4 Continuous and discontinuous change

One more step is needed to put together a full picture. Obviously, Austria is not near France, because Austria has a zero border to France. We can ask, however, whether there is a route from France to Austria such that at each step we always make a transition to a state in the current neighborhood - for these are the easy transitions. We could, for example, step from France into Germany (easy) and from Germany into Austria (easy). Roughly speaking, mathematicians call a path continuous if at any given small step it proceeds along the neighborhood relations of the space being considered; in other words, if you don't have to lift the pencil when drawing the path...

We can now ask an important question about the structure of phenotype space: Given any two phenotypes, is there always a continuous ("easy") path connecting them? That is, can we go from one neutral network to another that is not a neighbor (in the boundary sense defined above), such that at each step (mutation) along the way we always transit to a neighboring network, until we reach our target network? In the RNA case, the answer is no (Fontana and Schuster, 1998a). This means that certain shape transformations are irreducibly difficult to achieve, that is, they are discontinuous ("difficult") and there is no way around that difficulty. These transformations are fairly local changes in shape – nothing fancy – but they require the simultaneous rearrangement of several interactions (pairings) among building blocks, because any intermediate resulting from a serial rearrangement would be unstable or outright impossible (Figure 8A shows an example). To change three or four interactions at once by a small mutation is possible, but requires specially poised, hence rare, sequences and this makes the transition irreducibly difficult (that is, hard to find). I skip an exposition of the RNA shape transformations that fall into this category (but see Figure 8A for a quick example).

A few observations deserve emphasis. Analyzing the mapping from genotypes to phenotypes in terms of neutral networks enables a mathematically rigorous definition of continuity (or discontinuity) in evolutionary trajectories. First, this notion of (dis)continuity cross-cuts morphological (dis)similarity.

Some transitions between similar shapes are discontinuous (e.g. Figure 8A) and some transitions between dissimilar shapes are continuous. Second, the notion of discontinuity defined here is *not* related to sudden jumps in fitness or the discreteness of a change. The categories of discontinuity are caused by the genotype-phenotype map and thus remain the same regardless of the further evaluation of phenotypes in terms of fitness assignments. (Of course, the particular shapes observed at discontinuous transitions will depend on the fitness map, but not the nature of these transformations.) Third, the dynamical signature of this phenotype topology is punctuation (see Figure 8B). A population of replicating and mutating sequences under selection drifts on the neutral network of the currently best shape until it encounters a gateway to a network that conveys some advantage or is fitness-neutral. That encounter, however, is evidently not under the control of selection, for selection cannot distinguish between neutral sequences. While similar to the phenomenon of punctuated equilibrium recognized by Eldredge and Gould (1972) in the fossil record of species evolution, punctuation in evolving RNA populations occurs in the *absence* of externalities (such as meteorite impact or abrupt climate change in the species case), since it reflects the variational properties of the underlying developmental architecture (here: folding).

2.5 Summary

I summarize the main messages, implications and possible extensions.

Development matters (but it must be redundant). The processes that link genotype to phenotype typically generate a redundant mapping, that is, there are considerably fewer phenotypes than genotypes. Clearly, this depends on the level of resolution at which we define "phenotype". If, in the RNA case, we were to define shape as "atomic coordinates", there could be no redundancy; every genotype would have a unique phenotype and the topic of this paper is mute. I rather believe that it is precisely the onset of extensive redundancy that *defines* a meaningful level of resolution for phenotype.

Neutrality enables change. The mathematical and statistical analysis of the mapping from RNA sequences to shapes has produced the concept of a neutral network, that is, a mutationally connected set of genotypes that map to the same phenotype. I believe that this concept, or some variant of it,

holds for many genotype-phenotype mappings beyond the simple RNA case. Neutral networks are key to change. It is hard to understand how evolution could be successful at all, if developmental processes don't give rise to neutral networks. The popular image of a rugged evolutionary landscape, where populations get stuck at local optima, is certainly incomplete and perhaps even wrong. Populations may be pinned at the phenotypic level, but they constantly change at the genetic level, drifting on neutral networks, thereby dramatically increasing their chances for phenotypic innovation. [[Develop.]]

Variations versus innovations - the limits of selection. A neutral network has many neighbors (in the boundary sense defined above). A transition from a given network to any of its neighbors corresponds to the acquisition of an "easy" new phenotype. We have seen, however, that certain phenotypic changes *must* go through border-bottlenecks, involving a transition from one network to another that touches the former only very rarely. We have distinguished mathematically between such phenotypic changes in terms of continuous and discontinuous changes. I'd like to belabor that point by suggesting that the continuous/discontinuous distinction reflects an intuitive distinction between *variation* that occurs readily (a better mouse trap) and a true *innovation*, that is, a new phenotype that is not readily achievable and that requires long periods of drift (Wagner, 2001). The former variation is typically already present in any reproducing population that drifts over a neutral network and explores its boundary by mutation. Such omni-accessible variation is therefore easily available to selection for adaptive responses. But in the case of an innovation, selection can do nothing to coach a population over a neutral network in the direction of the rare boundary segment that marks an innovation, since selection cannot distinguish between genetic variants of the same phenotype. Thus, innovations are not under the control of selection, although their fate is. David Krakauer (personal communication) has argued that the definition of innovation should take into account how consequential a new phenotype is for subsequent evolution and not be limited to how hard it is to achieve under given developmental mechanisms. A further investigation of this issue may require models in which the genotype-phenotype map itself evolves.

The limits of what is knowable. Suppose that what I called innovations (discontinuous changes) actually correspond to what biologists identify as morphological or functional innovations in organisms. (It is an empirical question whether this assumption is correct.) In that case, it might

be impossible to experimentally demonstrate which genetic change caused the innovation (Wagner, 2001). Experiments can only be done with species that are alive today, and there is no guarantee that any recent species has a genotype poised to replicate the evolutionary transition in question. Recall (by means of Figure 5) that the consequences of a genetic change depend on the genetic context. The same change in a different context will have a different effect. If a species has drifted on a neutral network away from the tiny boundary region associated with the innovation, its genotype may no longer permit to prove that a genetic difference suspected to have caused the innovation actually did cause it. At this stage, this is mere speculation.

3 Beyond genotype and phenotype

I have described a paradigm of change in a rather specific molecular and biological setting. Underneath the surface, the notions of genotype and phenotype appear as convenient idealizations even in biology (Griesemer, 2000). Needless to say, they are not translatable to technological, economic and social realms. I'd like, therefore, to conclude the RNA case with a little extension for readers from other disciplines.

In social, economic and technological domains, we rely on systems consisting of many heterogenously interacting components whose collective action gives rise to ordered system behavior. When we wish to change that behavior, we hit on a simple fact whose consequences I have described in this contribution: Behavior is not a thing, behavior is the property of a thing. It follows that the only way to change behavior is to change the thing that generates it. The change of a property is necessarily indirect.

As an example, consider a computer program. A computer program implements a function. A function is not a thing, so you can't alter it directly. To alter the function you must alter the program text. The mapping from program to function is what computer scientists call the semantics of the programming language in which the program is written. Substitute for computer program your favorite social/technological/economic organization, for function the relevant qualitative behavior of that organization and for semantics the (microscopic) dynamical principles that govern the interactions among the parts of the organization.

At a high level of abstraction, the structure of the situation may be represented by a mapping from systems within a given class – computer programs, molecules, electronic circuit diagrams, urban transportation systems, firms in a given industry – to behaviors of those systems. An input to this mapping is the specification or description of a particular system configuration. The mapping generates the behavior or function of that system configuration. The mapping arises from the dynamical principles, the rules and processes that are *constitutive* for a given class of systems. In social contexts, institutions would certainly be a component ingredient of this mapping. The unfolding of this constitutive dynamics in the context of a particular system configuration yields the behavior associated with that configuration. (Example: the rules of the game of Go are constitutive of Go in that Go doesn't exist without these rules. As such they are an ingredient of the mapping from a board configuration to its, say, strategic value.)

When we wish to change behaviors of systems, we often have a spatial metaphor in mind, such as going from "here to there", where "here" and "there" are positions in the space of behaviors. But what exactly is the nature of this space? Who brought it to the party? It is a popular fallacy to assume that the *space* of behaviors is there to begin with. This is a fallacy even when all possible behaviors are known in advance. How does this fallacy arise? When we are given a set of entities of any kind, we almost always can cook up a way of comparing two such entities, thereby producing a definition of similarity (or distance). A measure of similarity makes those entities hang together naturally in a familiar metric space. The fallacy is to believe that a so-constructed space is real. It isn't, because that measure of similarity is not based on available real-world operations, since we cannot act on behaviors directly. We only have system-editors, we don't have property-editors. Seen from this operational angle, that which structures the space of behaviors is not the degree of similarity among behaviors but a rather different relation: *accessibility* of one behavior from another in terms of system-reconfigurations. This brings the mapping from system configurations to behaviors into the picture. The structure of behavior-space is induced by this mapping. It cannot exist independently of it.

In the previous sections I have described how this space-structure arises in the context of RNA molecules. With the above paragraphs in mind, substitute "system" for "RNA sequence" and "behavior" for "RNA shape". The RNA model is, however, a limiting (and highly idealized) case in which the mapping

from sequences (system configurations) to shapes (behaviors) is exogenous. In Go, the rules are not determined by the board configuration; they are given independently. While the limiting model is useful to illustrate the induced character of behavior-space, the most intriguing cases are those in which the mapping is itself endogenously generated by the system. This is the case in all of organismal biology, where genes also code for products that are responsible for developmental processes. The endogeneity of the mapping from system configurations to behaviors is certainly a feature of economic and social systems as well. Nothing prevents the rules of the game (and thus the game itself) from changing. It may be harder to change a rule than to change a game configuration, but this is a question of time scales, not of principle. In the case of an endogenous mapping, the structure of phenotype (behavior) space remains an induced one. However, the concept of phenotype space becomes significantly more subtle. Not only can a change in the mapping rearrange the accessibility structure between phenotypes, it can also bring into existence phenotypes that were not possible before. We lack a formal grip on this radical form of change. In the mid sixties, C. H. Waddington expressed his frustration with the mathematical apparatus of theoretical biology: *"The whole real guts of evolution - which is, how do you come to have horses and tigers, and things - is outside the mathematical theory."* (Quoted by Gould (2002, p. 584).) The challenge of formalizing the emergence of new classes of biological objects remains.

4 Constructive systems

Complex biological phenotypes rest on networks of chemical reactions in which molecules transform one another potentially generating molecules not yet present in the system. The intriguing aspect of chemistry lies in its generative power. Imagine seeding a test tube with a solution of four chemical substances. Their reactions generate new substances which can be used for further reactions producing further substances. A few initial molecular species hold the key to a large, potentially infinite, set of molecular species. However, this large set of different molecules is initially available only as a possibility. The molecules in this set are not causally effective until they are made. The mathematical apparatus of dynamical systems is not well equipped to deal with such a situation, since it requires *a priori* knowledge of all variables and their couplings (Fontana and Buss, 1996). There is a

difference between chemical *kinetics*, which describes concentration changes in terms of a nonlinear dynamical system, and chemistry proper, which describes the generative interrelations among molecules. I'd like to point out an analogy between a chemical system and logic. A chemical reaction has the flavor of an implication or rule of inference, where reactants and products function like premises and conclusions, respectively. Many functional systems, far removed from logic proper, such as technologies, cognition, or even semiotics, are based on some sort of consequence relations among their elements. The problem, of course, is that we lack a theory of these relations.

Several years ago, biologist Leo Buss and I proposed a very crude and abstract model capturing the generative aspect (and only that) of chemistry as an attempt to define and formalize the notion of "molecular organization" (Fontana and Buss, 1994a,b, 1996). In mathematics, a theory of transformations is provided by λ-calculus (Church, 1932, 1941). This calculus consists of a notation and rules to express functions that act on the very expressions representing them. Imagine a container filled with such functions that meet one another randomly. When two functions f and g meet, one is applied to the other, thereby generating a new function $h = f(g)$. This is the constructive part and it basically represents *modus ponens*: together, f and g imply h. In addition, each function has a finite lifetime after which it disappears (it is "forgotten", "destroyed", "removed" - the appropriate metaphor depends on what one takes these functions to stand for: molecules, information, knowledge, ...). A function f is treated like a "particle", meaning that the system may contain many copies of function f. In chemical parlance, a function f has a concentration or abundance. In a more general setting, the concentration of f may be interpreted as its "relevance". The concentrations of f and g determine the likelihood of them meeting and interacting in a well-stirred soup.

Consider now the dynamics of such a "gas" of functions. In the early stages, the gas is highly constructive, in the sense that almost every interaction produces a function that wasn't in the gas before. Now pick a function h in this gas. For h to persist, it must be the product of some interaction, say, between f and g, otherwise h will eventually disappear due to its finite lifetime. However, if f and g are not themselves maintained in the gas, the production pathway for h will disappear and h along with it. The maintenance of h therefore requires the maintenance of f and g which, in turn, requires the maintenance of further functions needed to produce f and g.

This can be resolved only if the web of dependencies loops back on itself in a constructive version of what is known as feedback, that is, when the construction processes induced by the functions in the gas permit the continuous regeneration of these same functions. This is known as (collective) catalysis (Maturana and Varela, 1980; Farmer et al., 1982; Kauffman, 1993). The earliest reference to this notion of feedback that I could find is in Kant's Critique of Judgement: "an organized product of nature is one in which all is end and, reciprocally, means too".

By extensive simulations of this model, we arrived at the following characterization of a self-maintaining function gas.

1. Once the gas has become self-maintaining, the expressions representing its constituent functions exhibit syntactical patterns. These patterns characterize all expressions maintained in the system. In fact, they define a grammar, that is, lawful arrangements of identifiable subexpressions (components). The grammar is invariant with respect to interactions, meaning that new functions produced from interactions within a self-maintaining gas conform to its grammar.

2. A few laws characterize all relationships of transformation among the functions allowed by the grammar of a particular self-maintaining gas. These laws constitute an abstract algebra. Grammar and algebraic laws provide a complete description of the system, independent of the original calculus in which the functions were specified and expressed.

The most salient feature of this model is the existence of a higher-order description that replaces the microscopic list of all components and their interactions, once the gas has become self-maintaining. This higher-order description identifies a self-maintaining function-gas as an object in its own right. We called such an invariant entity an organization. An organization is held together by specific invariant relationships of transformation among its components. The components turn over all the time, while grammar and algebra persist.

The connection with algebra allows to clarify some noteworthy behaviors of such organizations with respect to change at their component level. Self-maintaining organizations repair themselves when components are removed.

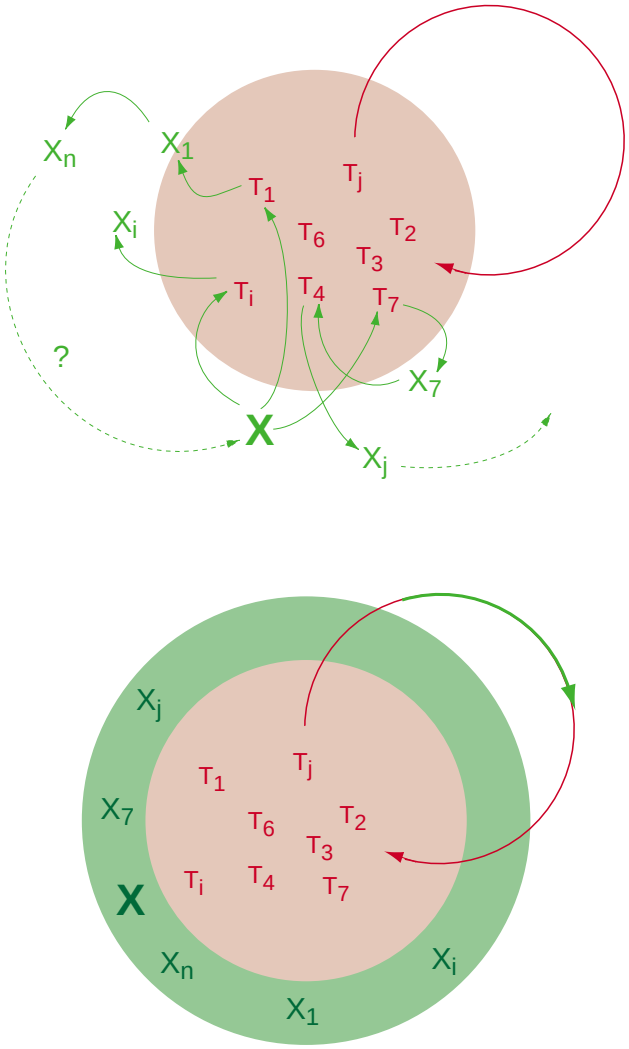

Figure 9: **The extension of a self-maintaining organization.** A self-maintaining organization is schematically represented by the red set containing "red" components. The "red" organization is perturbed by a "green" component X, spawning a trail of consequences X_i. If that trail gives rise to a pathway that loops back to reproduce the original perturbing agent X, the "red" organization is extended in a self-maintaining fashion by a "green" layer (bottom).

This robustness is an immediate consequence of self-maintenance. Constructive feedback provides component homeostasis by regenerating a missing component through interactions among components located upstream in its production pathway. The interesting questions concern the diversity of minimal component sets, or "generator sets", that guarantee the regeneration of an organization in the event of component loss.

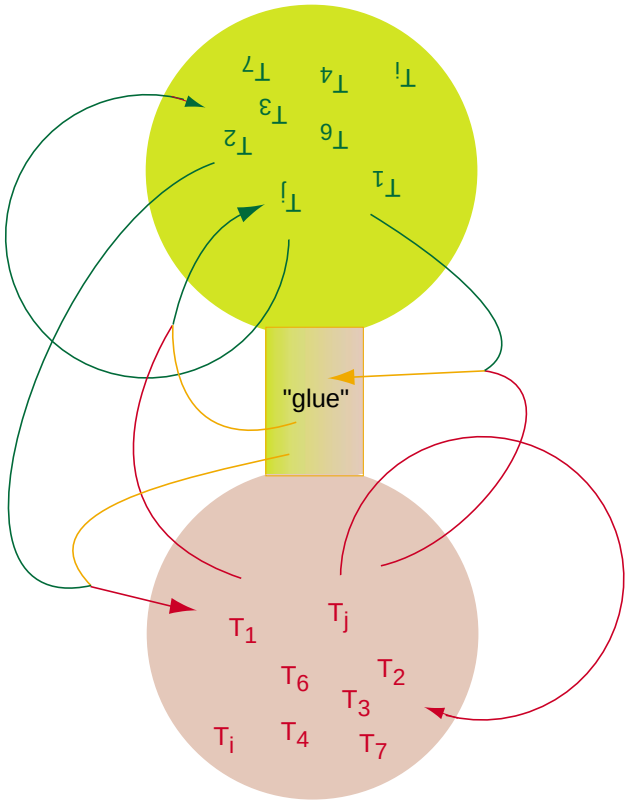

Figure 10: **Merger of self-maintaining organizations into a higher-order unit.** In this cartoon of an organizational merger, two autonomous organizations are "glued" together within a higher-order organization through the products generated by their "cross-talk" (center). If one of the component organizations disappears, so does the glue.

The situation is different with respect to the addition of a function q that does not belong to an organization. The function q generates a cloud of new functions from its interactions with the organization as well as interactions

among the newly generated functions and between them and the organization (for a schematic, see Figure 9). For q to be persistently integrated into the organization requires, as before, a constructive feedback loop in which some of the consequences caused by q eventually re-produce q. Otherwise q (and its trail of consequences) will disappear. The existence of such feedback loops between perturbing agent and organization is constrained by the organization. Note that if integration is successful, it is not just q that has been added, but an entire layer of new functions required to maintain q in the organization (Figure 9).

Rather than perturbing an organization with a single function, we can add a whole other organization (Figure 10). In some cases the merged organizations don't drive each other out of existence, but integrate stably. As in the prior case, successful integration is dependent on feedback loops. Interactions between components of different organizations produce new functions that belong to neither organization. These functions and their consequences establish a "glue" that integrates both self-maintaining organizations into a higher order unit, within which they continue to persist as autonomous entities (Figure 10).

Understanding the "laws of change" governing constructive systems founded on consequence relations is a wide open problem in chemistry, biology, economics and the social and cognitive sciences. Understanding a constructive system requires a theory linking the structure of its constituent agents to their potential actions (like the theory of computation does for a certain class of mathematical functions). The challenge is to exploit a theory of components, once available, to characterize the possible self-maintaining networks that these components can sustain and the changes these networks undergo in response to component perturbations. The challenge is to state the possible and its topology.

Acknowledgements

My research is supported by Mr. Jim Rutt's Proteus Foundation and by core grants to the Santa Fe Institute from the John D. and Catherine T. MacArthur Foundation, the National Science Foundation, and the U.S. Department of Energy, and by gifts and grants from individuals, corporations, other foundations, and members of the Institute's Business Network for Complex Systems

Research. I'd like to thank Woody Powell and John Padgett for many useful comments.

References

S B Carroll, J K Grenier, and S D Weatherbee. *From DNA to diversity.* Blackwell Science, Malden, Massachusetts, 2001.

A. Church. A set of postulates for the foundation of logic. *Annals of Math. (2)*, 33:346–366, 1932.

A. Church. *The Calculi of Lambda Conversion.* Princeton University Press, Princeton, 1941.

M. Eigen. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58:465–523, 1971.

N. Eldredge and S. J. Gould. Punctuated equilibria: an alternative to phyletic gradualism. In T. J. M. Schopf, editor, *Models in Paleobiology*, pages 82–115. Freeman, Cooper & Co., San Francisco, CA, 1972.

J. D. Farmer, S. A. Kauffman, and N. H. Packard. Autocatalytic replication of polymers. *Physica D*, 22:50–67, 1982.

W. Fontana and L. W. Buss. 'The arrival of the fittest': Toward a theory of biological organization. *Bull. Math. Biol.*, 56:1–64, 1994a.

W. Fontana and L. W. Buss. What would be conserved 'if the tape were played twice'. *Proc. Natl. Acad. Sci. USA*, 91:757–761, 1994b.

W. Fontana and L. W. Buss. The barrier of objects: From dynamical systems to bounded organizations. In J. Casti and A. Karlqvist, editors, *Barriers and Boundaries*, pages 56–116. Addison-Wesley, Reading, MA, 1996.

W. Fontana and P. Schuster. Continuity in evolution: On the nature of transitions. *Science*, 280:1451–1455, 1998a.

W. Fontana and P. Schuster. Shaping space: The possible and the attainable in RNA genotype-phenotype mapping. *J. Theor. Biol.*, 194:491–515, 1998b.

S. J. Gould. *The Structure of Evolutionary Theory*. Belknap/Harvard University Press, Cambridge, MA, 2002.

S. J. Gould and R. Lewontin. *Proc. R. Soc. Lond B*, 205:581–598, 1979.

J. R. Griesemer. Reproduction and the reduction of genetics. In P. Beurton, R. Falk, and H-J. Rheinberger, editors, *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*, pages 240–285. Cambridge University Press, Cambridge, MA, 2000.

L. He, R. Kierzek, J. SantaLucia, A.E. Walter, and D.H. Turner. Nearest-neighbour parameters for G-U mismatches. *Biochemistry*, 30:11124, 1991.

M. A. Huynen. Exploring phenotype space through neutral evolution. *J. Mol. Evol.*, 43:165–169, 1996.

M. A. Huynen, P. F.Stadler, and W. Fontana. Smoothness within ruggedness: The role of neutrality in adaptation. *Proc. Natl. Acad. Sci. USA*, 93:397–401, 1996.

J. A. Jaeger, D. H. Turner, and M. Zuker. Improved predictions of secondary structures for RNA. *Proc. Natl. Acad. Sci. USA*, 86:7706–7710, 1989.

S. A. Kauffman. *The origins of order*. Oxford University Press, New York., 1993.

M. Kimura. Evolutionary rate at the molecular level. *Nature*, 217:624–626, 1968.

M. Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK, 1983.

D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, 288:911–940, 1999.

H. Maturana and F. J. Varela. *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel, Boston, 1980.

J. Maynard-Smith, R. Burian, S. A. Kauffman, P. Alberch, J. Campbell, B. Goodwin, R. Lande, D. Raup, and L. Wolpert. Developmental constraints and evolution. *Quart. Rev. Biol.*, 60:265–287, 1985.

G. B. Müller and G. P. Wagner. Novelty in evolution: Restructuring the concept. *Annu. Rev. Ecol. Syst.*, 22:229–256, 1991.

R. Nussinov and A. B. Jacobson. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc. Natl. Acad. Sci. USA*, 77(11): 6309–6313, 1980.

E. A. Schultes and D. P. Bartel. One sequence, two ribozymes: Implications for the emergence of new ribozyme folds. *Science*, 289:448–452, 2000.

P. Schuster and W. Fontana. Chance and necessity in evolution: Lessons from RNA. *Physica D*, 133:427–452, 1999.

P. Schuster, W. Fontana, P. F. Stadler, and I. Hofacker. From sequences to shapes and back: A case study in RNA secondary structures. *Proc. Roy. Soc. (London) B*, 255:279–284, 1994.

B. M. R. Stadler, P. F. Stadler, G. Wagner, and W. Fontana. The topology of the possible: Formal spaces underlying patterns of evolutionary change. *J. Theor. Biol.*, 213:241–274, 2001.

D. H. Turner, N. Sugimoto, and S. Freier. RNA structure prediction. *Annual Review of Biophysics and Biophysical Chemistry*, 17:167–192, 1988.

E. van Nimwegen and J. P. Crutchfield. Metastable evolutionary dynamics: Crossing fitness barriers or escaping via neutral paths? *Bull. Math. Biol.*, 62:799–848, 2000.

C. H. Waddington. Canalization of development and the inheritance of acquired characters. *Nature*, 3811:563–565, 1942.

A. Wagner. Does evolutionary plasticity evolve? *Evolution*, 50(3:1008–1023, 1996.

G. P. Wagner. What is the promise of developmental evolution? Part II: A causal explanation of evolutionary innovations may be impossible. *Journal of Experimental Zoology (Mol Devol Evol)*, 291:305–309, 2001.

G. P. Wagner and L. Altenberg. Complex adaptations and the evolution of evolvability. *Evolution*, 50:967–976, 1996.

A. E. Walter, D. H. Turner, J. Kim, M. H. Lyttle, P. Muller, D. H. Mathews, and M. Zuker. Coaxial stacking of helices enhances binding of oligoribonucleotides and improves prediction of RNA folding. *Proc. Natl. Acad. Sci.*, 91:9218 – 9222, 1994.

M. S. Waterman and T. F. Smith. RNA secondary structure: A complete mathematical analysis. *Mathematical Biosciences*, 42:257–266, 1978.

M. Zuker and D. Sankoff. RNA secondary structures and their prediction. *Bull.Math.Biol.*, 46(4):591–621, 1984.

M. Zuker and P. Stiegler. Optimal computer folding of larger RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9:133–148, 1981.