J. theor. Biol. (2001) **213,** 241–274
doi:10.1006/jtbi.2001.2423, available online at http://www.idealibrary.com on IDEAL®

The Topology of the Possible: Formal Spaces Underlying Patterns of Evolutionary Change

Bärbel M. R. Stadler*, Peter F. Stadler*†, Günter P. Wagner‡ and Walter Fontana†§‖

*Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien, Währingerstraße 17, A-1090 Wien, Austria, †Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, U.S.A. ‡Ecology and Evolutionary Biology, Yale University, New Haven, CT, U.S.A. and §Institute for Advanced Study, Princeton, NJ 08540, U.S.A.

(*Received on* 19 *December* 2000, *Accepted in revised form on* 2 *August* 2001)

The current implementation of the Neo-Darwinian model of evolution typically assumes that the set of possible phenotypes is organized into a highly symmetric and regular space equipped with a notion of distance, for example, a Euclidean vector space. Recent computational work on a biophysical genotype–phenotype model based on the folding of RNA sequences into secondary structures suggests a rather different picture. If phenotypes are organized according to genetic accessibility, the resulting space lacks a metric and is formalized by an unfamiliar structure, known as a pre-topology. Patterns of phenotypic evolution—such as punctuation, irreversibility, modularity—result naturally from the properties of this space. The classical framework, however, addresses these patterns by exclusively invoking natural selection on suitably imposed fitness landscapes. We propose to extend the explanatory level for phenotypic evolution from fitness considerations alone to include the topological structure of phenotype space as induced by the genotype–phenotype map. We introduce the mathematical concepts and tools necessary to formalize the notion of accessibility pre-topology relative to which we can speak of continuity in the genotype–phenotype map and in evolutionary trajectories. We connect the factorization of a pre-topology into a product space with the notion of phenotypic character and derive a condition for factorization. Based on anecdotal evidence from the RNA model, we conjecture that this condition is not globally fulfilled, but rather confined to regions where the genotype–phenotype map is continuous. Equivalently, local regions of genotype space on which the map is discontinuous are associated with the loss of character autonomy. This is consistent with the importance of these regions for phenotypic innovation. The intention of the present paper is to offer a perspective, a framework to implement this perspective, and a few results illustrating how this framework can be put to work. The RNA case is used as an example throughout the text.

1. Introduction

The Neo-Darwinian model views evolutionary change as resulting from the spontaneous generation of genetic variation and the fixation of variants in the population through natural selection and genetic drift. It provides a useful framework for studying the evolution of phenotypic adaptation, the evolution of gene sequences and the process of speciation; for recent overviews, see Futuyma (1998) and Graur & Li (2000). Yet,

‖Author to whom correspondence should be addressed. Walter Fontana, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, U.S.A.

0022–5193/01/220241 + 34 $35.00/0

many important evolutionary phenomena do not result naturally from the current implementation of the Neo-Darwinian model. These phenomena comprise patterns and processes of phenotypic evolution (Schlichting & Pigliucci, 1998), such as the punctuated mode (the partially discontinuous nature) of evolutionary change (Eldredge & Gould, 1972), developmental constraints or constraints to variation (Maynard-Smith *et al*., 1985; Schwenk, 1995), innovation (Müller & Wagner, 1991), directionality in evolution and phenotypic stability or homology. Many of these issues were debated extensively in the last two decades, but their relationship to the mechanistic theory of evolutionary change, as represented in population genetics, remains unclear and tense.

Before selection can determine the fate of a new phenotype, that phenotype must first be produced or "accessed" by means of variational mechanisms. Phenotypes are not varied directly in a heritable fashion, but through genetic mutation and its consequences on development. We shall take development fairly broadly and refer to it as the genotype–phenotype map (Lewontin, 1974; Wagner & Altenberg, 1996; Fontana & Schuster, 1998a). The evolutionary accessibility of new phenotypes depends on this map, since it determines how phenotypes vary with genotypes. Its structure therefore bears on how a biological system evolves. In the early days of population genetics, insufficient knowledge justified ignoring the relationship between genotype and phenotype. This pragmatic approach has resulted in the habit of representing the accessibility of phenotypic and genetic states by means of metric spaces or even stronger structures, such as the Euclidean vector space of quantitative genetics or the Hamming graph of possible haplotypes in population genetics. This habit has become a deeply embedded assumption in the mathematical structure of classical population genetic theory, yielding models in which biological organization at the phenotypic and genetic level is extremely fluid. The phenomena cited above suggest that this fluidity is largely a fiction and point at profound asymmetries in the accessibility of phenotypic and genetic states.

We argue here for the need of a mathematical theory of evolution based on spaces that are less structured than metric spaces. The motivation for this apparently simple step comes from studies in which RNA folding from sequences to secondary structures is used as a biophysically realistic, yet extremely simplified toy-model of a genotype–phenotype map (Fontana & Schuster, 1998a, b). These studies show that the space derived from organizing the set of possible RNA shapes (phenotypes) in terms of mutational accessibility exhibits a weak and rather unfamiliar structure, a so-called pre-topology, as explained in Section 4. That topology provides a natural framework for understanding punctuated change, directionality and modularity in simulated populations of evolving RNA molecules.

The classical way of addressing these phenomena under the assumption of highly symmetric phenotype or genotype spaces consists in resorting to "fitness landscapes" conveniently constructed to yield the right asymmetries. If not reflected upon, this practice eventually becomes the claim that these phenomena are caused by the structure of the fitness landscape in conjunction with natural selection. In contrast, we argue here that the asymmetries underlying these phenomena can be rooted in the structure of the genotype–phenotype map itself, and thus can be logically prior to fitness assignments. This shift has two consequences. It grounds patterns of phenotypic evolution in biophysical principles and mechanisms rather than arbitrary and convenient assumptions about fitness. It provides a far more natural mathematical setting in which to address these patterns.

The present work offers, in essence, but a perspective. In conjunction with Cupal *et al.* (2000), it connects the intuitions underlying Fontana & Schuster (1998a) with the proper mathematical structures and vocabulary. Our goal is three-fold. First, we argue that many of the recalcitrant phenomena in evolutionary biology, like punctuated innovation, developmental constraints, homology and irreversibility, are but statements about the accessibility topology of phenotype space. Second, we review in a rigorous, yet hopefully accessible fashion the main results of the mathematical theory of pre-topological spaces to a degree that we understand them as relevant to our present concerns. We then extend and apply these instruments, illustrating the concepts by means of the RNA case. Third, we suggest a few

directions of how this abstract framework might be utilized to model phenotypic evolution.

2. Accessibility Structures in Biology

2.1. METRIC SPACES

Accessibility structures, frequently called configuration spaces, are an important conceptual construct in evolutionary biology, computer science and physics which often deal with combinatorial objects, such as genetic sequences, network routings or spin systems. One typically considers the collection of all possible objects (configurations) in that class together with a suite of "variation operators" representing processes which transform one object into another. In genetics, such operators may represent various types of mutation, like base pair substitution or recombination. In computer science, the operator may be more abstract, such as the permutation of the itinerary of a traveling salesman. In physics, it may be the flip of a spin. Variation operators define neighborhoods by establishing which objects are accessible from which other objects. For instance, the nearest neighbors of a DNA sequence with respect to point mutations consist of all one-error mutants of that sequence.

In many cases, the variational operators support a natural notion of "distance" which permits upgrading the notion of a set to that of a "metric space". A distance measure, or metric, is formally a mapping d from pairs of elements of a set X to the positive real numbers, $d: X \times X \to \mathbb{R}_0^+$, satisfying four axioms for all $x, y, z \in X$:

(D0) $d(x, x) = 0$.
(D1) $d(x, z) \leqslant d(x, y) + d(y, z)$.
(D2) if $d(x, y) = d(y, x) = 0 \quad$ then $x = y$.
(D3) $d(x, y) = d(y, x)$.

A well-known example of a metric space is the set of all binary strings of fixed length n that can be interconverted by point mutations alone. Connecting each sequence with its n immediate neighbors yields the hypercube as a graph. The hypercube is a highly regular topological space where distance is the number of positions in which two sequences differ (Hamming distance). This distance is an appropriate measure of genetic accessibility between sequences.

Metric accessibility topologies have far-reaching consequences for evolutionary dynamics. Every element can be reached from any other element by a series of mutations and the variational operator (e.g. point mutation) does not bias the production of variants. Accessing element y from x is as easy (or difficult) as accessing x from y. This same symmetry is oftentimes assumed to also hold for the effects of mutations on the phenotype. In that case, selection becomes the only process that can give a direction to evolution. The problem, however, is that phenotypic variation may well be biased even in the absence of any variational bias at the genetic level. Concepts like developmental constraint and homology express this fact. Since they conflict with the assumption of a metric phenotype (and/or genotype) space, these concepts are difficult to integrate with the existing mathematical framework.

2.2. NON-METRIC SPACES

The notion of distance allows an intuitive construction of the notion of "neighborhood" in terms of "small distance". The notion of distance is so familiar that one is easily fooled into believing that it precedes the concept of "neighborhood". Yet, neighborhood is the weaker and more primitive concept. To work with spaces that support a notion of neighborhood but not of distance runs against common sense. Some examples may soothe the pain.

2.2.1. *RNA Shape Space*

In RNA, both genotype (polymer sequence) and phenotype (polymer structure) are properties of a single molecule. The folding of RNA sequences into secondary structures¶ (henceforth

¶Let i, j, k, l denote positions of bases in the linear sequence and (i, j) a base pair. The secondary structure of an RNA sequence is defined as the set P of allowed base pairs (here Watson–Crick pairs plus GU) which minimize free energy, subject to a no-knot condition requiring that if (i, j) and (k, l) are both in P, then $i < k < j$ implies $i < l < j$ (i.e. base pairs do not cross). The secondary structure is computed with an implementation (Hofacker *et al.*, 1994) of a dynamic programming algorithm (Nussinov & Jacobson, 1980; Waterman, 1978; Zuker & Stiegler, 1981) widely used in laboratories to assist in the prediction of secondary structures. The procedure is based on empirical energy parameters (Turner *et al.*, 1998; Walter *et al.*, 1994).

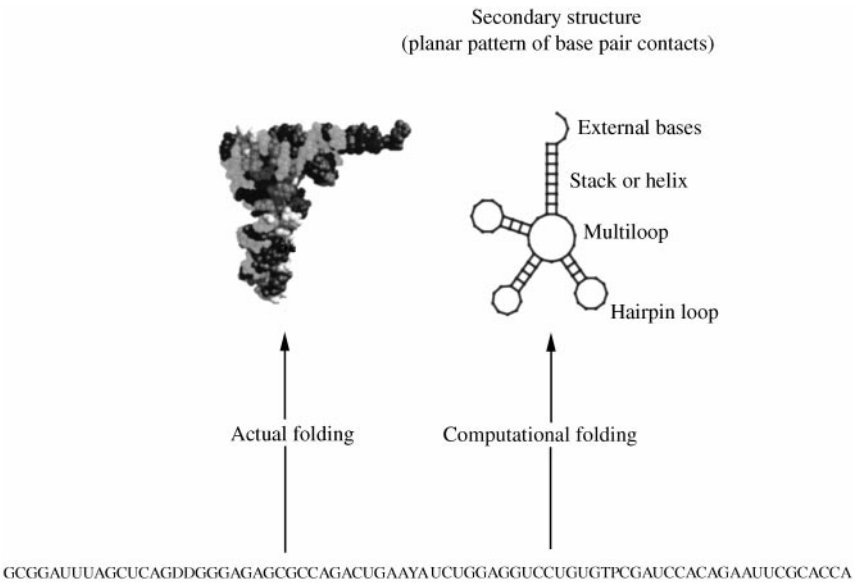


FIG. 1. RNA folding. A secondary structure (right-hand side) is a coarse-grained description of the three-dimensional shape (left-hand side) of an RNA molecule. The secondary structure does not refer to spatial coordinates, but only to the planar topology of base pair contacts. It can be viewed as a graph consisting of structural elements called cycles or loops: a hairpin loop occurs when one base pair encloses a number of unpaired positions, a stack consists of two base pairs with no unpaired positions, while an interior loop has two base pairs enclosing unpaired positions. An internal loop is called a bulge, if either side has no unpaired positions. Finally, multiloops are loops delimited by more than two base pairs. A position that does not belong to any loop type is called external, such as free ends or joints. Despite its abstract quality, the secondary structure is not a fictitious entity. It represents a crucial folding stage on the path towards the tertiary structure of an RNA molecule.

shapes) (Fig. 1), inspires a simple biophysically grounded genotype–phenotype map that is computationally and experimentally tractable. Simulated populations of replicating and mutating RHA sequences under selection exhibit many phenomena known from organismal evolution: neutral drift, punctuated change, plasticity, environmental and genetic canalization, and the emergence of modularity. Laboratory experiments have also generated phenomena consistent with these patterns (Spiegelman, 1971; Lenski & Travisano, 1994; Szostak & Ellington, 1993; Schultes & Bartel, 2000). The RNA model can therefore illuminate the extent to which these patterns of phenotypic evolution are rooted in statistical regularities of the genotype–phenotype map.

It is an important fact about RNA folding that the shapes realized by sequences of fixed length n do not occur with the same frequency. Only a tiny fraction of shapes is "typical", in the sense of being realized significantly more often than others.** As a consequence, (simulated) evolutionary histories exhibit statistical regularities that can be understood in terms of the statistical properties of typical shapes.

We single out one such statistical feature that is of special interest in the present context. Many sequences have the same (typical) shape α as their minimum free energy structure. We call such sequences "neutral" (in the sense of "equivalent") with respect to α. A structure α therefore identifies an equivalence class of sequences. A one-error mutant of a sequence that shares the same

**More precisely, as sequence length goes to infinity, the fraction of such typical shapes tends to zero (their number grows nevertheless exponentially), while the fraction of sequences folding into them tends to one. Consider a numerical example: In the space of GC-only sequences of length $n = 30$, 1.07×10^9 sequences fold into 218 820 shapes. 22 718 shapes (10.4%) are typical in the sense of being formed more frequently than the average number of sequences per shape. 93.4% of all sequences fold into these 10.4% shapes (Grüner *et al.*, 1996a, b; Schuster, 1997).

minimum free energy structure as that sequence is called a "neutral neighbor". By "neutrality" of a sequence we mean the fraction of its $3n$ one-error mutants that are neutral. (Again, the term neutrality refers here to the phenotype—the minimum free energy structure—of RNA sequences, and should not be confused with fitness-based neutrality). Any given sequence folding into a typical shape has a significant fraction of neutral neighbors, and the same holds for these neighbors. In this way, jumping from neighbor to neighbor, we can map an extensive mutationally connected network of sequences that fold into the same minimum free energy structure (Schuster *et al*., 1994; Reidys *et al*., 1997). Such networks were termed "neutral networks" (Schuster *et al*., 1994). The possibility of changing a sequence while preserving the phenotype is a key factor underlying evolvability. The evolutionary role of neutrality has for the most part been viewed conservatively as buffering the phenotypic effects of mutations. Yet, neutrality critically enables phenotypic change by permitting phenotypically silent mutations to set the context for subsequent mutations to become phenotypically consequential. Stated differently, neutrality shapes the accessibility structure of phenotype space.

To observe this, let us first ask what is meant by phenotype *space* in the case of RNA. At the outset we are given a *set*, not a space, of possible shapes (on sequences of length n). To turn this set into a space, we must define relationships of nearness between shapes. One approach would be to simply define a distance measure between shapes based on a morphological comparison and then derive a notion of neighborhood in terms of small distance. This would assume a metric space of shapes to begin with. The problem with this procedure is that it does not reflect *evolutionary* accessibility among shapes, because the variational operators underlying the definition of shape distance do not correspond to physical events or natural processes. In evolution, a shape is modified through mutations in the underlying sequence, rather than by direct modification of the shape, and the phenotypic effect of a mutation is determined by the folding map. An evolutionarily meaningful relation of nearness between shapes must be mediated by the folding map and not be independent of it. The interesting case arises when the genotype–phenotype map is many-to-one, as it is in RNA. A robust notion of nearness among two shapes then must reflect the mutual adjacency of the corresponding neutral networks as determined in the mutational neighborhood structure of genotype space (Fontana & Schuster, 1998a, b) (see Fig. 2).

More precisely, the nearness of shape β to shape α should correlate with the likelihood of a transition from α to β through, say, a single point mutation. In the simplest case, this likelihood will be given by the fraction of boundary shared by the neutral genotype sets of β and α relative to the total boundary of the neutral set of α. Let us write $S(\alpha)$ for the set of all sequences folding into α, and $\partial S(\alpha)$ for the set of all sequences obtained by one point mutation from sequences in $S(\alpha)$. $\partial S(\alpha)$ is the boundary of $S(\alpha)$ in sequence space. For any two structures α and β, $S(\beta) \cap \partial S(\alpha)$ describes all those sequences folding into β which are neighbors of sequences folding

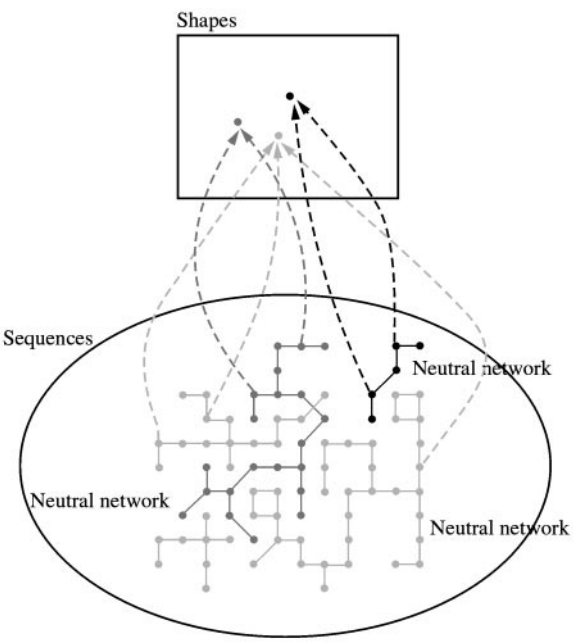


FIG. 2. Accessibility topology of shape space. In this schematic representation of the map from genotypes (sequences) to phenotypes (shapes), nearness of phenotype *green*: (light grey) to phenotype *red*: (medium grey) is determined by the size of the joint boundary between the red and green neutral networks relative to the size of the red network, $A(green \curvearrowleft red)$. In this picture, a random step off the red network is likely to end on the green network. Hence, phenotype *green* is near *red*. However, a random step off green is unlikely to end in red. Hence, *red* is not near *green*.

into α. The accessibility of β from α, $A(\beta \curvearrowleft \alpha)$, now becomes the frequency ratio

$$A(\beta \curvearrowleft \alpha) = \frac{|S(\beta) \cap \partial S(\alpha)|}{|\partial S(\alpha)|}, \qquad (1)$$

where $|X|$ stands for the number of elements (cardinality) of set X.

Note that $A(\beta \curvearrowleft \alpha)$ is *not* a distance measure and a so-organized shape space is not a metric space. In general, accessibility lacks symmetry: $A(\beta \curvearrowleft \alpha) \neq A(\alpha \curvearrowleft \beta)$, because the neutral sets (and hence the boundaries) of α and β can vastly differ in size. To pin down ideas with a cartoon, suppose we organize the United States of America in terms of accessibility based on relative shared boundary size. In this topology, Pennsylvania is near New Jersey—a random step out of New Jersey is likely to end up in Pennsylvania—but New Jersey is not near Pennsylvania—a random step out of Pennsylvania is unlikely to end up in New Jersey. Consider, for example, an RNA structure β that differs from α by the presence of a small stacking region. The formation of a stacking region implies the formation of an energetically costly loop. (To make a stack, the RNA sequence must bend back on itself, thereby creating a constrained loop region). A stack cannot be initiated with just one isolated base pair because that base pair cannot offset the destabilization resulting from the loop it creates. A minimum of three contiguous base pairs is required on average to balance the cost of a small hairpin loop. This is a thermodynamic all or none situation. Triggering a transition from α to β (creating a small stack) requires, therefore, specially poised sequences with the potential of establishing three contiguous base pairs in a single point mutation. Such sequences can be found by neutral drift on the genotype network of α, but they constitute only a tiny fraction of α's neutral genotype set. Yet, they make up all the boundary that α shares with β. Thus, β is hard to access from α or, in topological language, β is not near α.†† Consider now a sequence randomly picked on the neutral genotype network of β. While the stack in question is present, it is unlikely to be energetically well stabilized, since this would require special rather than random sequences on β's network. A point mutation that destroys any base pair of a marginally stable stack will therefore cause the entire stack to unwind, ending up with shape α. It becomes clear, then, that many sequences of network β border network α. In other words, α is easy to access from β. Thus, α is near β, but β is not near α.

Sequences that fold into β and that are located in the boundary of α's neutral network are oftentimes compatible with both structures;‡‡ a suitable mutation then tips the energy balance in favor of β. In RNA, the size of the intersection between the set of sequences compatible with α and the set of those compatible with β is a good predictor for the accessibility $A(\beta \curvearrowleft \alpha)$ (Weber, 1997). Further properties of a neutral network of a structure α, such as its degree of locality in sequence space and its density within the corresponding network of compatible sequences, are related to the average thermodynamic properties of α on the network and can to some extent be read off the structure itself (Göbel *et al.*, 1997).

In the topological sense of the word, an evolutionary path is "continuous" at a point in time at which a phenotypic change occurs, if the triggering mutation is between neighboring genotypes (in the topology of genotype space) and the new phenotype is near the old one (in the topology of phenotype space). Absent a notion of distance, a neighborhood structure in phenotype space has to be defined. We just argued informally that the appropriate neighborhood structure is the one induced by the genotype–phenotype map which determines the likelihoods for converting one phenotype into another by the application of a genetic operation. This is to be distinguished from an approach in which continuity is defined through the presence or absence of some discrete character, or even by a mere "jump" in fitness. RNA secondary structures are discrete objects to begin with and so is their change. What

††Just how hard is "not near"? A meaningful cutoff point must be defined, but we deliberately gloss over this question here. Details are found in Fontana & Schuster (1998b) and we shall briefly return to the issue in Section 5.2.

‡‡A sequence is compatible with a structure, if it is able to fold into that structure, although the latter may not be (and typically is not) the minimum free energy structure.

determines continuity is not the degree to which a modification is incremental, but the degree to which that modification is easy to achieve by virtue of the mechanisms underlying the genotype–phenotype relation. In our picture, a phenotypic change is discontinuous, if it constitutes a "jump" from a developmental perspective (fitness, for that matter, may not change at all), that is, if it is realizable by a continuous genetic change for only a small fraction of genotypes. At least in RNA, the discontinuity and the magnitude of change are not always congruent. The stack example illustrates how morphologically large changes (absence of a stack) can be continuous. Examples given in Fontana & Schuster (1998b) show how morphologically small changes, such as a simple shift between opposing strands of a stacking region, can be discontinuous.

Finally, the asymmetry of phenotype space can cause evolutionary change that is directional in the absence of directional selection. The size difference between a large and a small neutral network acts as a ratchet for a drift-induced discontinuous phenotypic transition that leaves fitness the same. If a phenotype α is near β, but β is not near α, a fitness-neutral transition from β to α is difficult to revert since the entry point will be rapidly lost by drift on the large α-network.

RNA shape space, if organized (statistically) in terms of shape variability through genetic mutation, is not a metric space. In Section 5, we shall see that its structure is even weaker than a topology.

2.2.2. *Subspecialization of Duplicated Genes—the DDC Model*

After a duplication, two copies of a gene can undergo different evolutionary fates. One copy may lose its function through a destructive mutation, becoming a pseudo-gene (Walsh, 1995). In this case, the functional situation of the genome reverts to the state preceding the duplication. In another scenario, one gene acquires a new function, while the other maintains the original one (Ohno, 1970). Finally, both genes may each specialize to a subset of the functions of the ancestral gene. There is an emerging consensus that the most frequent mode of evolution after gene duplication is functional subspecialization (Hughes, 1994). The causes of subspecialization, however, are unclear. One model assumes that the ancestral gene represents a compromise between the multiple functions it carries out, and that disruptive selection after duplication will drive each copy to optimize a subset of the ancestral functions (Hughes, 1994). An alternative model, the DDC model (Force *et al.*, 1999), explains subspecialization by variational biases in phenotype space similar to the ratchet-driven directional change in RNA secondary structure described above.

DDC stands for duplication, degeneration and complementation. The model considers a gene that is expressed in a variety of domains (organ tissues) where it participates in different developmental functions. Each expression domain is assumed to be regulated by a different set of modular enhancer domains. An enhancer domain is a short stretch of non-coding DNA that binds transcription factors which influence the expression of the gene. Enhancers often are modular, that is, for each expression domain, there is a physically and functionally distinct enhancer directing the expression in the corresponding domain. Some enhancers are phylogenetically highly conserved, and therefore seem to be tightly constrained. A mutation is likely to destroy the function of such an enhancer. As in the RNA case, a first asymmetry arises because a non-functional sequence is "near" an enhancer, but no enhancer is "near" a non-functional sequence. Thus, enhancers of duplicated genes will tend to degenerate. Since gene function is redundant after duplication, any degeneration of one enhancer will be phenotypically neutral as long as the other enhancer is maintained. The deleterious mutation will simply be complemented by the enhancer(s) of the duplicated gene. The degeneration of redundant enhancers will continue until either one gene has lost all its enhancers while the other copy has retained them, or until a complementary set of enhancers remains among the two genes. In the former case, one gene becomes a pseudo-gene. The latter case, however, enables the evolution of subspecialization (through mutations in the coding regions) which will be maintained as long as the functions served by each copy are required for survival and

reproduction. Examples consistent with this model are the expression of engrailed (*eng*) (Force *et al.*, 1999) and distal less (*Dlx*) (Quint *et al.*, 2000) paralogues in zebrafish. Since there are many more combinations of complementary enhancer sets enabling subspecialization than causing the loss of a gene, there is a strong bias towards evolving subspecialization (provided the ancestral gene has more than two expression domains). The model does not assume that subspecialization is favored by natural selection; it only assumes that mutations which eliminate an expression domain (a developmental function) from a gene are selectively neutral because of complementation and that the total loss of a function is selected against.

The DDC model is an elegant and genetically plausible model of how directionality can be the outcome of an evolutionary process without directional selection. Like in the previous RNA example, the main reason for this directional bias resides in the mutational accessibility structure of the phenotypic (functional) states involved.

2.2.3. *Unequal Crossover*

Asymmetric accessibility structures are not limited to phenotypic states, but can arise at the genetic level as well. Accessibility structures induced by homologous recombination (crossover at corresponding regions within chromosomes or sequences of fixed length) are topologically equivalent to the metric spaces induced by point mutations (Gitchoff & Wagner, 1996; Stadler & Wagner, 1998; Stadler *et al.*, 2000). The situation, however, differs with unequal crossover (where chromosomes are misaligned and the number of genes on a chromosome can change). Shpak & Wagner (2000) suggest that the genotype space induced by a model of unequal crossover is not metric. The problem here is again a lack of symmetry. Of course, distance measures on this genotype space can be defined, but any such measure would not reflect the accessibility structure induced by unequal crossover. This is analogous to the RNA case where any number of morphological similarity measures between shapes can be defined—but they do not reflect the mutational accessibility induced by the folding map.

3. Evolutionary Patterns and Phenotypic Accessibility

3.1. PUNCTUATED EQUILIBRIA

The term punctuated equilibrium was introduced to describe a pattern of phenotypic evolution inferred from the fossil record (Eldredge & Gould, 1972) in which a lineage spends a large amount of time in a state of stasis, that is, of no directional change, and then suddenly undergoes a phenotypic transition. A variety of mechanisms, ranging from sudden changes in the environment to speciation events that break up the homeostasis of the genotype (Maynard-Smith, 1983) can generate this pattern. It is worth noting, however, that some well-documented examples of punctuation, like the fossil record of *Olenus*, a trilobite, are character specific rather than involving the whole phenotype. This runs against the idea that punctuation is caused only by some general factor like the breakdown of genetic homeostasis during speciation (Wagner, 1989b). Computational models of RNA secondary structure evolution (Huynen *et al.*, 1996; Fontana & Schuster, 1998a) also show a pattern of punctuation. The population drifts on a neutral network in genotype space while maintaining the same phenotype α, until it encounters the neutral network of a new advantageous phenotype β. If β is not near α in the topological sense sketched previously (Section 2.2.1), a (finite) population will spend a long time drifting on the network of α. In the RNA model, punctuation correlates with a discontinuous phenotypic transition. Recall, however, that our definition of discontinuity does not hinge on "suddenness"; the phenomenology of "long periods of stasis ending in phenotypic change" is but a population dynamic manifestation of the topological structure of phenotype space induced by the genotype–phenotype map, and does not require exogenous events. It is therefore tempting to speculate that some of the punctuation events seen in phenotypic evolution are discontinuous phenotypic transitions in some appropriate developmental sense. At the same time, the gradual transitions typical for the Neo-Darwinian model of evolution correspond to continuous evolutionary trajectories connecting nearby phenotypes.

3.2. DEVELOPMENTAL CONSTRAINTS

Accessibility directly relates to the notion of developmental constraints which emphasize the limitations to phenotypic variation realizable in the neighborhood of a genotype. Turning a snail into a horse in a single step is not just discontinuous, but an impossible operation if "step" means a continuous genetic change, that is, one that remains in the neighborhood of a given genotype. There may, however, exist continuous paths in phenotype space that connect a snail with a horse. The existence (or absence) of such paths is a statement about the accessibility structure of phenotype space. The issue is important, because such paths are likely to show up as definite evolutionary trajectories.

We mention two well-documented examples of developmental constraints, because there is still some confusion about the existence of such constraints. The best understood example concerns patterns of phalanx reduction which is highly regular in amniotes and frogs (eutetrapods) and differs from the patterns in newts and salamanders (urodeles). This is caused by developmental differences in hand/foot development between urodeles and eutetrapods (Shubin & Alberch, 1986), as shown experimentally by Alberch & Gale (1983, 1985) in two landmark papers. The logic of the argument is that the last digit to develop is the first to be lost. In eutetrapods, the sequence of digit development is 4-3-2-(5)-1 with some limited variation in the timing of digit "5" development. Consequently, the first digit to be lost is digit "1" in frogs and amniotes. On the other hand, the developmental sequence in urodeles is (1,2)-3-4 in the forelimb and (1,2)-3-4-5 in the hind limb. Consequently, the first digits to be lost are "4" and "5" in the forelimb and hind limb, respectively. This pattern can be reproduced experimentally by decreasing the number of cells available for digit development and is thus not driven by natural selection. What is driven by natural selection, of course, is whether there is a loss at all. Another example of a developmental constraint underlies the fundamental difference between the endoskeleton of higher ray finned fish (teleosts) and that of fleshy finned fish (Wagner, 1999), including tetrapods. The endoskeleton of the former consists in four radials arranged along the anterior–posterior extension of the fin basis, while the endoskeleton of the latter is a complicated pattern of bones derived from a branching arrangement of skeletal analgen (Shubin & Alberch, 1986). The radials of the teleost paired-fin endoskeleton are developmentally derived from a cartilaginous disc that arises early in ontogeny and is later divided in two steps into four rods that ossify and form the radials (Grandel & Schulte-Merker, 1998). This mode of development constrains the pattern of adult osteology to a distinct and more restricted set of states than that of tetrapods and their fish relatives, lung fish and coelacanths.

3.3. HOMOLOGY AND EVOLUTIONARY INNOVATION

Different characters in two species are homologous if they have evolved from the same character in a common ancestor (Fitch, 2000). The notion of homology implies the existence of alternative phenotypic solutions to the same evolutionary problem. If there was only one solution, then all organs with the same function would be structurally identical, which they are not. For instance, there are obvious differences between an insect wing and a bird wing even though they both serve the same function. At the same time, homologous characters can serve radically different functions and still retain the same basic structure (Riedl, 1978). For instance, the limb of tetrapods are used for everything from swimming, running, digging to flying while remaining recognizably the same organ with the same basic plan that is clearly distinct from, say, arthropod appendages. Homology is a statement about what remains the same despite the forces of natural selection which act to adapt a character for different functions. Homology is, therefore, both a hypothesis about the existence of a common ancestral character and the accessibility of character states by mutation and selection (Wagner, 1989a, 1994, 1999). In the language of this paper, different homologues exist in distinct accessibility domains and two realizations of the same homologue are elements of the same accessibility domain.

The rather informal notion of evolutionary innovation describes the fact that certain phenotypic changes are difficult to achieve

and seem more important for the subsequent evolution of a character than others (Buss, 1987). The concept is closely related to that of homology (Müller & Wagner, 1991), since identifying "novelty" implies a definition of what constitutes more of the "same". An innovation may be characterized as a transformation of a phenotypic character that radically changes the set of subsequently accessible phenotypes (Galis, 2001). It is tempting to speculate that this notion is related to the notion of discontinuous change in the sense of Fontana & Schuster (1998a).

3.4. IRREVERSIBILITY AND DIRECTIONALITY

There are many examples of evolutionary reversibility, most notably the evolution of polygenic quantitative characters, such as body size (Roff, 1997). Yet, not all evolutionary transitions are readily reversible. For instance, the evolution of a genetically inactive Y-chromosome or the evolution of obligatory parthenogenesis seem to be irreversible (Bull & Charnov, 1985). Many examples of evolutionary irreversibility involve the loss of genetic information, since it is easier to lose a functional part of the genome, and the corresponding phenotype, than regaining it by mutation. Even if evolution as a whole is not directional, there is sufficient evidence suggesting that certain transformations are highly biased. One direction of the transformation is easy, like the loss of an enhancer element, but the inverse step is unlikely to occur. The same holds for the small stack example in the RNA case detailed in Section 2.2.1. Some directional trends in evolution are therefore explained more naturally by asymmetries in transition probabilities than by directional selection. These "entropic" or combinatorial biases are, again, reflected in the asymmetric accessibility between phenotypes.

3.5. CAVEATS

The definition of accessibility depends on the available operators of genetic change and the rate of mutation. In the RNA example (Section 2.2.1), we defined accessibility in terms of a *single* point mutation. This choice is motivated by simplicity and assumes sufficiently high replication accuracies where the replica of a genome is unlikely to contain two or more mutations with respect to its template. In general, the definition of accessibility should reflect the rate of mutation and the genetic moves, such as insertions and deletions or recombination, that are relevant in a given context.

A distinction must be made between the static topology of phenotype space as induced by the genotype–phenotype map under given genetic operators and the consequences of that topology for evolutionary dynamics. The extent of these consequences depends, for example, on population size. In an infinite population, evolution becomes essentially a kinetic problem in which the issue of innovation does not figure. Two neutral networks that share only a small fraction of their boundary affect phenotypic transition times differently in infinite and finite populations. This contribution is only concerned with developing a topological language for the static aspects of phenotype space, motivated by our belief that this topology profoundly affects evolution in finite populations. In this section, we have pointed at these consequences qualitatively, but a quantitative analysis must take population size into account. For the RNA case, see Weber (1997); Fraser & Reidys (1997); Forst (2000); Reidys *et al.* (2001).

Mutation rates, mutation operators, the environment and the genotype–phenotype map itself are subject to evolution and so is the topology of phenotype space. None of this is touched upon in the present contribution.

Having brought together a series of arguments of why it seems desirable, if not necessary, to introduce non-metric accessibility structures into the language of mathematical evolutionary theory, we now proceed to present some of the pertinent concepts.

4. Pre-topological Nearness and Neighborhood

4.1. TOPOLOGICAL CONCEPTS

This section provides a brief, yet rigorous, introduction to the mathematical structures needed to reason about the accessibility topology induced by genotype–phenotype maps, such as the RNA folding map. No original mathematics is provided here. The effort rather consists in making fairly abstract material accessible to the theoretical biologist who grapples with patterns of

phenotypic evolution, but is unfamiliar with topological concepts. We refer to the textbook of Gaal (1964) for proofs that are not central to our topic.

Sets augmented with relations among their elements are called "spaces". Spaces are distinguished by the degree of structure they possess. Figure 3 provides a highly simplified concept chart. Euclidean vector spaces are perhaps the most concrete, since they possess a rich algebraic structure and are close to our intuitive understanding of time and space: vectors are elements that can be added, multiplied with a scalar and projected onto each other. We exploit this structure when making drawings. If it is removed, the familiar notion of distance still remains intact and characterizes a metric space. The shapes of molecules—such as proteins or RNA—or the sequences of genes are examples of elements forming a metric space: shapes or sequences cannot be added, but their distance (or similarity) can still be quantified. If this traditional notion of distance is dropped, a notion of neighborhood still remains. Elements entertain relationships of nearness, but nearness is not a number anymore. Two elements may be near to a third, but there is not enough structure to always state which one is nearer. A space of this kind is a topological space. It has enough structure to support a notion of boundary that behaves in a familiar way like the boundary we draw around an area on a sheet of paper. More specifically, a set can be "closed" by including its boundary, and closing a set twice does not add anything further. Removing the structure underlying this behavior of boundary uncovers the weakest notion of nearness that characterizes a pre-toplogical space. Dropping the notion of neighborhood still saves convergence. On giving up convergence, we are left with a plain set.

Properties of an abstract space are often discussed in terms of a more concrete one. For example, topologies are derived top-down from metric spaces by using the notion of distance to define neighborhood (such as ε-balls in $\mathbb{R}^n$). This makes everything a notch more familiar. Our interest, however, is in those cases for which a notion of distance is not available. This bottom-up direction requires a more abstract axiomatic approach.

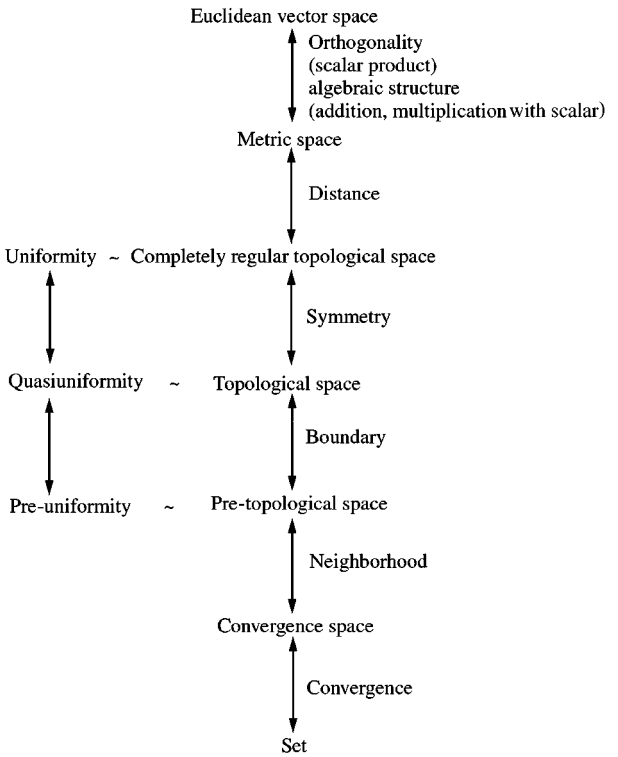


FIG. 3. Simplified topological concept chart. See main text for details.

4.2. NEARNESS

When a numerical distance measure is not available, nearness becomes a relationship that must be explicitly declared to hold between two elements of a set X.§§ The result is a list of pairs $(x, y) \in X \times X$ called a nearness relation U on $X (U \subseteq X \times X)$:

$$U = \{(x, y) | \text{“}y \text{ is } U\text{-near to } x\text{”}\}.$$

It is not required that $(x, y) \in U$ implies $(y, x) \in U$. In general, U is not symmetric.

We expect a formal nearness relation to capture the essential intuitions about nearness. Any nearness relation should, therefore, contain at the very least (x, x) for all $x \in X$, corresponding to the intuition that an element x is always near to itself. This is actually all one can say about a *particular* U. Consider two nearness relations U and U' (on the same set X). If $U \subset U'$, we can think of U as

§§This usage of "nearness" is not related to the notions of proximity or nearness in the sense of Pervin (1963); Herrlich (1974).

the result of applying a finer sieve, that is, a more stringent set of conditions which is satisfied by fewer elements compared to U'. Hence, U expresses a *finer* scale of nearness compared to U'. This affords a way of speaking about degrees of nearness (or levels of resolution), despite nearness not being a number. Furthermore, it seems natural to say that pairs of elements that are both U-near and U'-near are also $U \cap U'$-near. (Note that $U \cap U'$ is not empty, since (x, x) is contained both in U and in U' for all x.) Finally, consistency suggests that if y is U-near to x at one level of resolution, it should remain so at any coarser level.

Nearness is thus expressed by a *collection* $\mathcal{U}$ of relations U on X constrained to satisfy the following axioms:

(U1) $\Delta \subseteq U$ for all $U \in \mathcal{U}$, where $\Delta = \{(x, x) | x \in X\}$ is called the *diagonal.*
(U2) $U, U' \in \mathcal{U}$ implies $U \cap U' \in \mathcal{U}$.
(U3) $U \in \mathcal{U}$ and $U \subset U'$ implies $U' \in \mathcal{U}$.

The collection $\mathcal{U}$ is called a *pre-uniformity*.

4.3. NEIGHBORHOOD

An alternative way of structuring a set X with a notion of nearness consists in defining for each $x \in X$ a collection of subsets of X called neighborhoods of x. In analogy to nearness relations, a collection of *neighborhood systems* $\mathcal{N}(x)$ on X is formalized as a map $\mathcal{N}: X \to \mathcal{P}(X)$ [where $\mathcal{P}(X)$ is the powerset of X] such that for all $x \in X$:

(N1) $x \in N$ for all $N \in \mathcal{N}(x)$.
(N2) $N_1, N_2 \in \mathcal{N}(x)$ implies $N_1 \cap N_2 \in \mathcal{N}(x)$.
(N3) $N_1 \in \mathcal{N}(x)$ and $N_1 \subset N$ implies $N \in \mathcal{N}(x)$.

The Ns are neighborhoods, $\mathcal{N}(x)$ is a neighborhood system for element x, and the pair $(X, \mathcal{N})$ is called a *pre-topological space*. We speak of a *neighborhood basis*, if only (N1) and a weakened version of (N2)

(N2′) $N_1, N_2 \in \mathcal{N}(x)$ implies that there is a $N_3 \in \mathcal{N}(x)$ such that $N_3 \subseteq N_1 \cap N_2$

are satisfied.

It will be useful to extend the notion of neighborhood from individual elements to subsets of X.

Definition (*Neighborhood of a set*). Let $(X, \mathcal{N})$ be a pre-topological space and $B \subseteq X$. Then N is a neighborhood of B if and only if N contains a neighborhood N_x of each element $x \in B$.

The neighborhood system, $\mathcal{N}(B)$, for the set B is thus given by

$$\mathcal{N}(B) = \{N | N \in \mathcal{N}(x) \, \forall x \in B\} = \bigcap_{x \in B} \mathcal{N}(x). \quad (2)$$

4.4. FROM NEARNESS TO NEIGHBORHOOD (AND BACK)

A pre-topological neighborhood system $\mathcal{N}$ can be constructed from a pre-uniformity $\mathcal{U}$ in a natural way. For each $x \in X$, we define its neighborhood system $\mathcal{N}(x)$ to consist of the sets

$$U[x] = \{y \in X | (x, y) \in U\} \quad \text{for each } U \in \mathcal{U}. \quad (3)$$

It is easy to verify that the sets $U[x]$ satisfy the conditions (N1–N3) defining a neighborhood system. We call $\mathcal{N}_{\mathcal{U}}$ with $\mathcal{N}_{\mathcal{U}}(x) = \{U[x] | U \in \mathcal{U}\}$ the neighborhood system *induced* by the pre-uniformity $\mathcal{U}$.

Conversely, given a neighborhood system $\mathcal{N}$, we may construct a corresponding pre-uniformity as the collection $\mathcal{U}_{\mathcal{N}}$ of all sets U of the form

$$U = \{(x, y) | x \in X, y \in N_x \quad \text{for some } N_x \in \mathcal{N}(x)\}, \quad (4)$$

plus all sets U' containing some U [axiom (U3)]. Construction (4) says that a particular U is obtained by choosing for each $x \in X$, some neighborhood of x and (naturally) declaring its elements to be near x. The chosen neighborhoods are removed from the system $\mathcal{N}$ and the procedure is repeated to obtain a new U until all neighborhoods have been used up. $\mathcal{U}_{\mathcal{N}}$ is a *pre-uniformization* of the neighborhood system $\mathcal{N}$.

A pre-uniformity $\mathcal{U}$ and its induced pre-topology $(X, \mathcal{N})$ are very similar ways of structuring the set X. The pre-topology $\mathcal{N}_{\mathcal{U}_{\mathcal{N}}}$ induced by the pre-uniformization $\mathcal{N}_{\mathcal{U}}$ of $\mathcal{N}$ always coincides

with $\mathcal{N}$. This is shown in Appendix A.1.1. The converse, however, is not true in general. The relation between pre-uniformities and pre-topologies on X is not one-to-one. In general, different pre-uniformities give rise to the same pre-topology.

4.5. FROM PRE-TOPOLOGY TO TOPOLOGY

The concatenation of two nearness relations U' and U'' is defined by

$$U' \circ U'' = \{(x, y) | \exists z : (x, z) \in U' \text{ and } (z, y) \in U''\}. \tag{5}$$

$U' \circ U''$ contains both U' and U'' because each nearness relation contains the diagonal $\Delta = \{(x, x) | x \in X\}$ by virtue of axiom (U1). The concatenation of nearness relations enables us to lower the resolution of nearness: elements of U' and U'' are near on a finer scale than elements of $U' \circ U''$. We can think of the elements z in eqn (5) as "in between" x and y.

A pre-uniformity $\mathcal{U}$ such that

(UB) for each $U \in \mathcal{U}$ there is a $V \in \mathcal{U}$ with $V \circ V \subseteq U$,

is called a *quasiuniformity*. In essence, (UB) states that the structure of the universe X is such that for any two elements x and y, there is another element z in between [this may bottom out at some finite resolution where the only elements between (x, y) are x and y themselves.]

The condition (UB) has an interesting consequence which is best explained in the language of neighborhoods rather than nearness relations. In Appendix A.1.2, we show that the neighborhood equivalent of (UB) is

(N4) For each $N \in \mathcal{N}(x)$, there is an $N' \in \mathcal{N}(x)$ such that $N \in \mathcal{N}(y)$ for all $y \in N'$.

A pre-topology that satisfies (N4) is called a *topology*. The difference between the two spaces lies in the concept of *boundary*. Given a neighborhood system on X and a set $A \subseteq X$, we call $x \in X$ a boundary element of A if all the neighborhoods of x intersect both A and its complement $X \backslash A$. The boundary of A, ∂A, is the collection of all boundary elements of A:

$$\partial A = \{x \in X | \forall N \in \mathcal{N}(x) : N \cap A \neq \emptyset \text{ and } N \cap (X \backslash A) \neq \emptyset\}. \tag{6}$$

We can now define the *interior* and the *closure* of A as

$$\mathring{A} = A \backslash \partial A, \quad \bar{A} = A \cup \partial A. \tag{7}$$

By definition, a set A is *open* if it contains no boundary element. Stated positively, a set A is open if it contains a neighborhood of each of its elements, that is, for each $x \in A$, there is a neighborhood $N \in \mathcal{N}(x)$ such that $N \subseteq A$. A cartoon of the concept is given in Fig. 4. Open sets play a prominent role, because the collection of all open sets that contain x, $\mathcal{T}(x) = \{A | A \text{ is open and } x \in A\}$ constitutes a neighborhood basis at x (Gaal, 1964). In fact, if a neighborhood system $\mathcal{N}(x)$ satisfies (N4), then $\mathcal{T}(x)$ is a basis of $\mathcal{N}(x)$ and vice versa (Theorem IX′ in Alexandroff & Hopf, 1935).

Returning to the difference between a topology and a pre-topology, consider the behavior of the closure operation, $A \cup \partial A$. For the sake of simplicity, assume the open sets $\mathcal{T}(x)$ as the neighborhood basis of the topology. We close an arbitrary set A by adding all its boundary elements (6). What happens if we perform a closure twice? Intuitively, once a set has been closed, there are no further boundary elements and therefore nothing should happen. This is indeed how the boundary operation behaves in a topology. Yet, in a pre-topology, adding all boundary elements to a pre-topological neighborhood may result in the creation of further boundary elements.

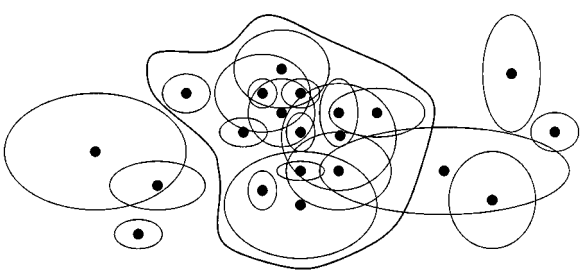


FIG. 4. Open sets. An open set contains a neighborhood of each of its elements.

Consider a scenario in which x is among the boundary elements ∂A just added to A. Suppose further that there exists an element y outside of A all of whose neighborhoods N_y contain the previous boundary element x. Owing to the inclusion of x, y has now become a boundary element of $A \cup \partial A$. When y is added at the next application of the boundary operator, this scenario may repeat itself and the set may continue growing. This cannot happen in a topology. If x is in the neighborhoods N_y of y, any N_y must contain at least one neighborhood of x (because the N_y are open sets). However, if x is a boundary element, then y must be one too, because its N_y will intersect A by virtue of containing a neighborhood of the boundary element x. Thus, once all boundary elements have been added to A, no further boundary elements can be created. In a topology, the boundary operator satisfies

$$\partial\partial A \subseteq \partial A,$$

which is equivalent to the idempotence, $\bar{\bar{A}} = \bar{A}$, of the closure operation (Albuquerque, 1941). The notion of boundary, in the sense of definition (6), does exist in a pre-topology, but its behavior is not as familiar as in a topology. Pre-topological spaces can be specified by equivalent axiom systems in terms of neighborhoods, closure, interior, or boundary. Their mutual translations are summarized in Appendix A.1.5.

Finally, symmetry results from the additional requirement:

(US) $U \in \mathcal{U}$ implies $U^{-1} = \{(x, y) | (y, x) \in U\} \in \mathcal{U}$.

It follows from (U2) that the symmetric relations $U \cap U^{-1}$ are also nearness relations. If (US) is satisfied, one speaks of a *semiuniformity*, and if both (US) and (UB) hold, we have a *uniformity*. In terms of neighborhoods, the symmetry axiom (US) implies two equivalent properties.||||

(R0) $x \in \overline{\{y\}}$ implies $y \in \overline{\{x\}}$ for all $x, y \in X$.
(S′) $x \in \bigcap \mathcal{N}(y)$ implies $y \in \bigcap \mathcal{N}(x)$.

Axiom (R0), introduced by Šanin (1943), plays an important role in topology as a notion of symmetry. Čech (1966, Theorem 23.B.3) proved that a pre-topological space is semiuniformizable if and only if it satisfies (S′), and, equivalently, (R0).

||||The equivalence is proven in Appendix A.1.4.

4.6. CONTINUITY

The debate on continuity in evolution would greatly benefit from a formal definition of the term. The notion of nearness is instrumental for this purpose. Before connecting nearness with continuity, however, we begin with the most general notion of continuity, which depends on even less structure than is available in pre-topological spaces.

The definitions of both nearness (Section 4.2) and neighborhood (Section 4.3) make use of the same generic structure. This structure deserves special emphasis. Let X be a set. A *filter* on X (Cartan, 1937; Gaal, 1964) is a subset $\mathcal{F}$ of the power set of X, $\mathcal{P}(X)$, with the following properties:

(F1) $\emptyset \notin \mathcal{F}$.

(F2) $F_1, F_2 \in \mathcal{F}$ implies the existence of a set $F_3 \in \mathcal{F}$ such that $F_3 \subseteq F_1 \cap F_2$.

(F3) If $F_1 \in \mathcal{F}$ and $F_1 \subseteq F_2$ then $F_2 \in \mathcal{F}$.

If $\mathcal{F}$ satisfies only (F1) and (F2), one speaks of a *filter basis* (which uniquely defines a filter). It is easy to verify that the neighborhood system $\mathcal{N}(x)$ of an element x in a pre-topological space $(X, \mathcal{N})$ is a filter on X, and that a pre-uniformity $\mathcal{U}$ of nearness relations is a filter on $X \times X$.

We say that $\mathcal{F}$ is *coarser* than $\mathcal{G}$ (or $\mathcal{G}$ is *finer* than $\mathcal{F}$) if $\mathcal{F} \subseteq \mathcal{G}$. Equivalently, $\mathcal{F}$ is coarser than $\mathcal{G}$ if for every $F \in \mathcal{F}$ there is $G \subseteq F$ such that $G \in \mathcal{G}$. (Note the reversal in the subset relation when passing from filters to their elements. See also the notion of "resolution" in the context of nearness relations, Section 4.2.)

A filter $\mathcal{F}$ is finest (or maximal) if it is contained in no other filter. This is equivalent to saying that

$$\text{for any } A \subset X \text{ either } A \in \mathcal{F} \text{ or } X - A \in \mathcal{F}, \quad (8)$$

which justifies the name "filter".

Filters are useful in defining convergence. Think of filters as generalizations of sequences.

Given a sequence $(x_n) = \{x_1, x_2, \dots,\}$, define the "ends" as $F_k = \{x_k, x_{k+1}, \dots\}$. It is straightforward to verify that the set of ends, $\{F_k | k \in \mathbb{N}\}$, satisfies (F1) and (F2) and is therefore the basis of a filter $\mathscr{F}$. (The basis here is like a series of telescopically nested tubes.) In the case of a sequence, we say that (x_n) converges to a limit point x, $x_n \to x$, if for all $\varepsilon > 0$ there is an integer n_ε such that $\|x_k - x\| < \varepsilon$ for all $k \geqslant n_\varepsilon$. The notion of filter enables us to speak of convergence without invoking a notion of distance $\|x_k - x\|$. Stated in terms of neighborhoods, the convergence $x_n \to x$ means that for every neighborhood N of x, there is an integer n_N such that $x_k \in N$ for all $k > n_N$. The phrase "$x_k \in N$ for all $k > n_N$" simply means that $F_{n_N} \subseteq N$. Recall that a neighborhood system constitutes a filter. Thus, (x_n) converges to x if and only if the filter $\mathscr{F}$ generated by the ends F_k of (x_n) is finer that the neighborhood filter of x, that is, $\mathscr{N}(x) \subseteq \mathscr{F}$. This replaces the notion of a distance becoming smaller and motivates the definition of (filter) convergence in a pre-topological space:

Definition (*Convergence*). Let $(X, \mathscr{N})$ be a pre-topological space and let $\mathscr{F}$ be a filter on X. Then $\mathscr{F}$ converges to x, in symbols: $\mathscr{F} \to x$ or $x \in \lim \mathscr{F}$, if and only if $\mathscr{N}(x) \subseteq \mathscr{F}$.

Filter convergence sets the stage for the notion of a continuous function.

Definition (*Continuity*). Let $f: (X, \mathscr{N}) \to (Y, \mathscr{M})$ be a function between two pre-topological spaces. We say f is continuous in $x \in X$ if for all filters $\mathscr{F}$ on X

$$\mathscr{F} \to x \text{ implies } f(\mathscr{F}) \to f(x). \tag{9}$$

Let us translate the definition of continuity into the language of neighborhoods:

Lemma 1. *Let f: $(X, \mathscr{N}) \to (Y, \mathscr{M})$ be an arbitrary function between two pre-topological spaces. Then the following propositions are equivalent*:

(i) *f is continuous in x.*

(ii) *For every neighborhood M of $f(x)$, there is a neighborhood N of x such that $f(N) \subseteq M$.*

(iii) $\mathscr{M}(f(x)) \subseteq f(\mathscr{H}(x))$.

Assertion (iii) follows directly from the definition of convergence and the definition of continuity. Observe that "$\mathscr{F} \to x$ implies $f(\mathscr{F}) \to f(x)$" becomes "$\mathscr{N}(x) \subseteq \mathscr{F}$ implies $\mathscr{M}(f(x)) \subseteq f(\mathscr{F})$". A fact about filters (given here without proof) asserts that an arbitrary function f preserves coarseness, that is, $\mathscr{F} \subseteq \mathscr{G}$ implies $f(\mathscr{F}) \subseteq f(\mathscr{G})$. Hence, $f(\mathscr{N}(x)) \subseteq f(\mathscr{F})$. But "$f(\mathscr{N}(x)) \subseteq f(\mathscr{F})$ implies $\mathscr{M}(f(x)) \subseteq f(\mathscr{F})$" is equivalent to "$\mathscr{M}(f(x)) \subseteq f(\mathscr{N}(x))$" which, in words, states that the neighborhood filter of $f(x)$ is coarser than the image of the neighborhood filter of x. This is equivalent to assertion (ii) which is but the definition of filter coarseness. At the same time, (ii) is the familiar neighborhood-based definition of continuity (Fig. 5).

In Appendix A.1.3, we rephrase continuity in terms of nearness relations or pre-uniformities.

4.7. FINITE SETS

Pre-topologies simplify considerably in the case of a finite universe X. There are only finitely many filters and every filter $\mathscr{F}$ is of the form

$$\mathscr{F} \equiv \dot{F} = \{F' | F \subseteq F'\}, \tag{10}$$

where F is a subset of X. (Such filters are called *discrete filters*.) This one-to-one correspondence between subsets of X and filters on X permits most properties to be stated in terms of subsets.

A particularly useful subset is the *vicinity* associated with the neighborhood filter $\mathscr{N}(x)$:

$$N(x) = \bigcap \mathscr{N}(x) = \bigcap \{N | N \in \mathscr{N}(x)\}. \tag{11}$$

The notion of vicinity can be used to establish a correspondence between pre-topological spaces

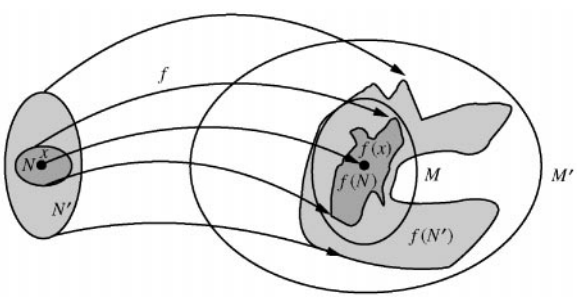


FIG. 5. Continuity. For each neighborhood M of $f(x)$, there is a neighborhood N of x such that $f(N) \subseteq M$.

$(X, \mathcal{N})$ and directed graphs (digraphs) $\Gamma(X, E)$ where X is the vertex set and E the set of *directed* edges from x to y, $E = \{(x, y) | x \in X, y \in N(x) \backslash \{x\}\}$.

Properties of pre-topological spaces can now be stated in more familiar graph-theoretical terms. For instance:

- By construction, the vicinity $N(x)$ consists of the forward-neighbors of x and x itself.
- A pre-topological space $(X, \mathcal{N})$ is finer than another pre-topological space on the same set X, $(X, \mathcal{M})$, if and only if, $\Gamma(X, E_{\mathcal{N}})$ is a subgraph of $\Gamma(X, E_{\mathcal{M}})$.
- A function $f: (X, \mathcal{N}) \to (Y, \mathcal{M})$ is continuous in x if and only if, $f(N(x) \subseteq M(f(x))$, that is, if f maps the vicinity of x into the vicinity of $f(x)$.
- Axiom (S′) makes $(X, \mathcal{N})$ a symmetric directed graph (two vertices are either connected by an edge in each direction or not at all). Symmetric digraphs are, of course, isomorphic to undirected graphs, which are, therefore, exactly the finite pre-topological spaces satisfying (R0).

A finite pre-topological space is topological if and only if, for each $x \in X$ and all $y \in N(x)$ there exists an $N(y)$ such that $N(y) \subseteq N(x)$. This means that the vicinities must be open sets.

- The *open vicinity* $T(x)$ of an element x, that is, the smallest open set containing x consists of all elements that can be reached from x along any number of forward-edges.

A concise characterization of directed graphs that express particular topologies on their vertex sets seems to be unknown. Some interesting results in this direction can be found in Cupal *et al.* (2000).

5. Pre-topologies and the Genotype–Phenotype Map

5.1. THE ACCESSIBILITY PRE-TOPOLOGY

Using the concepts reviewed in Section 4, we next consider the structure of phenotype space induced by a map f from genotypes G to phenotypes P. The folding from RNA sequences to secondary structures (Fig. 1 and Section 2.2.1) will serve as an example.

The genotype–phenotype map assigns to each genotype g a phenotype $\psi = f(g)$.¶¶ The central question is how to organize the set of phenotypes, that is, which neighborhood system is *natural* for phenotypes? The corresponding question for genotypes poses no difficulty, since physical processes exist which directly change genotypes and hence provide a natural neighborhood structure on the set of possible genotypes. Phenotypes, however, are not modified directly. Phenotypic innovation is the result of genetic modification mediated by development (the genotype–phenotype map). This reasoning motivated Fontana & Schuster (1998a) to consider a notion of phenotypic neighborhood induced by the genotype–phenotype map which differs fundamentally from a notion of nearness among phenotypes based solely on the comparison of their morphological features.

The induced neighborhood structure on the set of phenotypes reflects "accessibility" of one phenotype by another through mutations in the genotype of the former. An interesting situation arises when the genotype–phenotype map is many-to-one, which is typically the case in a realistic setting. The notion of nearness of a phenotype ψ to another should be a robust property, independent of a particular genotype giving rise to ψ—it should, in a sense, reflect a feature that is common to *all* genotypes whose phenotype is ψ. In a many-to-one map, phenotypes denote equivalence classes of genotypes (the set of genotypes sharing the same phenotype). Nearness among phenotypes, then, must reflect the mutual adjacency of these equivalence classes as determined in the given neighborhood structure of genotype space (Fontana & Schuster, 1998a, b).

We address this intuition formally by first asking a seemingly unrelated question: What kind of neighborhood system $\mathcal{M}$ on the set of phenotypes makes the genotype–phenotype map everywhere continuous?

From Lemma 1, we know that for f to be everywhere continuous, we must have for all

¶¶To improve clarity of exposition, we shall ignore the dependency of the phenotype on the environment. The inclusion of an environment does not affect the essence of the arguments presented here.

phenotypes ψ and all genotypes g that $\mathscr{M}(\psi) \subseteq f(\mathscr{N}(g))$. When several genotypes g_i give rise to the same phenotype ψ, the requirement for continuity becomes

$$\mathscr{M}(\psi) \subseteq f(\mathscr{N}(g_1)) \text{ and}$$

$$\mathscr{M}(\psi) \subseteq f(\mathscr{N}(g_2)) \text{ and } \ldots \tag{12}$$

for all $g \in f^{-1}(\psi)$. Define the neighborhood system

$$\mathscr{A}(\psi) := \bigcap_{g \in f^{-1}(\psi)} f(\mathscr{N}(g))$$

$$= \{S | S \in f(\mathscr{N}(g)) \forall g \in f^{-1}(\psi)\}. \tag{13}$$

In compliance with (N3), $\mathscr{A}(\psi)$ is meant to include all sets containing a set described in eqn (13), but we shall not explicitly notate this fact. Requirement (12) now becomes

$$\mathscr{M}(\psi) \subseteq \mathscr{A}(\psi) \quad \text{for all } \psi. \tag{14}$$

We shall see that $\mathscr{A}(\psi)$ has a simple interpretation. By its definition (13), $\mathscr{A}(\psi)$ is just the collection of sets S containing the image of some neighborhood shared by all genotypes g with phenotype ψ:

$$\mathscr{A}(\psi) = \{S | \exists N_g \in \mathscr{N}(g) \text{ such that}$$

$$f(N_g) \subseteq S \forall g \in f^{-1}(\psi)\} = \cdots, \tag{15}$$

which is just the collection of images of neighborhoods shared by all g with phenotype ψ [plus supersets by virtue of (N3)]:

$$\cdots = \{f(N) | N \in \mathscr{N}(g) \forall g \in f^{-1}(\psi)\}$$

$$= f\left(\bigcap_{g \in f^{-1}(\psi)} \mathscr{N}(g)\right). \tag{16}$$

The collection $\bigcap_{g \in f^{-1}(\psi)} \mathscr{N}(g)$ is the set of neighborhoods shared by all genotypes g with phenotype ψ, $\{N | N \in \mathscr{N}(g) \forall g \in f^{-1}(\psi)\}$. In the case of RNA, $f^{-1}(\psi)$ is the so-called neutral set (or neutral network when all sequences folding into the same structure are mutationally connected), and $\bigcap_{g \in f^{-1}(\psi)} \mathscr{N}(g)$ is the neighborhood system of the neutral set, $\mathscr{N}(f^{-1}(\psi))$. (See the definition for the neighborhood system of a set in Section 4.3.) In sum, we have

$$\mathscr{M}(\psi) \subseteq \mathscr{A}(\psi) = \bigcap_{g \in f^{-1}(\psi)} f(\mathscr{N}(g))$$

$$= f\left(\bigcap_{g \in f^{-1}(\psi)} \mathscr{N}(g)\right) = f(\mathscr{N}(f^{-1}(\psi))).$$

In words, a phenotype ϑ is contained in a neighborhood N_ψ of phenotype ψ ($N_\psi \in \mathscr{A}(\psi)$) if and only if there is a neighborhood of $g \in f^{-1}(\psi)$ which contains a genotype h folding into ϑ. This is straightforward for maps between finite sets, where the neighborhood structure is determined by the vicinity (the smallest neighborhood, see Section 4.7). In genotype space, the vicinity of the neutral set of ψ comprises all sequences obtained by a single point mutation from sequences folding into ψ. With respect to phenotypes, the vicinity of ψ, $A(\psi)$, therefore consists of all structures ϑ that can be accessed through a single point mutation from sequences folding into ψ:***

$$A(\psi) = \bigcup_{g \in f^{-1}(\psi)} f(N(g))$$

$$= \{\vartheta | \exists g \in f^{-1}(\psi) \text{ and } h \in N(g) \text{ such that}$$

$$\vartheta = f(h)\}. \tag{17}$$

The pre-topology $\mathscr{A}$ on the set of phenotypes is the weakest notion of phenotypic accessibility—weakest in the sense that, according to eqn (17), for phenotype ϑ to be in the neighborhood of ψ, it suffices that ϑ be realized just once by some one-error mutant of a sequence folding into ψ. $\mathscr{A}$ is the finest pre-topology on the set of phenotypes P such that $f: (G, \mathscr{N}) \to P$ is a continuous function. We refer to $\mathscr{A}$ as the *accessibility pre-topology*††† of phenotype space or the *final* pre-topology generated by f from $(G, \mathscr{N})$.

***The reader may wonder, in a first moment, why the intersection in eqn (13) becomes a union in eqn (17). More generally, the intersection of filters can be written as the union of their elements. This is clarified in Appendix A.1.6.

†††The concept of accessibility of phenotypes developed here is not related to the notion of accessibility spaces in the sense of Whyburn (1970).

The most restrictive sense of accessibility arises by requiring that ϑ is in the neighborhood of ψ only if ϑ is realized in the genetic vicinity of *every* sequence with phenotype ψ. In the finite case, this translates to

$$C(\psi) = \bigcap_{g \in f^{-1}(\psi)} f(N(g)) = \{\vartheta | \forall g \in f^{-1}(\psi) : \exists\, h \in N(g) \text{ such that } \vartheta = f(h)\}. \quad (18)$$

In the infinite case, we cannot simply replace the intersection of the filters $f(\mathcal{N}(g))$ in eqn (13) by their union, since the union of two filters is, in general, not a filter (see Appendix A.1.6). Instead, we must use the filter arising from the intersections of the individual neighborhoods. We use the notation

$$\mathcal{F} \vee \mathcal{G} = \{F \cap G | F \in \mathcal{F},\ G \in \mathcal{G}\} \quad (19)$$

for the coarsest filter that is finer than both $\mathcal{F}$ and $\mathcal{G}$. Note that $\mathcal{F} \vee \mathcal{G}$ exists only if $F \cap G \neq \emptyset$ for all $F \in \mathcal{F}$ and $G \in \mathcal{G}$; otherwise, $\mathcal{F}$ and $\mathcal{G}$ are called disjoint. Since $f(g) \in N$ for all $N \in f(\mathcal{N}(g))$ and all $g \in f^{-1}(\psi)$, no intersections are empty and the neighborhood filter

$$\mathcal{C}(\psi) = \bigvee_{g \in f^{-1}(\psi)} f(\mathcal{N}(g)), \quad (20)$$

exists. $\mathcal{C}$ is the coarsest pre-topology that is finer than $f(\mathcal{N}(g))$ for all $g \in f^{-1}(\psi)$. We call it the *shadow pre-topology*, because phenotype $\vartheta \in \mathcal{C}(\psi)$ "follows" ψ like a "shadow", being the image of a neighbor of every g that folds into ψ.

5.2. STATISTICAL NEIGHBORHOOD SYSTEMS

The accessibility pre-topology $\mathcal{A}$ of the previous Section 5 was constructed from the requirement that the genotype–phenotype map be continuous everywhere. This seems too strong a requirement, resulting in a rather weak neighborhood structure. One single genotype poised for a transition from α to β suffices to make β accessible from α. The shadow pre-topology $\mathcal{C}$ errs on the other extreme, as it requires that every genotype of α be mutable into a genotype of β. In the computational RNA genotype–phenotype model, the $\mathcal{C}$-pre-topology turns out to be trivial, since the $\mathcal{C}$-neighborhoods of α only contain α.

The notion of accessibility described in Section 2.2 emphasized the *likelihood* of a transition from phenotype α to phenotype β by mutation of genotypes underlying α. This affords a way of interpolating between the extreme versions of accessibility, $\mathcal{A}$ and $\mathcal{C}$. The likelihood of a phenotypic change is proportional to the number of genotypes with phenotype α that are adjacent to genotypes with phenotype β (Fontana & Schuster, 1998b; Cupal *et al.*, 2000), as expressed in eqn (1). An example distribution of such numbers is shown in Fig. 6. In the simplest case, a probabilistic version of accessibility introduces a cutoff point. If $A(\beta \curvearrowleft \alpha)$ [eqn (1)] is below that cutoff, β is not accessible from α. Depending on the cutoff point, a range of accessibility structures can be constructed on phenotypes. The appropriate cutoff value should be determined by biological factors, such as mutation rate, population size, or the chosen time frame (see Caveats in Section 3.5).

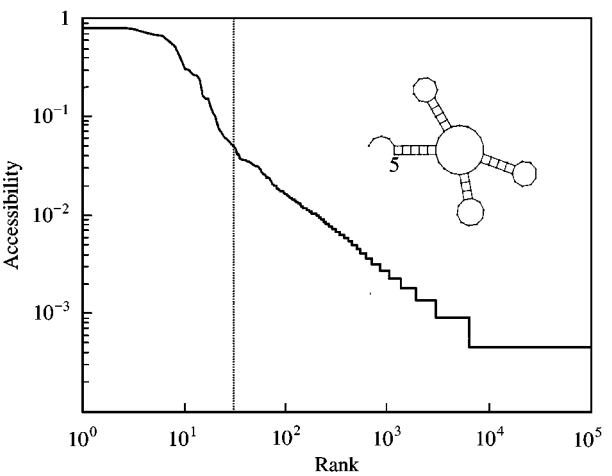


FIG. 6. Accessibility distributions. The one-error mutants of a sample of 2199 sequences folding into the tRNA clover-leaf reference structure (length $l = 76$, inset) were folded. 28% had the same structure as the reference. 72% folded into 141 907 distinct shapes. The curve shown is a log–log plot of the rank ordered $A(\beta \curvearrowleft \text{tRNA})$ [eqn (1)] for each of the 141 907 shapes β. The vertical line is meant to separate regions with different scaling, suggesting a natural cutoff point above which a shape β is regarded as being near the tRNA shape. For details, see Fontana & Schuster (1998b).

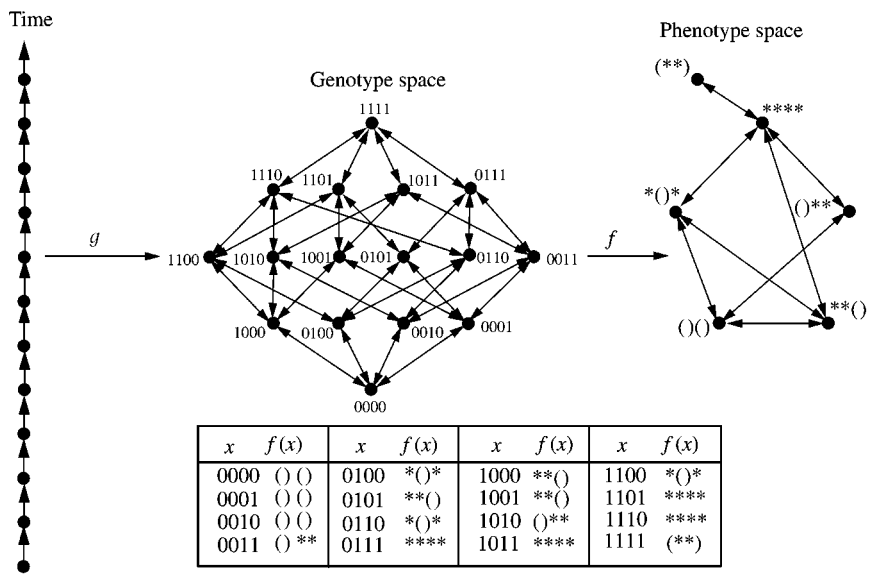
The limiting pre-topologies constructed in Section 5 served the purpose of formally justifying the idea of a phenotype space topology defined in terms of the mutational adjacency of genotypic equivalence classes. A rigorous treatment of a "statistical topology" (Fontana & Schuster, 1998b), however, must be based on consistent probabilistic notions of neighborhood and nearness which are well beyond the scope of this contribution. Probabilistic convergence spaces (Richardson & Kent, 1996) or fuzzy topology (Mordeson & Nair, 1998), in particular fuzzy pre-uniformities (Badard, 1984), may perhaps be useful in achieving this goal.

6. Continuity of Evolutionary Trajectories

An evolutionary trajectory can be viewed as a map from the "time axis" into the space of phenotypes. When analysing a series of paleontological samples or a series of shape transitions obtained from a computer simulation of RNA evolution, the time axis is inherently discrete with an obvious natural pre-topology. We simply number subsequent samples and define the vicinities on the time axis to be $N(t) = \{t, t+1\}$. The corresponding pre-topological space will be denoted by $(\mathbb{N}, \mathscr{T})$. Its graph is the directed path on the left in Fig. 7.

An evolutionary trajectory is the composition of two functions. First, a function $g: (\mathbb{N}, \mathscr{T}) \to (X, \mathscr{G})$ that assigns a genotype $g(t)$ to each point t in (discrete) time. The genotype space $(X, \mathscr{G})$ is a pre-topology induced by the genetic operators, such as point mutation in Fig. 7. This first function is then composed with a genotype–phenotype map $f: (X, \mathscr{G}) \to (Y, \mathscr{N})$. The space structure of the phenotypes Y is the accessibility pre-topology $\mathscr{N} = \mathscr{A}$ or $\mathscr{C}$, as discussed in Section 5.1, or a probabilistic version as discussed in Section 5.2.

An evolutionary trajectory, then, is a map $\tau: (\mathbb{N}, \mathscr{T}) \to (Y, \mathscr{N}): t \mapsto \tau(t) = f(g(t))$ whose first component—the time series of genotypes $g(t)$—is typically continuous, since genotypic changes occur by means of elementary genetic operators that determine the pre-topology $\mathscr{G}$ on X. This need not always be the case, however. For instance, if $\mathscr{G}$ is derived from point mutations (as in Fig. 7), then multiple mutations [that is, $g(t)$ and $g(t+1)$ differ in more than one sequence position], insertions, and deletions constitute discontinuities in $g: (\mathbb{N}, \mathscr{T}) \to (X, \mathscr{G})$. Yet, if we do not limit the case to continuity in the genetic



x	f(x)	x	f(x)	x	f(x)	x	f(x)
0000	()()	0100	*()*	1000	**()	1100	*()*
0001	()()	0101	**()	1001	**()	1101	****
0010	()()	0110	*()*	1010	()**	1110	****
0011	()**	0111	****	1011	****	1111	(**)

FIG. 7. Evolutionary trajectory. An evolutionary trajectory is the composition $f \circ g$ of the temporal sequence of genotypes and the genotype–phenotype map f. In the case of point mutations, the pre-topology $\mathscr{G}$ arranges the set X as a Hamming graph. For illustrative purposes, the phenotype space is endowed with the accessibility pre-topology $\mathscr{A}$. The genotype–phenotype map f is shown in the table.

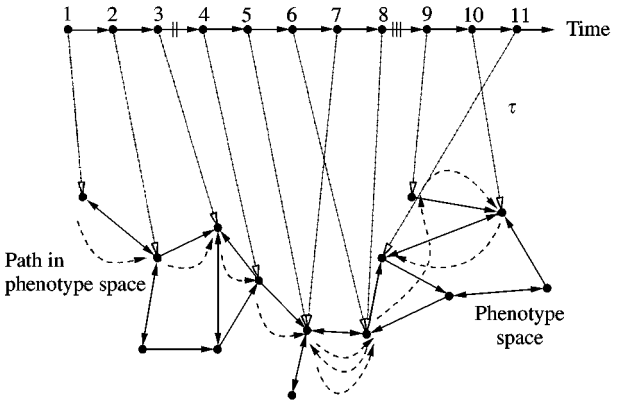


FIG. 8. Continuity of an evolutionary trajectory. A short trajectory $\tau:(\mathbb{N}, \mathscr{T}) \to (Y, \mathscr{N})$ is shown. Transitions from t to $t+1$ are continuous, or more precisely, τ is continuous at t, if the transition $[\tau(t) \curvearrowright \tau(t+1)]$ follows a directed edge in the pre-topology of phenotype space. In the present example, there are two non-continuous transitions, namely $\tau(3) \curvearrowright \tau(4)$ and $\tau(8) \curvearrowright \tau(9)$. Note that the transition $\tau(3) \curvearrowright \tau(4)$ becomes continuous in the topologization of $\mathscr{N}$ since $\tau(4)$ is reachable from $\tau(3)$ along a directed path. The transition $\tau(8) \curvearrowright \tau(9)$, however, remains discontinuous.

trace of the evolutionary trajectory, anything goes and nothing much can be said. If the genotype–phenotype map $f:(X, \mathscr{G}) \to (Y, \mathscr{N})$ is everywhere continuous ($\mathscr{N} = \mathscr{A}$), only genetic discontinuities can give rise to phenotypic discontinuities.

In practice, accessibility will be more restrictive than $\mathscr{N} = \mathscr{A}$ (and less restrictive than $\mathscr{N} = \mathscr{C}$). As discussed in Section 5.2, "effective" accessibility is better described by a pre-topology that is (much) finer than $\mathscr{A}$. As a consequence, f will not be everywhere continuous. It may even be the case that for any genotype g, there is at least one mutation of g that changes its phenotype in a discontinuous fashion, making f nowhere continuous. However, because the remaining mutations at g change its phenotype continuously, an evolutionary trajectory $\tau = f \circ g$—consisting of phenotypes constrained by selection—may still turn out to be continuous (Fig. 8). A transition at time t is continuous if and only if $\tau(t+1) \in N(\tau(t))$, that is, if it follows a directed edge in phenotype space.‡‡‡ We refer to Fontana & Schuster (1998a, b) for a classification and detailed discussion of discontinuities in evolutionary trajectories of simulated RNA populations.

‡‡‡In a continuous setting, the situation is qualitatively similar, albeit more difficult to visualize. The time axis will usually be the real axis $\mathbb{R}$ with the standard topology. Again, $\tau = f \circ g$ can be continuous even if neither g nor f are continuous everywhere.

7. Product Spaces and the Notion of Character

A formal explication of the character concept is a natural application of the present (pre)-topological framework. In evolutionary biology, the notion of character aims at identifying those phenotypic descriptors that are the quasi-independent units (Lewontin, 1978) of variation within and between species. The phenotypic variation of a character must to some extent decouple from the remaining organism, and the parsing of a phenotype into characters becomes a statement about the accessibility structure of phenotypic states. In this section, we shall argue that the factorization of a phenotype space (constructed on the basis of accessibility criteria) captures the essential formal properties of quasi-independence. Based on that argument, a notion of *structural independence* is proposed.

7.1. PRODUCT PRE-TOPOLOGIES

We start our discussion by introducing the notion of the (Cartesian) product of two pre-topological spaces $(X_1, \mathscr{N}_1)$ and $(X_2, \mathscr{N}_2)$. The *product pre-topology* (Carstens & Kent, 1969) on the Cartesian set product $X_1 \times X_2 = \{(x_1, x_2) | x_1 \in X_1, x_2 \in X_2\}$ is defined by the product of the neighborhood filters

$$\begin{aligned}\mathscr{N}(x_1, x_2) &= \mathscr{N}_1(x_1) \times \mathscr{N}_2(x_2) \\ &= \{M \subset X_1 \times X_2 \,|\, \exists N_1 \in \mathscr{N}_1(x_1), \\ &\qquad N_2 \in \mathscr{N}_2(x_2) : N_1 \times N_2 \subseteq M\}. \end{aligned} \tag{21}$$

We shall write the k-fold product as $\prod_k (X_k, \mathscr{N}_k)$ and restrict ourselves to a finite number of factors.

In the finite case, the product of pre-topological spaces $(X_1, \mathscr{N}_1) \times (X_2, \mathscr{N}_2)$ translates into the *strong product* of the associated graph representations (Section 4.7):

$$\Gamma(X, E) = \Gamma(X_1, E_1) \blacksquare \Gamma(X_2, E_2). \tag{22}$$

The vertex set X of $\Gamma(X, E)$ is $X_1 \times X_2$. The edge set E consists of all pairs $[(x_1, x_2), (x_1', x_2')) \neq ((x_1, x_2), (x_1, x_2)]$ such that $(x_1, x_1') \in E_1 \cup \Delta$ and $(x_2, x_2') \in E_2 \cup \Delta$ (Imrich & Klavžar, 2000, Chapter 5).

7.1.1. *Projectors*

An important class of maps associated with product spaces are the *projectors* $\mathbf{pr}_k$: $\prod_k (X_k, \mathcal{N}_k) \to (X_k, \mathcal{N}_k)$, defined as $\mathbf{pr}_k(x_1, x_2, \dots) = x_k$. The product map $f: (X, \mathcal{N}) \to \prod_k (Y_k, \mathcal{M}_k)$ is continuous in $x \in X$ (see Section 4.6) if and only if each of the maps $f_k = \mathbf{pr}_k \circ f$ is continuous in x (Fischer, 1959, Theorem 13).

7.1.2. *Isomorphism*

Two pre-topological spaces $(X, \mathcal{N})$ and $(X', \mathcal{N}')$ are *isomorphic* if there is a one-to-one map $\phi: (X, \mathcal{N}) \to (X', \mathcal{N}')$ such that both ϕ and ϕ^{-1} are continuous, that is, if and only if $\phi(\mathcal{N}(x)) = \mathcal{N}'(\phi(x))$ for all $x \in X$. We then write $(X, \mathcal{N}) \simeq (X', \mathcal{N}')$.

7.1.3. *Factorizability and Phenotypic Character*

The product $(X_1, \mathcal{N}_1) \times (X_2, \mathcal{N}_2)$ is trivial if one of the factors is a single point space, $(\{x\}, \dot{x})$. [Recall from Section 4.7 that $\dot{F}$ is the discrete filter, eqn (10).] Obviously,

$$(X_1, \mathcal{N}_1) \times (\{x\}, \dot{x}) \simeq (\{x\}, \dot{x}) \times (X_1, \mathcal{N}_1) \simeq (X_1, \mathcal{N}_1).$$

This suggests the following:

Definition (*Factorizability*). A pre-topological space $(X, \mathcal{N})$ is *factorizable* if it is isomorphic to a non-trivial product, in symbols

$$(X, \mathcal{N}) \simeq (\mathrm{X}_1, \mathcal{N}_1) \times (X_2, \mathcal{N}_2).$$

If the phenotype space $(Y, \mathcal{M})$ can be represented as a product space of the form

$$(Y, \mathcal{M}) \simeq (Y_1, \mathcal{M}_1) \times (Y_2, \mathcal{M}_2), \qquad (23)$$

then phenotype $y \in Y$ can be viewed as a "vector" (y_1, y_2). A phenotype (y_1', y_2') is accessible from (y_1, y_2) if y_1' is accessible from y_1 in the factor space $(Y_1, \mathcal{M}_1)$ and if y_2' is accessible from y_2 in the factor space $(Y_2, \mathcal{M}_2)$. Since the components y_1 and y_2 do not impose accessibility constraints on each other, they are *structurally independent*. This notion should not be confused with a stronger version of genetic independence. Structural independence implies that the factors y_1 and y_2 *can* be modified independently of each other, but it does not mean that they *always vary* independently. There still may be (continuous) mutations of the genotype g underlying (y_1, y_2) that affect both y_1 and y_2. Structural independence also does not imply statistical independence of the factors in a population. Fitness constraints, which by definition are not part of the genotype–phenotype map, may well cause covariations between the frequencies of the variants of two factors in a population.

A *primitive* character is one that cannot be subdivided into a collection of other characters. This suggests the following working definition:

Definition (*Primitive phenotype character*). If the phenotype space can be represented in the form $(Y, \mathcal{M}) = (Y_1, \mathcal{M}_1) \times (Y_2, \mathcal{M}_2)$ and $(Y_1, \mathcal{M}_1)$ is not factorizable, then $(Y_1, \mathcal{M}_1)$ is termed a *primitive phenotypic character*.

The usefulness of this definition depends on the ability to characterize and effectively compute factorizations of pre-topological spaces. In general, such a product decomposition will not be unique. This is analogous to the finite-dimensional case of vector spaces, $\mathbb{R}^n \simeq \prod_{j=1}^{n} \mathbb{R}[\mathbf{e}_j]$, where we may choose any set $\{\mathbf{e}_j \mid 1 \leqslant j \leqslant n\}$ of basis vectors. Yet, in the finite case, every connected graph has a unique prime factor decomposition with respect to the strong product defined in eqn (22) (Dörfler & Imrich, 1970; McKenzie, 1971). Polynomial time algorithms for computing the prime factor decomposition of an undirected graph are known (Feigenbaum & Schäffer, 1992; Imrich, 1998).

7.2. FACTORIZABILITY THEOREM—INFORMAL NARRATIVE

The problem with factorizing phenotype spaces is the identification of characters. Since descriptors of characters are often neither obvious nor simple, a notion of phenotypic dimension

that does not make a premature commitment with regard to the physical nature of a character seems desirable. Wagner & Laubichler (2000) propose to use certain partitions of the set of phenotypes (explained below) as representing such dimensions (and, thus, possible characters). A partition of a set X is a collection of disjoint subsets such that their union recovers X. Such a subset, then, corresponds to a character *state* (that is, a phenotype "value" along a character "axis"). These subsets are equivalence classes containing all phenotypes that agree on that character state. This is somewhat analogous to the notion of "schema" in genetic algorithm theory (Holland, 1993). For such characters to correspond to "dimensions", they must be "orthogonal". In fact, the partitions $\mathbf{P_i}$ considered in Wagner & Laubichler (2000) are orthogonal,§§§ that is, for each $P \in \mathbf{P_i}$ and each $Q \in \mathbf{P_j}$ there is a unique $x \in X$, such that $P \cap Q = \{x\}$.

An example from RNA shape space should help clarify these notions. The top of Fig. 9 depicts a shape consisting of two features, numbered 1 and 2. Suppose we decide to consider these two features as characters. The character states then consist of variations on these features, such as longer or shorter stacks and correspondingly shorter or longer loops. The characters 1 and 2 correspond to two partitions $\mathbf{P_1}$ and $\mathbf{P_2}$, respectively, that are sketched in Fig. 9. Take $\mathbf{P_1}$, for example. The columns correspond to disjoint subsets of RNA shape space in which the state of character 1 is fixed, while the rest of the shape varies. The first column of $\mathbf{P_1}$, for example, exhibits a particular manifestation (state) of character 1, distinct from column 2, and so on. The same holds for $\mathbf{P_2}$ and the states of character 2 (sketched at the bottom of Fig. 9). It is easy to see that $\mathbf{P_1}$ and $\mathbf{P_2}$ are orthogonal partitions of the set of RNA shapes. For example, consider the second column—label it P—in $\mathbf{P_1}$ and the first column—label is Q—in $\mathbf{P_2}$. The shape on the sequence segment reserved for character 2 is fixed in Q, while the shapes on the remaining segments

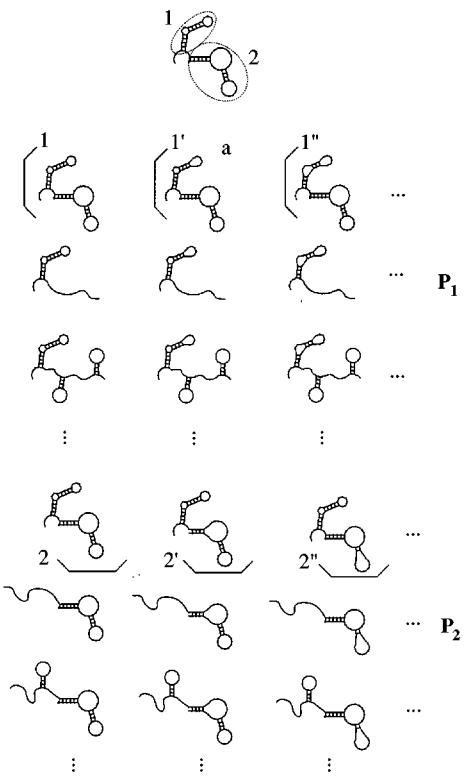


FIG. 9. Orthogonal partitions. See text for details.

vary. In set P, however, the shapes on segment 2 vary. Hence, there will be a phenotype in P whose shape on segment 2 coincides with the fixed shape on the same segment in Q. Conversely, there will be a phenotype in Q whose variable portion coincides with the fixed portion in P. Since the segments do not overlap, there will be exactly one phenotype in both P and Q that consists of the fixed P-state on segment 1 and the fixed Q-state on segment 2. Thus, $\mathbf{P_1}$ and $\mathbf{P_2}$ are orthogonal partitions. The same reasoning applies to any features—and any number of them—one wishes to choose.

The partitioning here really just slices the sequence into disjoint segments (two in our example) and declares any shape on a particular segment as a state of the same character. In this interpretation, a character is but a sequence segment (and a character state is a shape on that segment). This may be questionable. An alternative, more satisfactory interpretation is to think of a partition as a choice of "grain", that is, as setting the scale or size at which we choose to consider characters. An element of the partition

§§§Orthogonality is used here in a colloquial manner. In the context of partitions, the notion of orthogonality has to be generalized (Bailey, 1996). The correct term, defined and used in the formal section, is "orthogonally complementary".

(a column in Fig. 9) then corresponds to a possible character of that size (rather than a character state).

In the formal section below, we clarify the relationship between orthogonal partitions and the original phenotype space $(X, \mathcal{N})$. First, we show that if $(X, \mathcal{N})$ is factorizable, that is, if it has the same structure as the product of some factors (which need not be derived from the original space X), then there exist orthogonal partitions of X. This means that in studying the factorization of $(X, \mathcal{N})$, we might as well stick to (orthogonal) partitions derived from the original set of phenotypes. Second, using the partitions of X, we pass to the corresponding quotients and their quotient pre-topologies. That is, our units of analysis shift to the equivalence classes themselves (the columns in Fig. 9), rather than their elements. We then consider the (pre)-topology of the product of quotients and derive a simple condition under which this product has the same topological structure as the original space $(X, \mathcal{N})$. In other words, when this condition is fulfilled, we can faithfully represent the phenotype space in terms of phenotype "components" while preserving the original neighborhood structure. This nails the notion of *structural independence* alluded to above.

As an RNA example, consider the shape labelled *a* in Fig. 9 and shown in Fig. 10(a). Using the folding map and the techniques detailed in Fontana & Schuster (1998b) and Cupal *et al.* (2000), determine the shapes that are near *a* in the accessibility sense, that is, determine the vicinity of *a*, $N(a)$. Next, assign that shape to its image in the product space. Its image is given by the pair of equivalence classes determined by the orthogonal partitions $\mathbf{P_1}$ and $\mathbf{P_2}$ in Fig. 9, that is column 2 (set P) in $\mathbf{P_1}$ and column 1 (set Q) in $\mathbf{P_2}$. As representatives of these classes, we select the shapes with an unfolded configuration on the variable segment. Next, determine the vicinities of each equivalence class [proceeding in the same manner as for $N(a)$]. The vicinity of the pair, $N(P, Q)$, is given by the smallest set in eqn (21), provided we replace everything by equivalence classes. Next, we pass to the product space for each shape in $N(a)$. If RNA shape space is factorizable, we must recover $N(P, Q)$. This simply expresses a superposition principle for the variability of shape features: by analysing the variabilities of shape features separately and then combining them, we find the same accessibility structure as for the overall composite shape.

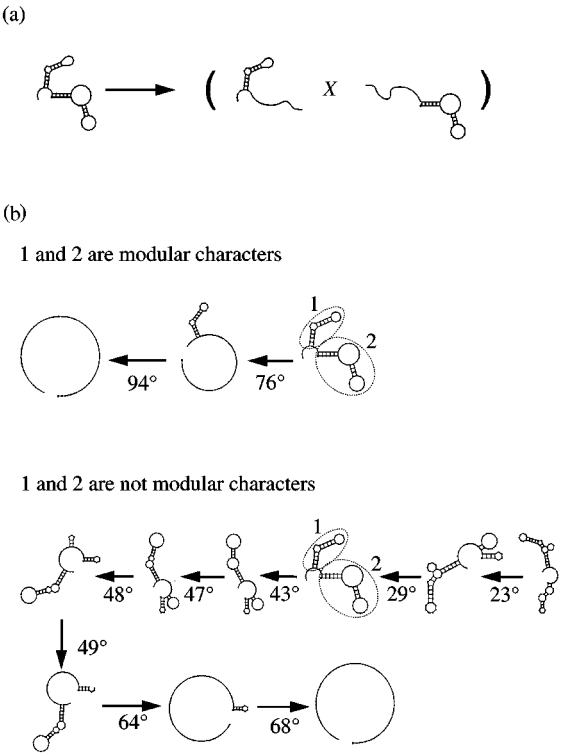


FIG. 10. Factorization and character. (a) An illustration of the product assigned to an RNA shape. (b) The figure shows the series of shape changes as the temperature is increased from 0 to 100°C for two sequences sharing the same shape at 37°C. The shape features of the first sequence are highly stable and react independently to changing temperature. This correlates with a high mutational stability of these features. Phenotypic changes upon mutation remain localized to a particular feature and are continuous in the sense of the accessibility pre-topology of Section 2.2. The same shape features are only marginally stable for the second sequence and melt in a highly interdependent fashion. Mutations easily destroy these features. For more details, see Ancel & Fontana (2000).

Product spaces emphasize characters as quasi-independent units of phenotypic variation. Changes in one character should not affect other characters in the same phenotype. Yet, two RNA sequences (genotypes) with the same phenotype can differ substantially in the degree of modularity of their shape features. This is illustrated in Fig. 10(b) which depicts the series of minimum free energy secondary structures at various temperatures for two sequences that share the same

phenotype at 37°C. The first series reveals a highly modular shape, in the sense that individual features remain stable across a large temperature range and melt independent of one another. This behavior correlates with a high degree of mutational stability (Ancel & Fontana, 2000). The opposite is the case for the second series in Fig. 10. Our definition of RNA shape space topology (Section 2.2) is based on the accessibility of one shape from another, say α, by averaging over all sequences that fold into α. However, the stability, and therefore autonomy, of phenotypic features is not an automatic property of a phenotype displaying these features. It rather depends on the underlying genotype. Which RNA shape transformations are continuous or discontinuous is determined by the accessibility likelihoods [eqn (1)] and a cutoff (Section 5.2). It is therefore important to consider the *local* properties of the genotype–phenotype map in the light of that topology. It turns out that the map is almost never continuous at a randomly chosen genotype. In contrast, the map is continuous at the specific genotype underlying the shape whose melting behavior is highly modular in Fig. 10(b). The inability to change a phenotype in a discontinuous fashion upon mutation is an indicator of phenotypic features that are sufficiently stabilized to warrant the label of characters. The framework proposed here would be right on target, if it turned out that the factorization condition (derived below) is in general not satisfied globally, but only in those regions of genotype space where the genotype–phenotype map is locally continuous (which are the regions associated with stable characters). The assessment of this possibility in the RNA case requires a computational analysis that constitutes a topic for future research.

7.3. FACTORIZABILITY THEOREM—FORMAL TREATMENT

Lemma 2. *Suppose* $(X, \mathcal{N})$ *is factorizable. Then there are two partitions* $\mathbf{P_1}$ *and* $\mathbf{P_2}$ *of* X *satisfying*

(i) *For each* $P_1 \in \mathbf{P_1}$ *and each* $P_2 \in \mathbf{P_2}$, *there is a unique* $x \in X$ *such that* $P_1 \cap P_2 = \{x\}$.

(ii) $\mathbf{P_1}$ *and* $\mathbf{P_2}$ *are non-trivial, i.e. they are neither the universal partition* $\{X\}$ *nor the identity partition* $\{\{x\} \mid x \in X\}$.

Proof. (i) Consider the isomorphism $\phi : (X, \mathcal{N}) \to (X_1, \mathcal{N}_1) \times (X_2, \mathcal{N}_2)$ and write $\phi_i = \mathbf{pr}_i \circ \phi$, $i = 1, 2$. Then

$$\mathbf{P_i} = \{\phi_i^{-1}(x_i) \mid x_i \in X_i\} \tag{24}$$

is a partition of X for $i = 1, 2$. Given $x_i \in X_i$, $i = 1, 2$, there is a unique $x \in X$ such that $\phi_1(x) = x_1$ and $\phi_2(x) = x_2$ since ϕ is invertible. In other words, there is a unique $x \in \phi_1^{-1}(x_1) \cap \phi_2^{-1}(x_2)$, and (i) follows.

(ii) First, we show that (i) implies that if one partition is universal, then the other one is the equality partition: Suppose $\mathbf{P}_1 = \{X\}$ is the universal partition. Then (i) implies $P_1 \cap P_2 = X \cap P_2 = P_2 = \{x\}$ for some $x \in X$, i.e. $\mathbf{P_2}$ is the equality partition. If $\mathbf{P_1}$ is the equality partition, then $P_1 \cap P_2 = \{x\} \cap P_2 = \{x\}$ for some $x \in X$ and all $P_2 \in \mathbf{P_2}$, hence $P_2 = X$ and $\mathbf{P_2}$ is the universal partition. If one of the partitions is trivial, one of them, say $\mathbf{P_1}$, is therefore the universal partition, and thus $\phi_1^{-1}(x_1) = X$ for all x_1. Since ϕ is invertible, this implies that X_2 consists of a single point, i.e. the product is trivial. □

A pair of partitions satisfying (i) in Lemma 2 is called *orthogonally complementary* (Wagner & Laubichler, 2000). For more general notions of orthogonality among partitions, we refer to Bailey (1996). Cast in this language, Lemma 2 states:

If $(X, \mathcal{N})$ *is factorizable, then there is a pair of non-trivial orthogonally complementary partitions of* X.

Next, we introduce some more notations. Let $(X, \mathcal{N})$ be a pre-topological space, and let $\mathbf{P}$ be a partition of X. For each $x \in X$, we denote the (equivalence) class to which x belongs with $[x]$. It is customary to write $X/\mathbf{P} = \{[x] \mid x \in X\}$. Moreover, for a set $M \subseteq X$, we write $[M] = \{[x] \mid x \in M\}$. Associated with $\mathbf{P}$ is the *canonical map*

$$\chi_{\mathbf{P}} : X \to X/\mathbf{P}, \quad x \mapsto [x], \tag{25}$$

which induces the *quotient pre-topology* on $X/\mathbf{P}$ with neighborhood systems

$$\mathcal{N}_{\mathbf{P}}[x] = \{[N] \mid N \in \mathcal{N}(x)\}. \tag{26}$$

This is the finest pre-topology on $X/\mathbf{P}$ such that the canonical map $\chi_{\mathbf{P}}$ is continuous (Fischer, 1959; Kent, 1969).

Lemma 3. *If $(X, \mathcal{N})$ is factorizable, then there is a pair of orthogonally complementary partitions $\mathbf{P_1}$ and $\mathbf{P_2}$ and pre-topologies $\mathcal{N}_1^{\psi}$ and $\mathcal{N}_2^{\psi}$ such that*

$$(X, \mathcal{N}) \simeq (X/\mathbf{P_1}, \mathcal{N}_1^{\psi}) \times (X/\mathbf{P_2}, \mathcal{N}_2^{\psi}). \quad (27)$$

Proof. Assuming factorizability and using the notation of Lemma 2, we see that there is a one-to-one correspondence between the elements $x_i \in X_i$ and the equivalence classes $\phi_i^{-1}(x_i) \in \mathbf{P_i}$ because ϕ is invertible. That is, the function

$$\psi_i : X_i \to X/\mathbf{P_i}, \quad x_i \mapsto \phi_i^{-1}(x_i)$$

is invertible. Defining the pre-topology on $X/\mathbf{P_i}$ by $\mathcal{N}_i^{\psi}(\psi_i(x_i)) = \psi_i(\mathcal{N}_i(x_i))$ implies that ψ_i is an isomorphism, i.e. $(X_i, \mathcal{N}_i) \simeq (X/\mathbf{P}_i, \mathcal{N}_i^{\psi})$ for $i = 1, 2$, and the lemma follows since the products of these two factor spaces are, of course, also isomorphic. □

By construction, the composition $\psi_i \circ \phi_i$ maps a point $x \in X$ onto its equivalence class $\phi_i^{-1}(\phi_i(x)) = [x]_i$, that is, $\psi_i \circ \phi_i = \chi_{\mathbf{P_i}}$. In order for $\chi_{\mathbf{P_i}}$ to be a projector of an isomorphism, it must be continuous by Fischer (1959, Theorem 13). Hence, the pre-topologies $\mathcal{N}_i^{\psi}$ must be coarser than the quotient pre-topologies, $\mathcal{N}_i^{\psi} \subseteq \mathcal{N}_{\mathbf{P_i}}$, $i = 1, 2$, since the quotient pre-topologies are the finest ones for which the characteristic maps $\chi_{\mathbf{P_i}}$ are continuous.

Lemma 3 shows that we can restrict ourselves to quotient maps and suitable pre-topologies on the sets $X/\mathbf{P_i}$. It appears natural to choose a pair of orthogonally complementary partitions $\mathbf{P_1}$ and $\mathbf{P_2}$ of X and to consider the pre-topological product space

$$(X/\mathbf{P_1} \times X/\mathbf{P_2}, \mathcal{N}_{\mathbf{P_1} \times \mathbf{P_2}}) = (X/\mathbf{P_1}, \mathcal{N}_{\mathbf{P_1}}) \times (X/\mathbf{P_2}, \mathcal{N}_{\mathbf{P_2}}). \quad (28)$$

The map

$$\xi : (X, \mathcal{N}) \to (X/\mathbf{P_1} \times X/\mathbf{P_2}, \mathcal{N}_{\mathbf{P_1} \times \mathbf{P_2}}),$$

$$x \mapsto ([x]_{\mathbf{P_1}}, [x]_{\mathbf{P_2}}) = (\chi_{\mathbf{P_1}}(x), \chi_{\mathbf{P_2}}(x)) \quad (29)$$

is invertible if and only if $\mathbf{P_1}$ and $\mathbf{P_2}$ on X are orthogonal complements. What we need to derive is when ξ is an isomorphism. As a first step, we need a more convenient characterization of the pre-topological structure of the product space.

Lemma 4. *The set*

$$[[\mathcal{N}(x)]] = \{[M]_{\mathbf{P_1}} \times [M]_{\mathbf{P_2}} \mid M \in \mathcal{N}(x)\} \quad (30)$$

is a filter basis of the neighborhood system $\mathcal{N}_{\mathbf{P_1} \times \mathbf{P_2}}([x]_1, [x]_2)$.

Proof. Let $N \in \mathcal{N}_{\mathbf{P_1} \times \mathbf{P_2}}(\alpha, \beta)$ with $\alpha \cap \beta = \{x\}$. By definition of the product pre-topology, there is $N_1 \in \mathcal{N}_{\mathbf{P_1}}(\alpha)$ and $N_2 \in \mathcal{N}_{\mathbf{P_2}}(\beta)$ such that $N_1 \times N_2 \subseteq N$. Since $\mathcal{N}_{\mathbf{P_1}}(\alpha)$ and $\mathcal{N}_{\mathbf{P_2}}(\beta)$ are, by definition, filter bases of the neighborhood filters in the quotient spaces, there exist sets, $M_1, M_2 \in \mathcal{N}(x)$ such that $[M_1]_{\mathbf{P_1}} \subseteq N_1$ and $[M_2]_{\mathbf{P_2}} \subseteq N_2$, respectively. From the filter axioms, we have $M = M_1 \cap M_2 \in \mathcal{N}(x)$ and hence $[M]_{\mathbf{P_1}} \in \mathcal{N}_{\mathbf{P_1}}(\alpha)$ and $[M]_{\mathbf{P_2}} \in \mathcal{N}_{\mathbf{P_2}}(\beta)$, and finally $[M]_{\mathbf{P_1}} \times [M]_{\mathbf{P_2}} \in \mathcal{N}_{\mathbf{P_1} \times \mathbf{P_2}}(\alpha, \beta)$. On the other hand, we have

$$[M]_{\mathbf{P_1}} \times [M]_{\mathbf{P_2}} \subseteq [M_1]_{\mathbf{P_1}} \times [M_2]_{\mathbf{P_2}} \subseteq N_1 \times N_2$$

$$\subseteq N \in \mathcal{N}_{\mathbf{P_1} \times \mathbf{P_2}}(\alpha, \beta)$$

$[[\mathcal{N}(x)]]$ is, therefore, a filter basis of the product space. □

Let us now consider the inverse map ξ^{-1}:

$$\xi^{-1} : (X/\mathbf{P_1} \times X/\mathbf{P_2}, \mathcal{N}_{\mathbf{P_1} \times \mathbf{P_2}}) \to X. \quad (31)$$

Define the pre-topology $\mathcal{N}^*$ on X as the final pre-topology generated by ξ^{-1}. By definition, the neighborhood filters $\mathcal{N}^*(x)$ have filter bases consisting of the sets

$$M^* = \xi^{-1}([M]_{\mathbf{P_1}} \times [M]_{\mathbf{P_2}})$$

$$= \{y \in X \mid \exists \alpha \in [M]_{\mathbf{P_1}}, \exists \beta \in [M]_{\mathbf{P_2}}, \text{ and}$$

$$[\alpha]_{\mathbf{P_1}} \cap [\beta]_{\mathbf{P_2}} = \{y\}\} \quad (32)$$

for each $M \in \mathcal{N}(x)$.

In sum, we first constructed product space (28) from orthogonal complement partitions on X, and then returned from the product space back to X via ξ^{-1}. Now, observe that

Lemma 5. $(X/\mathbf{P_1} \times X/\mathbf{P_2}, \mathcal{N}_{\mathbf{P_1}\times\mathbf{P_2}}) \simeq (X, \mathcal{N}^*)$.

Proof. We already know that $\xi^{-1}: X/\mathbf{P_1} \times X/\mathbf{P_2} \to X$ is invertible. Hence, it remains to show that $\xi = (\xi^{-1})^{-1}$ is continuous as map from $(X, \mathcal{N}^*)$ into the product space $(X/\mathbf{P_1} \times X/\mathbf{P_2}, \mathcal{N}_{\mathbf{P_1}\times\mathbf{P_2}})$. This follows immediately, however, since the image of the neighborhood basis $\{M^* \mid M \in \mathcal{N}(x)\}$ under ξ is just $\{\xi(M^*) = ([M]_{\mathbf{P_1}}, [M]_{\mathbf{P_2}}) \mid M \in \mathcal{N}(x)\} = [[\mathcal{N}(x)]]$, the neighborhood basis of the product space described in Lemma 4. □

The purpose of this exercise was to characterize the relationship between product space (28), or more precisely, the isomorphic pre-topological space $(X, \mathcal{N}^*)$, and the original space $(X, \mathcal{N})$. Note that ξ is an isomorphism if and only if $\mathcal{N}^* = \mathcal{N}$. Of course, this is not always the case. We show below, however, that the product space always has a coarser pre-topology than $(X, \mathcal{N})$. Formally,

Lemma 6. $\mathcal{N}^*(x) \subseteq \mathcal{N}(x)$.

Proof. There are two independent arguments to see this. (i) We can of course write $M \in \mathcal{N}(x)$ in the form

$$M = \{y \in X \mid [y]_{\mathbf{P_1}} \in [M]_{\mathbf{P_1}} \text{ and } [y]_{\mathbf{P_2}} \in [M]_{\mathbf{P_2}}\}.$$

Comparison with eqn (32) shows $M \subseteq M^*$, hence $\mathcal{N}(x)$ contains the filter basis $\{M^* \mid M \in \mathcal{N}(x)\}$ and the lemma follows. The second proof is even simpler: we simply note that $\xi: (X, \mathcal{N}) \to (X/\mathbf{P_1} \times X/\mathbf{P_2}, \mathcal{N}_{\mathbf{P_1}\times\mathbf{P_2}})$ and $\xi^{-1}: (X/\mathbf{P_1} \times X/\mathbf{P_2}, \mathcal{N}_{\mathbf{P_1}\times\mathbf{P_2}}) \to (X, \mathcal{N}^*)$ are continuous by construction. Hence, $\xi^{-1} \circ \xi = \mathrm{id}: (X, \mathcal{N}) \to (X, \mathcal{N}^*)$ is also continuous, implying that the pre-topology $\mathcal{N}^*$ is coarser than the pre-topology $\mathcal{N}$ on X. □

This result can be obtained in the even more general context of convergence spaces Carstens & Kent, (1969). We next seek conditions under which the neighborhood structure is preserved in the construction of the product space:

Lemma 7. *The following statements are equivalent*:

(i) $(X, \mathcal{N}) \simeq (X/\mathbf{P_1} \times X/\mathbf{P_2}, \mathcal{N}_{\mathbf{P_1}\times\mathbf{P_2}})$.
(ii) $(X, \mathcal{N}) \simeq (X, \mathcal{N}^*)$.
(iii) $\mathcal{N}(x) = \mathcal{N}^*(x)$ *for all* $x \in X$.
(iv) *For each* $x \in X$ *and each neighborhood* $N \in \mathcal{N}(x)$ *there is an* $M \in \mathcal{N}(x)$ *such that*

$$[M]_{\mathbf{P_1}} \times [M]_{\mathbf{P_2}} \subseteq \{([y]_{\mathbf{P_1}}, [y]_{\mathbf{P_2}}) \mid y \in N\}. \quad (33)$$

Proof. (i $\Leftrightarrow$ ii) follows from the transitivity of the isomorphy relation $\simeq$. (ii $\Leftrightarrow$ iii) is the fact that the identity $(X, \mathcal{N}) \to (X, \mathcal{N}^*)$ is continuous in both directions if and only if the pre-topologies coincide. Condition (iv), finally, is the set-wise rewriting of the condition $\mathcal{N}(x) \subseteq \mathcal{N}^*(x)$, which together with Lemma 6 implies (ii). □

Condition (iv) could be called the *rectangle condition*, since it requires that every neighborhood of $\mathcal{N}(x)$ contains the direct product $M_1 \times M_2$ of neighborhoods (a "rectangular" neighborhood) from the two quotient spaces.

We are now in a position to state the main technical result of this section:

Theorem 1. *A pre-topological space* $(X, \mathcal{N})$ *is factorizable if and only if there is a pair of non-trivial orthogonally complementary partitions that satisfy rectangle condition* (33).

Proof. It follows immediately from Lemma 7 that the condition is sufficient. Necessity follows from Lemmas 2 and 3, and the fact that we can rule out pre-topologies on $(X/\mathbf{P_i})$ that are strictly coarser than the quotient pre-topologies as candidates for factorizability: Their product $\mathcal{N}_1^{\psi} \times \mathcal{N}_2^{\psi}$ is then strictly coarser than $\mathcal{N}^*$ and therefore always strictly coarser than $\mathcal{N}$, contradicting that ξ is an isomrophism. □

So far, we have considered the factorization of the entire phenotype space. This appears to be too demanding a structure, as it would imply that the definition of a particular character is applicable to *all* phenotypes. A standard approach to

turn global concepts into local‖‖‖‖‖ ones leads to the following.

Definition. The pre-topological space $(X, \mathcal{N})$ is locally factorizable in $x \in X$ provided that for each neighborhood $N' \in \mathcal{N}(x)$, there is a neighborhood $N \subseteq N'$ such that the restriction $(N, \mathcal{N}_N)$ is factorizable.

The *trace* pre-topology $\mathcal{N}_N$ is defined by $\mathcal{N}_N(y) = \{N' \cap N \mid N' \in \mathcal{N}(y)\}$ for each $y \in N$. A *finite* pre-topological space is locally factorizable in x if and only if the subgraph induced by the vicinity $N(x)$ has a non-trivial prime-factor decomposition.

8. Discussion

The products of evolution are shaped by both the dynamics of selection in populations and the attainability of variants that selection can act upon. The mechanisms underlying the construction of phenotype from genotype mediate phenotypic innovation. This mediation, however, is biased. Even if mutational mechanisms generate genetic variation at random, the phenotypic consequences need not be random, for they depend on the genetic context in which a mutation is expressed. Underlying this bias is ultimately a genotype–phenotype relation that is strongly many-to-one, thereby inducing a non-trivial relationship of mutational accessibility among phenotypes. In Section 2.2.1, we illustrated this issue with the folding of RNA sequences (representing genotypes) into secondary structures (representing phenotypes).

When organized in terms of mutational accessibility, the set of possible RNA secondary structures lacks a metric and exhibits an unfamiliar topological structure in which a phenotype may be near another without implying the reverse. Yet, this weakly structured space naturally addresses some classic patterns of phenotypic evolution, such as punctuation and irreversibility and neatly expresses the consequences of folding constraints ("developmental constraints"). We presented a formal language that abstracts the insights obtained from the specific RNA case. We hope that this apparatus helps in preventing a bias in thinking and facilitates further questions that are relevant across diverse and complex genotype–phenotype relations.

The appropriate formal structure for RNA shape space is a pre-topology, a space whose neighborhoods are, in general, not related to open sets and thus represents a weaker structure than a topology. In Section 4, we reviewed this notion along with the necessary technical accessories for connecting pre-topological neighborhoods with continuity. While this material is well known to topologists, the original literature may not be easily accessible to a theoretical biologist. The section contains some original material when connecting finite (pre)-topological structures with graphs.

In Section 5, we applied these concepts to characterize the finest and coarsest accessibility pre-topologies resulting from the requirement that the genotype–phenotype map be everywhere continuous. In practice, accessibility of a phenotype from another is a likelihood (Fontana & Schuster, 1998b), see eqn (1). That likelihood is converted into a binary attribute by choosing a cutoff value below which a phenotype is regarded as inaccessible from the other. This leads to a range of pre-topologies with various degrees of discontinuity. Once phenotypes are equipped with a (pre)-topological structure, continuity for evolutionary trajectories can be meaningfully defined (Section 6). Since the fate of a phenotype in a population is constrained by selection, a locally highly discontinuous map can nevertheless give rise to locally continuous evolutionary trajectories (Fontana & Schuster, 1998a, b). The accessibility topology may be such that no continuous path exists between two particular phenotypes. The dynamical signature of such irreducible discontinuities is punctuation—long periods of phenotypic stasis with underlying drift in genotype space until a genetic context arises that enables the transition to occur in response to a small mutation. By definition, phenotypic accessibility topologies depend on what constitutes

‖‖‖‖‖This notion of locality is to be distinguished from the one intended in the final paragraph of the informal narrative. There, we meant a factorization of phenotype space as induced by the genotype–phenotype map in a restricted region of genotype space.

a likely mutation. In our RNA examples, we have always assumed accessibility to be relative to a single point mutation.

Given a phenotypic accessibility topology, we may ask whether it sustains a notion akin to "dimension" that can be related to the notion of "character" or "module". We explored this issue in Section 7 as the factorizability of pre-topologies into product spaces. This led us to define characters in terms of *structural* independence which is related to Lewontin's notion of quasi-independence (Lewontin, 1978). The notion of units as independent entities is always relative to a process with respect to which this independence is expressed, such as natural selection (Kim & Kim, 2001; Wagner *et al*., 2000) or the covariation of mutational effects (Lande, 1980; Houle, 2001). Structural independence is independence with respect to accessibility and thus expresses the most primitive notion of independence. Factorizability of a topology asserts the existence of phenotypic units that *can* be varied independently (that is, they do not impose constraints of accessibility on each other), it does not imply that they always have to vary independently under natural selection. The main result of Section 7 is a factorizability condition such that the product space remains isomorphic to the original, non-factorized accessibility topology. This condition may not be satisfied globally. In fact, our experience with the RNA map leads us to conjecture that the factorizability condition is satisfied only in special regions of genotype space—those in which the genotype–phenotype map is locally continuous. Equivalently, regions of genotype space with discontinuous phenotypic transitions cannot be factorized, which fits the intuition that innovation, that is, a sudden change in phenotype, requires the loss of character identity. We have not carried out the substantial computations required to corroborate this conjecture in RNA, but circumstantial evidence is presented in Ancel & Fontana (2000). If factorization is local, then different factorizations may exist in different parts of genotype space.

The topological framework presented here has direct consequences for the developmental explanation of major evolutionary transitions. Two kinds of such transitions can be distinguished. One refers to the integration of lower level replicators into higher level replicators, like the transition from single-celled organisms to multicellular forms (Buss, 1987; Maynard-Smith & Eörs Szathmáry, 1995; Michod, 1999). The other refers to the evolution of major multicellular body plans, as they originate from the differentiation of cell colonies, or the origin of new body parts (Müller & Wagner, 1991; Raff, 1996). The present discussion addresses the second kind of transition.

The origination of body parts or body plans is a rare evolutionary event and constitutes a plausible instance of a discontinuous transition in the sense of the present paper. From a topological point of view, a transition from an ancestral character to a derived one is discontinuous ("difficult to achieve") if the neutral network of the former is adjacent to the neutral network of the latter in only a small fraction of its boundary, that is, if many genotypic states are able to realize the ancestral state, but only a small fraction of them are poised to generate the derived phenotype. This implies that the actual transition to the derived phenotype occurred in a genotypic state that is unlikely to be realized in any extant species. This has implications for our ability to experimentally dissect the molecular mechanisms underlying morphological innovations. A major goal of evolutionary developmental biology is to understand how changes in the genetic regulation of development gave rise to new morphological characters and body plans (Hall, 1998; Raff, 1996; Wagner *et al*., 2000). There is considerable progress in identifying key molecular differences underlying morphological innovations, like the origin of butterfly eye spot patterns (Keys *et al*., 1999) or the origin of flower organs (Kramer & Irish, 1999), but the number of examples is small. The most stringent experimental test for a molecular explanation of a morphological innovation would be the induction of the derived character in a species representing the ancestral phenotype by genetic manipulation. If, however, the topological explanation of discontinuous evolutionary transitions is correct, this test cannot be done, since the right genetic constellation poised for the transition is extremely unlikely to exist in any extant species bearing the ancestral phenotype. Moreover, a genetic manipulation intended to mimic the molecular mechanism that

historically led to the derived phenotype is likely to have different effects in different genetic backgrounds. That is, a molecular developmental mechanism that can give rise to an innovation should not be expected to cause the derived phenotype in every genotype representing the ancestral phenotype.

A further mechanistic consequence of the topological interpretation of major evolutionary transitions is that natural selection does not provide a complete explanation for their occurrence. Natural selection is a sufficient explanation for the outcome of an evolutionary process, if the genetic variation contributing to the derived phenotype is easily accessible. Whether a transition occurs, then only depends (to a first approximation) on the strength and direction of selection. If major transitions, however, require specially poised genotypic/developmental realizations of the ancestral phenotype, then the transition critically depends on factors not under the control of selection, since different genetic realizations of the same ancestral phenotype lie on a neutral network and are not distinguishable by selection on phenotypes.

Pre-topological concepts might, in principle, find an empirical application in connection with the reconstruction of phenotype spaces or related constructs, such as a morphospace (Raup & Michelson, 1965; Raup, 1966; McGhee, 1999; Eble, 2000). A pre-topological neighborhood system can be generated from a so-called sub-basis, that is, a set of accessibility relations that do not necessarily fulfill all the axioms of a neighborhood system. A sample of possible relationships of genetic accessibility among phenotypes could be obtained from a phylogenetic reconstruction of character transformations (Donoghue, 1989), or from morphometric, phenetic and developmental considerations. Of course, the observed transformations will rarely, if ever, reflect all possible phenotypic transformations and hence provide at most a sub-basis for the construction of the pertinent topology. From a sub-basis, a neighborhood system can be constructed by adding those sets that are implied by (N2) and (N3), Section 4.3. In other words, one adds all the intersections among sets in the sub-basis, as well as all their supersets. This embedding process could be used to construct the coarsest (local) pre-topology consistent with the empirical data on accessibility. It does not presuppose a metric, but it does not exclude one either.

We dedicate this paper to Peter Schuster in celebration of his 60th birthday, B.M.R.S. is supported by the Austrian *Fonds zur Förderung der Wissenschaftlichen Forschung*, Project No. P-13887-MOB. The work at the Santa Fe Institute was supported by the Keck Foundation and by core grants from the John D. and Catherine T. MacArthur Foundation, the National Science Foundation, and the U.S. Department of Energy. W.F. and his research program are supported by Michael A. Grantham.

REFERENCES

ALBERCH, P. & GALE, E. A. (1983). Size dependence during the development of the amphibian foot. Colchicine-induced digital loss and reduction. *J. Embryol. Exp. Morph.* **76,** 177–197.

ALBERCH, P. & GALE, E. A. (1985). A developmental analysis of evolutionary trend: digital reduction in amphibians. *Evolution* **39,** 8–23.

ALBUQUERQUE, J. (1941). La notion de "frontiere" en topologie. *Portugaliae Math.* **2,** 280–289.

ALEXANDROFF, P. & HOPF, H. (1935). *Topologie.* Berlin: Springer.

ANCEL, L. & FONTANA, W. (2000). Plasticity, evolvability and modularity in RNA. *J. Exp. Zool.* (*Mol. Dev. Evol.*) **288,** 242–283.

BADARD, R. (1984). Fuzzy preuniform structures and the structures they induce: 1. Main results. *J. Math. Anal. Appl.* **100,** 530–548.

BAILEY, R. A. (1996). Orthogonal partitions in designed experiments. *Designs, Codes Cryptogr.* **8,** 45–77.

BIESTERFELDT, H. J. (1968). Uniformizations of convergence spaces. *Math. Ann.* **177,** 31–42.

BULL, K. J. & CHARNOV, E. L. (1985). On irreversible evolution. *Evolution* **39,** 1149–1155.

BUSS, L. W. (1987). *The Evolution of Individuality*. Princeton, NJ: Princeton University Press.

CARSTENS, A. M. & KENT, D. C. (1969). A note on products of convergence spaces. *Math. Ann.* **182,** 40–44.

CARTAN, H. (1937). Theorie des filtres. *C. R. Acad. Sci., Paris* **205,** 595–598.

ČECH, E. (1966). *Topological Spaces.* London: Wiley.

COOK, C. H. & FISCHER, H. R. (1967). Uniform convergence structures. *Math. Ann.* **173,** 290–306.

CUPAL, J., KOPP, S. & STADLER, P. F. (2000). RNA shape space topology. *Artif. Life* **6,** 3–23.

DONOGHUE, M. (1989). Phylogenies and the analysis of evolutionary sequences, with examples from seed plants. *Evolution* **43,** 1137–1156.

DÖRFLER, W. & IMRICH, W. (1970). Über das starke Produkt von endlichen Graphen. *Österreich. Akad. Wiss. Math.-Natur. Kl. S.-B. II* **178,** 247–262.

EBLE, G. J. (2000). Theoretical morphology: state of the art. *Paleobiology* **26,** 498–506.

ELDREDGE, N. & GOULD, S. J. (1972). no title. In: *Models in Paleobiology* (Schopf, T. J. M., ed.), pp. 82–115. San Francisco: Freeman.

FEIGENBAUM, J. & SCHÄFFER, A. A. (1992). Finding the prime factors of strong direct products of graphs in polynomial time. *Discr. Math.* **109,** 77–102.

FISCHER, H. R. (1959). Limesräume. *Math. Ann.* **137,** 269–303.

FITCH, W. M. (2000). Homology, a personal view on some of the problems. *Trends Genet.* **16,** 227–231.

FONTANA, W. & SCHUSTER, P. (1998a). Continuity in evolution: on the nature of transitions. *Science* **280,** 1451–1455.

FONTANA, W. & SCHUSTER, P. (1998b). Shaping space: the possible and the attainable in RNA genotype–phenotype mapping. *J. theor. Biol.* **194,** 491–515.

FORCE, A., LYNCH, M., PICKETT, F. B., AMORES, A., YAN, Y. L. & POSTLETHWAIT, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151,** 1531–1545.

FORST, C. V. (2000). Molecular evolution of catalysis. *J. theor. Biol.* **205,** 409–431.

FRASER, S. M. & REIDYS, C. (1997). Evolution of random catalytic networks. In: *Fourth European Conference on Artificial Life* (Husbands, P. & Harvey, I., eds). Cambridge, MA: MIT Press.

FUTUYMA, D. J. (1998). *Evolutionary Biology*. Sunderland, MA: Sinauer Associates.

GAAL, S. A. (1964). *Point Set Topology*. New York: Academic Press.

GALIS, F. (2001). Key innovations and radiations. In: *The Character Concept in Evolutionary Biology* (Wagner, G. P., ed.), pp. 581–605. San Diego, CA: Academic Press.

GITCHOFF, P. & WAGNER, G. P. (1996). Recombination induced hypergraphs: a new approach to mutation–recombination isomorphism. *Complexity* **2,** 37–43.

GÖBEL, U., FORST, C. V. & SCHUSTER, P. (1997). Structural constraints and neutrality in RNA. In: *Proceedings of the German Conference in Bioinformatics* 1996 (Lengauer, T., Löffler, M. & Schomburg, D., eds). *Lecture Notes in Computer Science*, vol. 1278, pp. 156–165. Berlin: Springer.

GRANDEL, H. & SCHULTE-MERKER, S. (1998). The development of the paired fins in zebrafish (*danio rerio*). *Mech. Dev.* **79,** 99–120.

GRAUR, D. & LI, W.-H. (2000). *Fundamentals of Molecular Evolution*. Sunderland, MA: Sinauer Associates.

GRÜNER, W., GIEGERICH, R., STROTHMANN, D., REIDYS, C., WEBER, J., HOFACKER, I. L., STADLER, P. F. & SCHUSTER, P. (1996a). Analysis of RNA sequence structure maps by exhaustive enumeration. I. Neutral networks. *Monatsh. Chem.* **127,** 355–374.

GRÜNER, W., GIEGERICH, R., STROTHMANN, D., REIDYS, C., WEBER, J., HOFACKER, I. L., STADLER, P. F. & SCHUSTER, P. (1996b). Analysis of RNA sequence structure maps by exhaustive enumeration. II. Structure of neutral networks and shape space covering. *Monatsh. Chem.* **127,** 375–389.

HALL, B. K. (1998). *Evolutionary Developmental Biology*, 2nd Edn. London: Chapman & Hall.

HERRLICH, H. (1974). A concept of nearness. *Gen. Topol. Appl.* **5,** 191–212.

HOFACKER, I. L., FONTANA, W., STADLER, P. F., BONHOEFFER, S., TACKER, M. & SCHUSTER, P. (1994). Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.* **125,** 167–188.

HOLLAND, J. H. (1993). *Adaptation in Natural and Artificial Systems*. Cambridge, MA: MIT Press.

HOULE, D. (2001). Characters as units of evolutionary change. In: *The Character Concept in Evolutionary Biology* (Wagner, G. P., ed.), pp. 109–140. San Diego, CA: Academic Press.

HUGHES, A. L. (1994). The evolution of functionally novel proteins after gene duplication. *Proc. Roy. Soc. Lond. B* **256,** 119–124.

HUYNEN, M. A., STADLER, P. F. & FONTANA, W. (1996). Smoothness within ruggedness: the role of neutrality in adaptation. *Proc. Natl Acad. Sci. U.S.A.* **93,** 397–401.

IMRICH, W. (1998). Factoring cardinal product graphs in polynomial time. *Discr. Math.* **192,** 119–144.

IMRICH, W. & KLAVŽAR, S. (2000). *Product Graphs: Structure and Recognition*. New York: Wiley.

KELLER, H. H. (1968). Die Limes-Uniformisierbarkeit der Limesräume. *Math. Ann.* **176,** 334–341.

KENT, D. C. (1968). A representation theorem for uniform convergence spaces. *Math. Ann.* **179,** 42–46.

KENT, D. C. (1969). Convergence quotient maps. *Fund. Math.* **65,** 197–205.

KEYS, D. N., LEWIS, D. L., SELEGUE, J. E., PEARSON, B. J., GOODRICH, L. V., JOHNSON, R. L., GATES, J., SCOTT, M. P. & CARROLL, S. B. (1999). Recruitment of hedgehog regulatory circuit in butterfly eyespot evolution. *Science* **283,** 532–534.

KIM, J. & KIM, M. (2001). The mathematical structure of characters and modularity. In: *The Character Concept in Evolutionary Biology* (Wagner, G. P., ed.), pp. 215–236. San Diego: Academic Press.

KRAMER, E. M. & IRISH, V. F. (1999). Evolution of genetic mechanisms controlling petal development. *Nature* **399,** 144–148.

LANDE, R. (1980). The genetic covariance between characters maintained by pleiotropic mutations. *Genetics* **94,** 203–215.

LENSKI, R. E. & TRAVISANO, M. (1994). Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc. Natl Acad. Sci. U.S.A.* **91,** 6808–6814.

LEWONTIN, R. C. (1974). *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press, New York.

LEWONTIN, R. C. (1978). Adaptation. *Sci. Am.* **239,** 156–169.

MAYNARD-SMITH, J. (1983). The genetics of punctuation and stasis. *Ann. Rev. Genet.* **17,** 11–25.

MAYNARD-SMITH, J., BURIAN, R., KAUFFMAN, S. A., ALBERCH, P., CAMPBELL, J., GOODWIN, B., LANDE, R., RAUP, D. & WOLPERT, L. (1985). Developmental constraints and evolution. *Quart. Rev. Biol.* **60,** 265–287.

MAYNARD-SMITH, J. & SZATHMÁRY, EÖRS (1995). *The Major Transitions in Evolution*. Oxford: W. H. Freeman.

MCGHEE, G. R. (1999). *Theoretical Morphology: The Concept and its Applications*. New York: Columbia University Press.

MCKENZIE, R. (1971). Cardinal multiplication of structures with a reflexive multiplication. *Fund. Math.* **70,** 59–101.

MICHOD, R. E. (1999). *Darwinian Dynamics*. Princeton, NJ: Princeton University Press.

MORDESON, J. N. & NAIR, P. S. (1998). *Fuzzy Mathematics: An Introduction for Engineers and Scientists*. Heidelberg, New York: Physica-Verlag.

MÜLLER, G. B. & WAGNER, G. P. (1991). Novelty in evolution: restructuring the concept. *Annu. Rev. Ecol. Syst.* **22,** 229–256.

NUSSINOV, R. & JACOBSON, A. B. (1980). Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc. Natl Acad. Sci. U.S.A.* **77,** 6309–6313.

OHNO, S. (1970). *Evolution by Gene Duplication*. New York: Springer-Verlag.
PERVIN, W. J. (1962a). Quasi-uniformization of topological spaces. *Math. Ann.* **147,** 316–317.
PERVIN, W. J. (1962b). Uniformization of neighborhood axioms. *Math. Ann.* **147,** 313–315.
PERVIN, W. J. (1963). Quasi-proximities for topological spaces. *Math. Ann.* **150,** 325–326.
PREUß, G. (1995). Semiuniform convergence spaces. *Math. Japonica* **41,** 465–491.
QUINT, E., ZERUCKA, T. & EKKER, M. (2000). Differential expression of orthologous *dlx* genes in zebrafish and mice: implications for the evolution of *dlx* homeobox gene family. *J. Exp. Zool.* (*Mol. Dev. Evol.*) **288,** 235–241.
RAFF, R. (1996). *The Shape of Life*. Chicago, IL: Chicago University Press.
RAUP, D. M. (1966). Geometric analysis of shell coiling: general problems. *J. Paleontol.* **40,** 1178–1190.
RAUP, D. M. & MICHELSON, A. (1965). Theoretical morphology of the coiled shell. *Science* **147,** 1294–1295.
REIDYS, C., FORST, C. V. & SCHUSTER, P. (2001). Replication and mutation on neutral networks. *Bull. Math. Biol.* **63,** 57–94.
REIDYS, C., STADLER, P. F. & SCHUSTER, P. (1997). Generic properties of combinatory maps—neutral networks of RNA secondary structures. *Bull. Math. Biol.* **59,** 339–397.
RICHARDSON, G. D. & KENT, D. C. (1996). Probabilistic convergence spaces. *J. Austral. Math. Soc. A* **61,** 400–420.
RIEDL, R. (1978). *Order in Living Organisms*: *A Systems Analysis of Evolution*. New York, NY: Wiley.
ROFF, D. A. (1997). *Evolutionary Quantitative Genetics*. New York: Chapman & Hall.
ŠANIN, N. (1943). On separation in topological space. *Dok. Akad. Nauk SSSR* **38,** 110–113.
SCHLICHTING, C. D. & PIGLIUCCI, M. (1998). *Phenotypic Evolution*: *A Reaction Norm Perspective*. Sunderland, MA: Sinauer Associates, Inc.
SCHULTES, E. A. & BARTEL D. P. (2000). One sequences, two ribozymes: Implications for the emergence of new ribozyme folds. *Science*, **289,** 448–452.
SCHUSTER, P. (1997). Landscapes and molecular evolution. *Physica D* **107,** 351–365.
SCHUSTER, P., FONTANA, W., STADLER, P. F. & HOFACKER, I. (1994). From sequences to shapes and back: a case study in RNA secondary structures. *Proc. Roy. Soc.* (*Lond.*). *B* **255,** 279–284.
SCHWENK, K. (1995). A utilitarian approach to evolutionary constraint. *Zoology* **98,** 251–262.
SHPAK, M. & WAGNER, G. P. (2000). Asymmetry of configuration space induced by unequal crossover: implications for a mathematical theory of evolutionary innovation. *Artif. Life* **6,** 25–43.
SHUBIN, N. & ALBERCH, P. (1986). A morphogenetic approach to the origin and basic organization of the tetrapod limb. *Evol. Biol.* **20,** 319–387.
SPIEGELMAN, S. (1971). An approach to experimental analysis of precellular evolution. *Quart. Rev. Biophys.* **4,** 213–253.
STADLER, P. F., SEITZ, R. & WAGNER, G. P. (2000). Population dependent Fourier decomposition of fitness landscapes over recombination spaces: evolvability of complex characters. *Bull. Math. Biol.* **62,** 399–428.
STADLER, P. F. & WAGNER, G. P. (1998). The algebraic theory of recombination spaces. *Evol. Comput.* **5,** 241–275.
SZOSTAK, J. W. & ELLINGTON, A. D. (1993). *In Vitro selection of functional RNA sequences*. In: *The RNA World* (Gesteland, R. F. & Atkins, J. F., eds), pp. 511–533. Plainview, NY: Cold Spring Harbor Laboratory Press.
TURNER, D. H., SUGIMOTO, N. & FREIER, S. (1988). RNA structure prediction. *Ann. Rev. Biophys. Biophys. Chem.* **17,** 167–192.
WAGNER, G. P. (1989a). The biological homology concept. *Ann. Rev. Ecol. Syst.* **20,** 51–69.
WAGNER, G. P. (1989b). The variance allocation hypothesis of stasis and punctuation. In: *Molecular Biology and Organisms* (Hoyningen-Huene, P. & Wuketits, F. M., eds.), pp. 161–185. Boston: Reidel.
WAGNER, G. P. (1994). Homology and the mechanisms of development. In: *Homology*: *The Hierarchical Basis of Comparative Biology* (Hall, B. K., ed.), pp. 273–299. San Diego, CA: Academic Press.
WAGNER, G. P. (1999). A research programme for testing the biological homology concept. In: *Homology* (Bock, G. R. & Cardew, G., eds), pp. 125–134. New York: John Wiley.
WAGNER, G. P. & ALTENBERG, L. (1996). Complex adaptations and the evolution of evolvability. *Evolution* **50,** 967–976.
WAGNER, G. P., CHIU, C.-H. & LAUBICHLER, M. D. (2000). Developmental evolution as a mechanistic science: the inference from developmental mechanisms to evolutionary processes. *Am. Zool.* **40,** 108–120.
WAGNER, G. P. & LAUBICHLER, M. D. (2000). Character indentification in evolutionary biology: the role of the organism. *Theory Biosci.* **119,** 20–40.
WALSH, J. B. (1995). How often do duplicated genes evolve new functions? *Genetics* **139,** 421–428.
WALTER, A. E., TURNER, D. H., KIM, J., LYTTLE, M. H., MULLER, P., MATHEWS, D. H. & ZUKER, M. (1994). Coaxial stacking of helices enhances binding of oligoribonucleotides and improves prediction of RNA folding. *Proc. Natl Acad. Sci. U.S.A.* **91,** 9218–9222.
WATERMAN, M. S. (1978). Secondary structure of single-stranded nucleic acids. *Studies on Foundations and Combinatorics*, *Advances in Mathematics Supplementary Studies*, Vol. 1, pp. 167–212. New York: Academic Press.
WEBER, J. (1997). Dynamics of neutral evolution. Ph.D. Dissertation, Friedrich-Schiller-Universität Jena, Germany. Available electronically from http://www.tbi.univie.ac.at.
WEIL, A. (1937). *Sur les espaces à structures uniformes et sur la topologie générale*. Paris: Hermann.
WHYBURN, G. T. (1970). Accessibility space. *Proc. Am. Math. Soc.* **24,** 181–185.
WYLER, O. (1974). Filter space monads, regularity, completions. In: *Topo 72—General Topology and its Applications, Lecture Notes in Mathematics*, Vol. 378, pp. 591–637. Berlin: Springer.
ZUKER, M. & STIEGLER, P. (1981). Optimal computer folding of larger RNA sequences using thermodynamics and auxiliary information. *Nacl. Acids Res.* **9,** 133–148.

APPENDIX A: Pre-topological Spaces

A.1. PRE-UNIFORMIZATION

This lemma states that every pre-topological space has a pre-uniformization.

Lemma A.1. *The neighborhood structure $\mathcal{N}_{\mathcal{U}_\mathcal{N}}$ induced by the pre-uniformization $\mathcal{U}_\mathcal{N}$ of $\mathcal{N}$ coincides with $\mathcal{N}$.*

Proof. Since $U' = \{(x, y) | x \in X, y \in N_x\} \in \mathcal{U}_\mathcal{N}$ for any choice $N_x \in \mathcal{N}(x)$, we obtain every neighborhood N_x in the form $N_x = \{(x, y) | y \in N_x\}[x] = U'[x]$. By construction, each set $U \in \mathcal{U}_\mathcal{N}$ is a superset of a U', hence $U[x]$ is a superset of some $N_x \in \mathcal{N}(x)$, and therefore a neighborhood of x. □

A more abstract treatment of the relationships between pre-topological spaces and various types of uniform structures can be found in Weil (1937), Cook & Fischer (1967), Biesterfeldt (1968), Kent (1968), Keller (1968), Wyler (1974) and Preuß (1995). In particular, Lemma A.1 is Theorem 8 by Biesterfeldt (1968).

A.2. CONDITION (UB) AND OPEN SETS

We rephrase (UB), Section 4.5, by projecting down to neighborhoods (the pre-topology induced by $\mathcal{U}$) according to eqn (3): For each $U[x] \in \mathcal{U}[x]$ $(= \mathcal{N}(x))$, there is a $V[x] \in \mathcal{U}[x]$ such that $(V \circ V)[x] \subseteq U[x]$. The definition of concatenation (5) implies that $V \subseteq V \circ V$ and hence $V[x] \subseteq (V \circ V)[x] \subseteq U[x]$. More explicitly:

$$(V \circ V)[x] = \{y \mid \exists z \in V[x] \text{ and } y \in V[z]\} = \bigcup_{z \in V[x]} V[z]. \tag{A.1}$$

Write $N = U[x]$, $N' = V[x]$, $N'' = (V \circ V)[x]$. By eqn (A.1), N'' is the union of neighborhoods of the elements in N'. Since N'' contains a neighborhood of each $y \in N'$, it constitutes by (N3), a neighborhood for each $y \in N'$, that is, $N'' \in \mathcal{N}(y)$. We also have $N'' \subseteq N$, hence $N \in \mathcal{N}(y)$ for all $y \in N'$. Thus, (UB) has the following implication for neighborhoods:

(N4) For each $N \in \mathcal{N}(x)$, there is an $N' \in \mathcal{N}(x)$ such that $N \in \mathcal{N}(y)$ for all $y \in N'$.

Further material on the relation between (quasi-)uniformities and neighborhood axioms can be found in Pervin (1962a, b).

A.3. PRE-UNIFORMITIES AND CONTINUITY

Let $f: X \to Y$ and $U \subset X \times X$. With the notation

$$(f \times f)(U) = \{(f(x), f(y)) | (x, y) \in U\} \tag{A.2}$$

we define uniform continuity in the following way:

Definition. Let $f: (X, \mathcal{U}) \to (Y, \mathcal{V})$, where $\mathcal{U}$ and $\mathcal{V}$ are pre-uniformities on X and Y, respectively. Then f is *uniformly continuous* if

$$\forall V \in \mathcal{V} : \exists U \in \mathcal{U} : (f \times f)(U) \subseteq V. \tag{A.3}$$

Theorem A.1. *If $f: (X, \mathcal{U}) \to (Y, \mathcal{V})$ is uniformly continuous, then $f: (X, \mathcal{N}_\mathcal{U}) \to (Y, \mathcal{M}_\mathcal{V})$ is continuous with respect to the induced pre-topologies on X and Y.*

Proof. Recall that $N_x \in \mathcal{N}_\mathcal{U}(x)$ if and only if there is $U \in \mathcal{U}$ such that $U[x] \subseteq N_x$. Thus, f is continuous with respect to $\mathcal{N}_\mathcal{U}$ and $\mathcal{M}_\mathcal{V}$, if for each $V \in \mathcal{V}$ there is a $U \in \mathcal{U}$ such that $f(U[x]) \subset V[f(x)]$. Assume that f is uniformly continuous, that is, for each $V \in \mathcal{V}$, there is a $U \in \mathcal{U}$ such that $(f \times f)(U) \subseteq V$. Hence,

$$\begin{aligned}(f \times f)(U)[f(x)] &= \{(f(x), f(y)) | (x, y) \in U\}[f(x)] \\ &= \{f(y) | (x, y) \in U\} \\ &= \{f(y) | y \in U[x]\} = f(U[x]) \\ &\subseteq V f[(x)],\end{aligned}$$

and the theorem follows. □

A.4. THE EQUIVALENCE OF AXIOMS (R0) AND (S′)

Theorem A.2. *If $(X, \mathcal{N})$ is a pre-topological space then*
(R0) *$x \in \overline{\{y\}}$ implies $y \in \overline{\{x\}}$ for all $x, y \in X$.*
(S′) *$x \in \bigcap \mathcal{N}(y)$ implies $y \in \bigcap \mathcal{N}(x)$.*
are equivalent.

Proof. We first observe

$$\overline{\{y\}} = \{z \mid \forall N \in \mathcal{N}(z): \{y\} \cap N \neq \emptyset\}$$

$$= \{z \mid \forall N \in \mathcal{N}(z): y \in N\}$$

$$= \{z \mid y \in \cap \mathcal{N}(z)\}.$$

Therefore, $x \in \overline{\{y\}}$ is equivalent to $y \in \cap \mathcal{N}(x)$, and the theorem follows immediately. □

A.5. AXIOM SYSTEMS FOR PRE-TOPOLOGICAL SPACES

Pre-topological spaces can be defined in a variety of seemingly unrelated, yet equivalent ways. In the main text, we have focussed on neighborhood as the basic concept. The neighborhood operator $\mathcal{N}$ assigns the collection $\mathcal{N}(x) \subseteq \mathfrak{P}(X)$ to each point $x \in X$. Alternative constructions make use of *Closure*, *Interior*, and *Boundary* operators which assign to each set $A \subset X$ its closure $\bar{A}$, its interior A°, and its boundary ∂A, respectively. These operators are related with one another in a fairly simple way, summarized in Table A1.

The closure, interior and boundary constructions can be used to uniquely specify a pre-topological space if and only if the axioms (1)–(3) in Table A2 are satisfied. In each case, the fourth axiom (T) is a necessary and sufficient condition for the pre-topology to be a topology.

A.6. INTERSECTION AND UNION OF FILTERS

Filters are sets that can be intersected and united in the usual way. Owing to the filter axiom (F3), the intersection of two filters can be expressed as the union of their elements. The union of two filters, however, is not the intersection of their elements—it is in general, not even a filter. Yet, the intersection of the elements of two filters does form a filter which we denote by $\mathcal{F} \vee \mathcal{G}$ ($\neq \mathcal{F} \cup \mathcal{G}$). The following should help in making this clear.

Let $\mathcal{F}$ and $\mathcal{G}$ be filters.

Lemma A.1. $\mathcal{F} \cup \mathcal{G} = \{F \cup G \mid F \in \mathcal{F}, G \in \mathcal{G}\}$

Proof. (i) $F \cup G \in \mathcal{F} \cap \mathcal{G}$, because $F \in \mathcal{F}$ implies $F \cup G \in \mathcal{F}$ (since $F \cup G$ is a superset of F) and similarly, $G \in \mathcal{G}$ implies $G \cup F \in \mathcal{G}$.

TABLE A1
Alternatives to neighborhood

		Neighborhood		Closure		Interior		Boundary
$\mathcal{N}(x)$	=			$\{A \mid x \notin \overline{X \backslash A}\}$	=	$\{A \mid x \in A^\circ\}$	=	$\{A \mid x \in A \backslash \partial A\}$
$\bar{A}$	=	$\{x \mid \forall N \in \mathcal{N}(x): A \cap N \neq \emptyset$	=			$X \backslash (X \backslash A)^\circ$	=	$A \cup \partial A$
A°	=	$\{x \mid \exists N \in \mathcal{N}(x): N \subseteq A\}$	=	$X \backslash \overline{X \backslash A}$	=			$A \backslash \partial A$
∂A	=	$\{x \mid \forall N \in \mathcal{N}(x):$	=	$\bar{A} \cap \overline{X \backslash A}$	=	$X \backslash (A^\circ \cup (X \backslash A)^\circ)$		
		$A \cap N \neq \emptyset, (X \backslash A) \cap N \neq \emptyset\}$						

TABLE A2
Axiom systems for neighborhood alternatives

	Neighborhoods	Closure	Interior	Boundary
1	$\forall N \in \mathcal{N}(x): x \in N$	$\bar{\emptyset} = \emptyset$	$\emptyset^\circ = \emptyset$	$\partial \emptyset = \emptyset$
2	$N \in \mathcal{N}(x), N \subset N' \Rightarrow N' \in \mathcal{N}(x)$	$A \subseteq \bar{A}$	$A^\circ \subset A$	$\partial A = \partial (X \backslash A)$
3	$N, N' \in \mathcal{N}(x) \Rightarrow N \cap N' \in \mathcal{N}(x)$	$\overline{A \cap B} = \bar{A} \cap \bar{B}$	$(A \cap B)^\circ = A^\circ \cap B^\circ$	$A \cap B \cap \partial (A \cap B) = A \cap B \cap (\partial A \cup \partial B)$
T	$N \in \mathcal{N}(x) \Rightarrow \exists N' \in \mathcal{N}(x):$ $\forall y \in N': \exists N_y \in \mathcal{N}(y): N_y \subseteq N$	$\bar{\bar{A}} = \bar{A}$	$(A^\circ)^\circ = A^\circ$	$\partial \partial A \subseteq \partial A$

(ii) If $H \in \mathscr{F} \cap \mathscr{G}$, then $H \in \mathscr{F}$ and $H \in \mathscr{G}$. Thus, H can trivially be written in the form $H = F \cup G$ with $F = H \in \mathscr{F}$ and $G = H \in \mathscr{G}$. □

Let us write $\mathscr{F} \vee \mathscr{G}$ for the object obtained from intersecting the elements of $\mathscr{F}$ and $\mathscr{G}$,

$$\mathscr{F} \vee \mathscr{G} := \{F \cap G \mid F \in \mathscr{F}, G \in \mathscr{G}\},$$

provided the intersections $F \cap G$ are non-empty for all $F \in \mathscr{F}$ and $G \in \mathscr{G}$. Otherwise, $\mathscr{F} \vee \mathscr{G} = \emptyset$ and we call $\mathscr{F}$ and $\mathscr{G}$ disjoint.

Observe that $F \cap G \subseteq F, G$ implies $F, G \in \mathscr{F} \vee \mathscr{G}$ for all $F \in \mathscr{F}$ and $G \in \mathscr{G}$, provided $\mathscr{F}$ and $\mathscr{G}$ are not disjoint. Thus, $\mathscr{F} \cup \mathscr{G} \subseteq \mathscr{F} \vee \mathscr{G}$.

However, since nothing ensures that the intersection $F \cap G$ of some $F \in \mathscr{F}$ and $G \in \mathscr{G}$ is an element of $\mathscr{F}$ or $\mathscr{G}$, $\mathscr{F} \cup \mathscr{G}$ is typically not a filter. $\mathscr{F} \vee \mathscr{G}$ is the coarsest filter that is finer than both $\mathscr{F}$ and $\mathscr{G}$, because if there was an $F \in \mathscr{F}$ and a $G \in \mathscr{G}$ such that $F \cap G \notin \mathscr{F} \vee \mathscr{G}$, the intersection axiom (F2) for filters would be violated.