Entropy statistics as a framework to analyse technological evolution

Eindhoven Centre for Innovation Studies

Entropy Statistics as a Framework to Analyse Technological Evolution

K. Frenken & A. Nuvolari

Eindhoven Centre for Innovation Studies, The Netherlands

Working Paper 02.15

Department of Technology Management

Technische Universiteit Eindhoven, The Netherlands

September 2002

Entropy Statistics as a Framework to Analyse Technological Evolution*

Koen Frenken

Department Economic Geography and Urban Research centre Utrecht (URU)
Faculty of Geographical Sciences
Utrecht University
P.O. Box 80115, 3508 TC, Utrecht, The Netherlands
e-mail: k.frenken@geog.uu.nl, www: http://econ.geog.uu.nl/department/home.html

Alessandro Nuvolari

Eindhoven Centre for Innovation Studies (ECIS)
Faculty of Technology Management
Eindhoven University of Technology
P.O. Box 513, 5600 MB, Eindhoven, The Netherlands
e-mail: a.nuvolari@tm.tue.nl , www: http://www.tm.tue.nl/ecis/

20 September 2002

Manuscript, third version

(Second version 30 January 2002, First version presented at 2nd EMAEE September 2001)

* We are grateful to John Kanefsky for generously providing us with the updated version of his dataset on British steam engines, and to Paolo Saviotti for the use of his data on aircraft and helicopters previously collected for a project funded by the ESCR. We thank Nick von Tunzelmann for helpful discussions. An earlier draft of this chapter was presented at the Second European Meeting on Applied Evolutionary Economics in Vienna, September 2001. We thank the participants to the workshop and, especially, Arnulf Grübler for their comments.

1. INTRODUCTION

Many scholars have suggested that important similarities exist between technological development and biological evolution and that, for this reason, evolutionary models can provide us with fairly adequate representations of technical change (Nelson and Winter, 1982; Basalla, 1988; Mokyr, 1990). However, as it has been repeatedly pointed out by those who endorse the adoption of an evolutionary approach, there are also substantive differences between biological evolution and technological evolution (Freeman, 1991; Nelson, 1995). Therefore, evolutionary models should always be employed with caution, taking into account the specificities of the processes of mutation and selection under study.

The issue we are considering here concerns evolutionary processes of a special kind, namely the way complex entities evolve through processes of mutation and selection. Recent evolutionary theorising in biology and artificial intelligence has stressed that complex entities evolve in ways that are different from non-complex ones in important respects. This claim has also important implications for models of technological evolution, as a technological artefact is a complex evolving entity *par excellence* (Rosenberg, 1976).

Following Simon's (1969 [1996]) work on the design of artificial systems, we describe a technological artefact as a man-made system constituted by interconnected components that are intended to collectively perform a number of functions. The complexity of an artefact is due to the interdependencies between components, which causes only some combinations of elements to work well together, in the sense that these combinations are capable of achieving satisfactory levels of performance. In Simon's view, a good deal of what we call innovative activities, consists in trying to improve the general performance of the artefact, by finding out progressively better configurations of its constituting elements.[1]

Until recently, however, formal treatments of system interdependencies for the understanding of technological innovation have been scarce. This has changed with the introduction of 'complexity' models from natural sciences in the realm of (evolutionary) economics. In this respect, Kauffman's (1993) NK-model of evolutionary biology has proven extremely promising and has already been adopted in a large number of contributions in the innovation and organization literature.[2] The NK-model represents the design process of a complex technological artefact as a trial-and-error process that is bound to end up in a local optimum. Though the NK-model has received considerable attention, considerably less effort has so far been put into empirical applications.[3] In this chapter, we set out a framework based on entropy statistics, which allows a relatively straightforward application of the NK model to empirical studies of technological change.

We consider the examples of the early development of the steam engine (1760-1800), the development of the aircraft (1913-1984), and the development of the helicopter (1940-1983) to illustrate the way in which the NK-model can be employed in empirical studies of technical change by means of entropy statistics. As we will see, the interpretative accounts that we were able to produce using the NK-model in combination with the entropy methodology, emendate the received histories of the technologies we are examining in this chapter in important respects. This suggests that other historical studies of technology could indeed highly benefit from the adoption of the type of approach we propose in this study.

[1] Bradshaw (1992) provides an interesting account of Wrights' development of early aircraft technology using the concepts of Simon.

[2] See, Kauffman and Macready (1995), Levinthal (1997), Frenken *et al.* (1999a), Auerswald *et al.* (2000), Gavetti and Levinthal (2000), Kauffman *et al.* (2000), Marengo *et al.* (2000), Rivkin (2000), Valente (2000), Fleming (2001), Fleming and Sorenson (2001), Frenken (2001). See also the discussion by Simon (2002) on the relationship between Kauffman's NK-model and Simon's (1969 [1996]) early work.

[3] Notable expectations are Fleming (2001) and Fleming and Sorenson (2001).

The remainder of the chapter is organized as follows. Section 2 contains an exposition of Kauffman's (1993) NK-model and a number of generalisations developed hereafter. Section 3 presents in detail our entropy methodology. Section 4 applies the entropy framework to data on steam engines, aircrafts, and helicopters and discusses the results in the light of received histories of these three technologies. Section 5 draws conclusions.

2. TECHNOLOGICAL DEVELOPMENT AS A SEARCH PROCESS ON RUGGED LANDSCAPES

Many scholars have recognised that interdependencies between components in technological artefacts are the prime source of design complexity (Simon, 1969 [1996]; Rosenberg, 1976; Sahal, 1985; Vincenti, 1990; Ziman, 2000). The existence of interdependencies between components implies that the functioning of a system cannot be fully understood from the functioning of its individual components. Depending on the precise combination of the components that make up a system, a component will function in a different way. And, each time one achieves to improve the functioning of one component, new problems can arise in other components requiring to be accordingly re-designed. In this context, Rosenberg (1976) introduced the concept of "technical imbalances" between components that trigger sequences of problems and solutions over time.[4]

The existence of system interdependencies is what we understand to be the nature of complexity in the development of new technological designs. In this perspective, the design task consists essentially in the combinatorial problem of assembling the right set of components together in a functioning system. The space of *all* the possible combinations between *all* the possible configurations of *all* the components of a system is called the "design space" of a technology (Bradshaw, 1992). Assume that a technology can be described by N components, or more generally, dimensions ($i=1,\ldots,N$). Along each dimension i, there exist A_i possible states or configurations, called "alleles", which can be coded as "0", "1", *et cetera*. Each possible design can then be written as a string of alleles $s_1s_2 \ldots s_N$ and is part of N-dimensional design space S, for which holds:[5]

$$s \in S\,;\;\; s = s_1 s_2 \ldots s_N\;;\;\; s_i \in \{\,0\,,1\,,\ldots,A_i - 1\,\} \qquad (2.1)^6$$

The combinatorial nature of a design space implies that the size of a design space increases exponentially for linear increases in N. The size of the design space S is given by the product of the number of alleles along each dimension:

$$S = \prod_{i=1}^{N} A_i \qquad (2.2)$$

[4] See also David (1975) in his discussion on localised technological change.

[5] The combinatorial nature of the design space of a system requires that dimensions are orthogonal to one another. Therefore, one dimension of a system cannot correspond with an allele of another dimension in the same system. For example, the description of alleles of the engine dimension as gasoline ("0"), electric ("1"), and steam ("2") implies that the type of battery used in electric engines cannot count as another dimension in the description of the vehicle as a system. The choice for a type of battery only constitutes a dimension for electric vehicles, and not for vehicle technologies in general.

[6] Note that, since the first allele is labelled "0", the description of alleles of an element ranges from 0 to A_i-1, while the number of alleles ranges from 1 to A_i .

In the case of binary strings (*i.e.* when all dimensions contain only two alleles "0" and "1"), the size of design space equals $S=2^N$ meaning that the number of possible designs doubles for each dimension added. As technological artefacts are typically made up by many dimensions and many alleles per dimension, they have enormous design spaces. Exploring the whole design space would be obviously very expensive. Instead designers will usually apply search rules that allows them to economise by examining only subsets of the design space. Thus only a small part of the design space will be actually searched, and an even smaller part of the design space will be actually commercialised on product markets.

2.1 THE NK-MODEL

Kauffman and Levin (1987) and Kauffman (1993) developed the NK-model to examine the properties of evolving complex systems with varying degrees of complexity. Complexity stems from interdependencies between the constituting dimensions of a system, such as genes in biological organisms and components in technological artefacts. The interdependencies between dimensions in a complex system are called "epistatic relations". An epistatic relation between components implies that when a component mutates, the mutation affects not only the functioning of the component itself but also the functioning of all the components that are "epistatically related" to it. The ensemble of epistatic relations in a technological system is called a technology's architecture (Henderson and Clark, 1990).[7]

The NK-model is restricted to particular types of system architectures that can be expressed by single parameter *K*, which stands for the number of other components that affect the functioning of each component. For example, the class of systems for which holds $K=1$ refers to systems with an architecture in which the functionality of each component depends on the choice of allele of the component itself and on the choice of the allele of one other component. The K-parameter can be considered an indicator of a system's complexity with $K=0$ being the least complex and $K=N-1$ the most complex architecture. When $K=0$ each technical dimension is independent from any other dimensions. Optimisation can then proceed by optimising each individual dimension separately, which will lead automatically to the global optimum. For increasing values of *K*, it will become increasingly harder to globally optimise the system design as interdependencies exist between dimensions. The number of local optima in which one can end up, increases with the value of *K*.[8]

Consider as an explanatory example, a system for which holds $N=3$ and $K=1$ with an architecture as specified in figure 1. In the figure mutations in the components in the columns affect the functioning of the component in the row, they are indicated with "x". Vice versa, the symbol "-" denotes that there is no epistatic relation between the component in the row and the one in the column. The architecture in figure 1 specifies the following epistatic relations between the three components in the system. In the figure when mutations in the components in the columns affect the functioning of the component in the row, they are indicated with "x". Vice versa, the symbol "-" denotes that there is no epistatic relation between the component in the row and the one in the column. In our case, the functioning of the first component w_1 changes only when the component itself or the second component is mutated. The functioning of the second component w_2 changes only when the component itself or the first component is

[7] A system's architecture has also been termed the system's internal structure (Simon 1969 [1996]; Saviotti 1996).

[8] The K-value is an indicator of the complexity of a system's architecture and does not exactly coincide with the system's computational complexity, which can be expressed in the computational time that is required to optimise a complex system. On this, see Frenken *et al.* (1999a).

mutated. And, the functioning of the third component w_3 changes only the component itself or the first component is mutated.

FIGURE 1 ABOUT HERE

Following Kauffman (1993), we construct a fitness landscape by drawing randomly the value of the fitness w_i of component i from the uniform distribution between 0 and 1. A random value is drawn for w_i each time component i itself is mutated and each time another component that epistatically affects component i, is mutated. System fitness W is derived as the mean value of the fitness values of all components:

$$W(s) = \frac{1}{N} \cdot \sum_{i=1}^{N} w_i \ (s_i) \tag{2.3}$$

A simulation of a fitness landscape is given in figure 2. The circled strings are local optima or "peaks" on a "rugged fitness landscape". For these local optima it holds that all neighbouring strings, i.e. the strings that can be reached by a mutation in one component, have a lower fitness W. In the simulation in figure 2, this property holds for strings 011 and 101 as their system fitness values W(011) and W(101) exceed the values of their neighbouring strings. Local optima reflect complementary alleles as the collective fitness exceeds the value of neighbouring strings.

FIGURE 2 ABOUT HERE

Using the concepts of design space and fitness landscape, the design process can be modelled as a local search process based on trial-and-error. Local search proceeds by means of a mutation in one, randomly chosen, dimension (a *trial*). A mutation means that a designer moves to a neighbouring string in the design space. The newly found string is accepted when system fitness W increases, while it is rejected when system fitness decreases (*error*). Acceptance of a mutation implies that search continues from the newly found string, and rejection implies that search continues from the previous string. In this way, a designer can search for improvements in an incremental way until a local optimum is found that can no longer be improved by means of a mutation in one dimension.[9] Trial-and-error search can thus be considered as an "adaptive walk" over a fitness landscape towards a local optimum, and search will only halt when a local optimum is reached. Following the metaphor of the fitness landscape, this type of search in complex technological systems can be considered a process of "hill-climbing".

It should be stressed that we used the relative simple case of $N=3$ in the example above for explanatory purposes. In real-world Research and Development activities, the number of design dimensions N is generally much larger. Consequently, local search takes place in much larger design spaces containing many more local optima for the same value of K. The probability to end up in a local optimum is correspondingly much higher.

An important property of the NK-model holds that the number of local optima in a fitness landscape is a function of the complexity K of a system architecture. When complexity is absent ($K=0$), the fitness values of each dimension are not affected by mutations in other

[9] Allowing for mutation in several dimensions at the same time would allow a designer to escape local optima. However, the more dimensions that are allowed to be mutated at the same time, the higher the search costs involved as the number of possible moves increases exponentially with the number of dimensions that is allowed to be mutated at the same time. One can thus argue that designers are expected to search in only few dimensions at the same time. On this issue, see Frenken *et al.* (1999a) and Kauffman *et al.* (2000).

dimensions. Therefore, the global optimum of a system of $K=0$ can always be found by local search through trial-and-error as described above. Put another way, fitness landscapes of $K=0$-systems always contain only one optimum (which is by definition the global optimum). For systems with a positive K-value, the fitness values of dimensions are affected by mutations in dimensions that are epistatically related. As a result, the fitness landscape will generally contain multiple local optima. Kauffman (1993) has shown that the expected number of local optima increases for increases in K. This means that it becomes increasingly harder to find the global optimum for systems with higher complexity.

A second property of the NK-model holds that the fitness of local optima decreases for increases in K. One can understand this outcome as reflecting the detrimental effects of a higher number of conflicting constraints between components. The higher K, the more difficult it becomes to improve the fitness of one component without lowering fitness of other components. Consequently, the system fitness of local optima is generally quite low. Kauffman (1993) showed that the value of local optima decreases for increases in system complexity K. Furthermore, the variance of fitness value of local optima also decreases for increases in K, which means that the differences in fitness of local optima becomes smaller for systems with higher complexity. In the context of competing technological designs, this result suggests that the higher a technology's complexity, the smaller the performance differences between locally optimal designs, the more persistent design variety will be.[10]

2.2 GENERALISATIONS OF THE NK-MODEL

The NK-model can be generalised in a number of respects to represent a wider range of phenomena. The first generalisation concerns the representation of the relation between a system's "genotype" (the set of design dimensions) in relation to the "phenotype" (the set of functions a system performs). The NK-model is based on the idea that each component of the system performs an "own" sub-function within the system with regard to the attainment of one overall function on which external selection operates (Kauffman, 1993, p. 37). Each component i is conceived to have a particular fitness value w_i that reflects its functional contribution to the system as a whole. The fitness of the system as a whole is derived as the average of the fitness of individual components.

Altenberg (1995, 1997) describes a generalised (biological) model of complex systems that contains N dimensions ($i=1,...,N$) and F functions ($f=1,...,F$) and for which holds that the N does not necessarily equal F. In biological systems, for which the original and generalised NK-model were both initially conceived, an organism's N genes are the system's components and an organism's F traits are the system's functions in which natural selection operates. The string of genes constitutes an organism's genotype and the set of traits constitutes an organism's phenotype. The genotype of an organism is the level at which mutations take place, which are transmitted in its offspring. The phenotype is the level at which natural selection operates in terms of its relative fitness.

Analogously, a technological artefact can be described in terms of its N components and the F functions it performs. The string of alleles describes the "genotype" of a technological system, and the list of functions describes the "phenotype" of this system. Typical functions of technological artefacts include cost-related criteria (fuel-efficiency, maintenance cost, etc.) and performance related criteria (power, speed, weight, safety, etc.).[11]

[10] For more properties of the NK-model, see Kauffman (1993), Altenberg (1997), and Frenken *et al.* (1999a).

[11] This perspective on fitness differs from the NK-model applied to process technology where fitness is expressed only by a single cost criterion (Auerswald *et al.*, 2000; Kauffman *et al.*, 2000).

In Altenberg's generalised NK-model, the architecture of a complex system is represented by a "genotype-phenotype matrix" of size FxN with:

$$M = [m_{fi}],\ f = 1,...,F,\ i = 1,...,N \quad (2.4)$$

As in the NK-model, an epistatic relation is represented by "x" when function *f* is affected by component *n* and by "–" when function *f* is not affected by the component *n*. An example of a matrix for *N=3* and *F=2* is given figure 3. Note that *K=2* in this example as the fitness of all functions is affected by the choice of allele in two dimensions.

FIGURE 3 ABOUT HERE

The way in which fitness landscape are constructed for generalised genotype-phenotype matrices follows the same logic as the original NK-model discussed in the previous section. For each component that is mutated, all functions that are affected by this component are assigned a new randomly drawn fitness value w_f from the uniform distribution between 0.0 and 1.0. Total fitness *W* is again derived as the mean of the fitness values of all functions:

$$W(s) = \frac{1}{\mathrm{F}} \cdot \sum_{f=1}^{\mathrm{F}} \mathrm{w}_{\mathrm{f}}(s) \quad (2.5)$$

A simulation of the fitness landscape example of the genotype-phenotype matrix of figure 3 is given in figure 4 for all possible combinations between two alleles of three components. Local optima are again circled reflecting the combinations in which component technologies are complementary.

FIGURE 4 ABOUT HERE

The meaning of the concepts of fitness landscape and local optima remains entirely the same in Altenberg's (1997) generalised NK-model. Moreover, the properties of the NK-model discussed above, that relate the number of local optima, the fitness of local optima, and the variance in fitness of local optima to complexity *K*, remain intact. The main difference compared to the original NK-model is that in the generalised model the number of dimensions *N* is not necessarily equal to the number of functions *F*. Altenberg's (1997) model can therefore be considered as an important generalisation of the original NK-model of complex systems by Kauffman (1993).[12]

A second generalisation of the NK-model can be introduced by specifying a more general fitness function that translates the fitness levels of individual functions w_f into one overall assessment value *W*. The fitness function in equation (2.5) specified that each function is weighted equally. As an empirical specification of fitness (performance) of a technology, this equation obviously does not account for the general case in which users may apply different weights to the various functions of the artefact. Generally, users do not assign the same weight to each function but rank some functions higher than other ones. Allowing for different values of weights for each function, we get:

[12] Altenberg's generalised NK-model also allows one to model search by adding new components to a system that would increases *N* while keeping the number of selection criteria *F* constant. On this, see Altenberg (1995) and Frenken (2001, chapter 4).

$$W(s) = \sum_{f=1}^{F} \beta_{\mathrm{f}} \cdot \mathrm{w}_{\mathrm{f}}(s) \qquad (2.6)$$

$$\sum_{f=1}^{F} \beta_{\mathrm{f}} = 1 \;,\; \beta_{\mathrm{f}} > 0 \qquad (2.7)$$ [13]

A selection environment can then be defined by the set of weights $\{\beta_1, \beta_2, \ldots, \beta_F\}$ that is applied by users of the technology. The concept of a fitness landscape does not change when total fitness is computed as a weighted sum instead of as the average of the fitness values of functions. However, the values of total fitness of each design *W(s)* will be different depending on the values of the weights that are applied.

A final generalisation can be introduced by allowing for heterogeneity among users. So far, we implicitly assumed that the fitness of each design takes on only one value for all users. Put another way, we assumed that each user of a particular design applies the same set weights and thus assigns the same fitness value *W(s)* to a design. However, depending on the specific use of the design , different users may well apply different weights, and thus assign different fitness values to one and the same design (Lancaster, 1966, 1979; Saviotti and Metcalfe, 1984; Saviotti, 1996).[14] In this case of heterogeneous demand, different users have different valuations of the same technological design, as they weight the levels of functions differently. As Lancaster (1979, p. 17) expressed it:

> "Differences in individual reactions to the same good are seen as expressing different preferences with respect to the collection of characteristics possessed by that good and not different perceptions as to the properties of the good."

The weights assigned to functions as specified above $\{\beta_1, \beta_2, \ldots, \beta_F\}$ reflect one homogeneous user group. When there exist more than one user group, we can characterise each different user group by a different set of weights. For a *G* number of user groups *g* (*g=1,…,G*), we have *G* sets of weights. For each user group, the fitness W_g of a design is given by:

$$W_g(s) = \sum_{f=1}^{F} \beta_{\mathrm{fg}} \cdot \mathrm{w}_{\mathrm{f}}(s) \qquad (2.8)$$

$$\sum_{f=1}^{F} \beta_{\mathrm{fg}} = 1 \;,\; \beta_{\mathrm{fg}} \geq 0 \qquad (2.9)$$ [15]

This specification of the selection environment includes the specification given above for a homogeneous selection environment as the special case in which *G=1*.

---

[13] This is a relatively simple function sometimes applied in multi-criteria analysis of project selection (Nijkamp *et al.*, 1990). This function implies that a loss in fitness of one function can infinitely be substituted by an increase in other functions. Various alternative functions exist to derive a fitness value or "utility" from a collection of characteristics (Lancaster, 1966, 1979).

[14] Compare the SCOT approach of Pinch and Bijker (1984), who stress the interpretative flexibility of the meaning and use of artefacts. Here, the sociology of technology meets evolutionary economics.

[15] Note that in the case of heterogeneous user groups, some weights can equal zero, while in the case of a homogeneous user group in formula 2.7, all weights are by definition positive. A zero weight in a homogeneous user population would imply that the feature does not count as a function for anyone.

Different user groups or market segments are thus characterised by the different ranking of functions. When heterogeneity in preferences is more dispersed, it is less likely that one design is optimal for all user groups. In that case, product differentiation is expected to occur. In the extreme case, given a sufficiently large design space, a different design may be found for each different user group. When a user group exists for which no design is yet optimised, and this is known to designers, this in itself can spur the search for innovations in particular components in order to find a new design capable of fitting their particular demand ("induced innovation").

2.3 IMPLICATIONS

Regarding the patterns of technological evolution one can expect to emerge, a number of implications follow from the previous discussion of the NK-model and its generalisations. To summarise the implications, one can distinguish between technologies that are subject to homogenous demand and technologies that are subject to heterogeneous demand:

- When demand is homogeneous ($G=1$) and complexity is absent ($K=0$) there exists only one local optimum, which can be found by local trial-and-error. When demand is homogeneous ($G=1$) and system complexity is present ($K>0$), the expected number of local optima becomes a function of the complexity parameter K. Thus, even when demand is homogeneous, technological variety is expected to emerge as different designers will come up with different locally optimal, solutions. What is more, the variety in technological designs can be quite persistent for highly complex architectures as the variance in the fitness values of local optima has been shown to decrease as a function of K.

- When demand is heterogeneous ($G>1$) and complexity is absent ($K=0$), there is only one global optimum, which is the same for each user group, since each function can be optimised independently from other functions. The existence of heterogeneous demand is not a sufficient condition for design variety to emerge. When demand is heterogeneous ($G>1$) and complexity is present ($K>0$), design variety is expected to emerge for two reasons. First, as in the case of homogeneous demand, trial-and-error may lead different designers to come up with different local optima. Second, the heterogeneity in preferences may render different designs to be globally optimal for different user groups. Note that in the case of heterogeneity in preferences, the design variety is expected to be even more persistent than in the case of homogeneity in preferences where variety may slowly disappear as sub-optimal design loose ground due to small differences in fitness.

Note that design variety that is expected to emerge following the generalised NK-model, is always limited by the extent to which scale economies and network externalities are realised in the production and use of a single design (Arthur, 1989). When one design s is produced and used in much higher numbers than alternative designs, lower price and higher willingness-to-pay may attract users that previously preferred an alternative design. However, a greater degree of heterogeneity of preferences as expressed by the different weights users assign to the various functions, will in turn render it less likely that one design will attract all users. Moreover, radical innovation may at all times lead to the introduction of complete new designs attracting new users or users that previously adopted another design.

Patterns of technological evolution thus depend crucially on the complexity of a technology's architecture, the number of functions that can be distinguished, and the degree of heterogeneity of demand. In the three cases we will discuss below (steam engines, aircrafts, and helicopters), the complexity and number of selection criteria is generally estimated to be quite high (Rosenberg, 1982). Moreover, all three technologies have been used in a wide range of user contexts. In our view, one can expect an empirical analysis to show technological variety to emerge in the course of their evolution.

3. ENTROPY STATISTICS AS A FRAMEWORK TO ANALYSE TECHNOLOGICAL EVOLUTION

In the previous section we proposed a formalisation of artefact complexity and discussed its implications for the patterns of technological evolution that are expected to emerge. A straightforward way to analyse empirical data on artefact designs in terms of the ruggedness of fitness landscapes is to apply entropy statistics. Entropy statistics can be computed using frequency distributions of technological designs coded in the N-dimensional design space and they allow one to map both the degree of technological variety (by means of entropy indices), and the nature of technological variety (by means of mutual information indices). In this way, evolutionary trends in the development of a technology can be consistently outlined.

The entropy index refers to the degree of randomness in the choice of technological designs as reflected by the skewness of a distribution. A skewed distribution reflects a situation in which designers hardly differ in their choice of design, while a flat distribution reflects a situation in which designers have come up very many different designs. As such, entropy can be used as an indicator of technological standardisation and to what extent a dominant design can be said to have emerged (Frenken *et al.*, 1999b). The more skewed a distribution, the lower the entropy (randomness) of a distribution.

To understand to what extent the variety indicated by entropy can indeed be said to reflect local optima on a rugged fitness landscape, a second indicator called mutual information is introduced (Frenken, 2000, 2001). Mutual information indicates the extent in which particular alleles along different dimensions co-occur in the technological designs offered on the market. Statistically, mutual information thus indicates the degree of dependence between different design dimensions. The existence of local optima would imply that particular alleles along one dimension typically co-occur with particular alleles along other dimensions, which would result in a high value of mutual information (dependence). When alleles along different dimensions are more or less randomly combined, mutual information is low (independence).

3.1 ENTROPY

The entropy concept was developed in late nineteenth century thermodynamics to describe randomly moving particles (Prigogine and Stengers, 1984). When many particles are randomly moving through a state space, like particles of a gas in a box, the resulting distribution of all particles is completely flat. The flat distribution follows from the fact that at all times each particle has an equal probability to be present in any area in the box. The flat distribution is characterised by maximum entropy (randomness). When particles behave in a non-random way, some areas in the box will be filled with more particles than other areas, and the resulting distribution is skewed. In that case, the entropy of the distribution is lower compared to the case in which all particles move randomly. In the extreme case when all particles cluster in one area of the box, entropy is lowest.

Entropy is thus a macroscopic measure at the level of a distribution that indicates the degree of randomness in the micro-dynamics underlying a frequency distribution. As such, entropy can also be used as a variety measure of frequency distributions of technological designs. Following Saviotti (1996), we refer to a distribution of technological designs as the "product population". Maximum entropy corresponds to the case in which all designs occur at the same frequency. Such a complete flat distribution would occur when designers would randomly move around in state space, which has been called here the "design space". In that case, designers pick randomly the various alleles of each component. In this hypothetical case, any product design has an equal probability to occur, and the product population would be characterised by even frequencies of all designs. This hypothetical situation refers to a situation in which designers do not "learn" about the functional properties of different designs, and simply choose the alleles configuration at random (analogous to the randomly moving particles in a box explained above). A skewed distribution occurs when some designs are dominating the product population. In that case, the frequency of some designs is high, while the frequency of most designs is low or zero. Such a distribution has a low entropy value indicating that design variety is low. In this case, designers have not chosen a design at random, but have somehow learned which designs are most demanded, for example, by applying a local search strategy of hill-climbing. In the extreme case in which all designers choose to offer one and the same design on the market, entropy will be minimum.

The entropy measure thus indicates the degree of design variety in a product population. To describe a product population as a frequency distribution of designs, let each design be coded again as a string of N alleles ($i=1,\dots,N$). Each of the N dimensions is labelled here as X_i with each dimension containing A_i alleles again coded as "0","1", etc. The relative frequency of design s in the product population is denoted as p_s. The entropy value of an N-dimensional distribution is then given by (Theil, 1967; 1972; Langton, 1990):

$$H\,(X_1,\dots,X_N) = -\sum_{s_1=0}^{A_1-1}\dots\sum_{s_N=0}^{A_N-1} p_s \cdot \ln p_s \qquad (3.1)^{16\ 17}$$

Entropy is zero when all products present in the population are designed according to one and the same design. This design would have a frequency of one in the product population, which implies that the entropy of the product population equals:

$$H_{\min} = -1 \cdot \ln(1) = 0$$

Entropy is positive otherwise. The larger the entropy value, the larger the design variety in the product population. The maximum entropy is limited by the size of design space S. When all S possible combinations of alleles have an equal frequency, we obtain a uniform distribution in which each design has frequency $p_s = 1/S$. The entropy of this distribution equals:

$$H_{\max} = -S \cdot \left(\left(\frac{1}{S}\right) \cdot \ln\left(\frac{1}{S}\right) \right) = -\ln\left(\frac{1}{S}\right) = \ln(S)$$

[16] In information theory, entropy is computed using the logarithm of two instead of the natural logarithm taken here (Theil, 1967, 1972; Frenken *et al.*, 1999b).
[17] $0 \cdot \ln(0) \equiv 0$.

This value is the maximum possible entropy value for a distribution of product designs with a design space of *S* possible designs.

Similarly, the design variety along one dimension *i* can be computed. The one-dimensional or marginal entropy indicates the variety in a product population with respect to one design dimension only, and is given, for each dimension, by:

$$H(X_i) = -\sum_{s_i=0}^{A_i-1} p_{s_i} \cdot \ln p_{s_i} \tag{3.3}$$

As we will see, the one-dimensional entropy formula can be used to compute the mutual information index, which is equal to the difference between the sum of one-dimensional entropy values and the N-dimensional entropy value.

3.2 MUTUAL INFORMATION

In information theory, the measure that indicates the degree of dependence (co-occurrence of alleles) in a frequency distribution is the measure of mutual information *T*. Mutual information is given by (Theil, 1967, 1972; Langton, 1990):

$$T(X_1, \ldots, X_N) = \sum_{s_1=0}^{A_1-1} \cdots \sum_{s_N=0}^{A_N-1} p_s \cdot \ln \frac{p_s}{\prod_{i=1}^{N} p_{s_i}} \tag{3.4}$$

The mutual information value *T* indicates the extent in which alleles along different dimensions are co-occurring in the distribution of designs. The mutual information value equals zero when there is exist no dependence between any of the dimensions. In that case, the joint frequency of alleles of components p_s corresponds exactly to the frequency that could be expected from the product of the marginal frequencies $\prod_{i=1}^{N} p_{s_i}$. When the product of marginal frequencies does not correspond to the joint frequency, there is dependence between dimensions. Mutual information is thus derived by the weighted sum of dependence values for each design. It can be proven that the weighted sum of dependence values is non-negative for any frequency distribution, i.e. $T \geq 0$ (Theil, 1972). The greater the difference between the joint frequency and the product of marginal frequencies, the higher the value of the mutual information, the more alleles along particular dimensions co-occur in "design families".

The mutual information measure is directly related to the concept of entropy as mutual information can be derived from the multi-dimensional and marginal entropy values. In the general case of an N-dimensional distribution (*N>1*) the mutual information equals the sum of marginal entropy values minus the N-dimensional entropy value (Theil and Fiebig, 1994: 12):

$$T\ (X_1,,...,X_N) = \left(\sum_{i=1}^{\mathrm{N}} H(X_i)\right) - \ H\ (X_1,...,X_N) \tag{3.5}$$

From this equation, it is can derived that the mutual information equals zero if entropy equals zero, and that mutual information equals zero if entropy is maximum (see Appendix).

Similarly, one can compute the mutual information between each pair of dimensions to indicate dependence between two dimensions

$$T\ (X_i, X_{i+1}) = H(X_i) + H(X_{i+1}) - H\ (X_i, X_{i+1})$$

The two-dimensional mutual information values indicate the dependence between a pair of dimensions and are thus informative with regard to the importance of epistatic relations among the pair of dimension in question. A high mutual information between two dimensions suggests that an important epistatic relation exists between the two dimensions, since designers offer predominantly alleles in particular opposite combinations (*e.g.*, either combination 00 or combination 11). These combinations reflect complementarities between alleles in the two dimensions as particular alleles are often co-occurring with particular alleles along the other dimension.

3.3 ENTROPY AND MUTUAL INFORMATION AS INDICATORS OF EVOLUTION

To explain the connection between entropy and mutual information indicators and the exploration of rugged fitness landscapes, one should keep in mind the relationship between entropy en mutual information. Recall formula (3.5), which expresses mutual information as the sum of marginal entropy values minus multi-dimensional entropy, which can be rewritten as:

$$\left(\sum_{i=1}^{\mathrm{N}} H(X_i)\right) = \ T\ (X_1,...,X_N) + \ H\ (X_1,...,X_N)$$

From this formula, it can readily be seen that, *given a value for the sum of marginal entropy (ΣH_i)*, mutual information can increase only at the expense of (total) entropy, and *vice versa*. This relationship is illustrated in Table 1 in which three different frequency distributions of designs are listed (for *N=3*). In all three cases, the sum of marginal entropy values is the same ($\Sigma H_i = 3 \cdot ln(2) = ln(8)$), because in all three cases the two alleles along each dimension occur at the same frequencies. However, the three-dimensional entropy and mutual information values differ for each distribution. Case 1 corresponds to a uniform distribution with maximum entropy and zero mutual information. Case 2 shows a multi-modal distribution with four designs have positive and equal frequencies. Three-dimensional entropy equals ln(4) while three-dimensional mutual information is only ln(2). Finally, case 3 shows a bi-modal fifty-fifty distribution in which two opposite designs are present in the product population (000 and 111). In this case, three-dimensional entropy equals only ln(2) while three-dimensional mutual information adds up to ln(4). The latter case is characterised by such high mutual information because knowledge of one allele along one dimension of a design would allow one to perfectly predict the alleles along the two other dimensions.

TABLE 1 AROUND HERE

When entropy and mutual information are applied to frequency distributions of consecutive years of technological evolution, a very different picture may emerge. In that case, the value of ΣH_i in a particular year will differ from the value of ΣH_i in other years. Over time, the value of ΣH_i may increase or decrease, or show no trends. An increasing trend would indicate a growing variety in alleles used along each design dimension. Following the formula such an increase in the value ΣH_i implies that entropy and mutual information can increase *both* at the same time. In that case, we have a pattern of increasing design variety as indicated by the rise in $H(X_1 \ldots X_N)$ *and* of increasing differentiation of designs in families as indicated by the rise in $T(X_1 \ldots X_N)$. Such a process indicates the progressive development of a growing number of design families akin to "speciation" in biology (Saviotti, 1996; Levinthal, 1998). The reverse pattern can also take place. When ΣH_n is falling, entropy and mutual information may decrease at the same time (for example when a product family totally disappears).

The evolutionary development of a complex technology, following the generalised NK-model as discussed earlier, is expected to be characterised by both increasing degree of variety (entropy) and an increasing degree of differentiation (mutual information). Such a development process can be understood from the multi-dimensional and complex nature of technological artefacts on the one hand, and from the existence of heterogeneous demand at the other hand.

4. APPLICATIONS

We will confront our thesis of growing design variety and differentiation into design families using data on early steam engines (1760-1800), aircraft (1913-1984) and helicopters (1940-1983). For each technology we will first provide a short summary of the "standard" historical account of its development, then present the data and results, and finally discuss what new insights can be derived from our analysis.

4.1 Steam engines

Early steam engine history[18]

Historians of technology have described the development of steam power technology as a "linear" succession of technological breakthroughs. The main contours of what might be called the "traditional" account[19] of early steam engine development concern the design sequence of Savery-Newcomen-Watt-Trevithick that has taken place during the eighteenth century.

In the late seventeenth century mining activities begun to be severely hampered by flooding problems. Following the scientific investigations of Torricelli and Pascal, there were several attempts of using atmospheric pressure to lift water out of mines. The Savery engine can be considered as the first successful effort in this direction. The engine was developed in the period 1695-1702. In the Savery engine, steam was first admitted and then condensed inside a "receiving" vessel by pouring cold water over its outside. Following steam condensation, atmospheric pressure drove water up into the vessel. The engine suffered of two major shortcomings, which limited its practical utilization: restricted height of operation and high fuel consumption due to the need of recreating steam inside the vessel at each stroke.

[18] A more elaborated account can be found in Frenken and Nuvolari (2002).
[19] In this respect, Dickinson (1938) can be considered as an exemplar reference.

The Newcomen engine, developed in 1712, resolved the problem of limited height of operation. The Newcomen engine was constituted by a piston-cylinder arrangement connected with one hand of a working beam. Steam was admitted from the boiler into the cylinder by means of a valve. Then a cold of jet of water was sprayed into the cylinder condensing the steam. At this point, because of the creation of a partial vacuum, atmospheric pressure pushed the piston down, lifting the pump rod at the other end of the beam. The use of the cylinder-piston arrangement together with the beam made it possible the use of the engine for an effective mine drainage. Furthermore, the Newcomen engine was robust, highly reliable and based on a fairly simple working principle. The Newcomen engine, however, did not solve the problem of high fuel consumption. Neither did the engine design deliver smooth motion, preventing the use of this kind of engine in applications in which a smooth rotary motion was needed.[20]

James Watt in the 1770s and in the 1780s tackled successfully these two problems. In his engine condensation was carried out in a separate vessel and not in the cylinder. This design implied that there was no longer the need of re-heating the cylinder at each stroke, which greatly contributed to fuel-efficiency. After the invention of the separate condenser, Watt conceived a number of modifications to his engine in order to allow the effective transformation of reciprocating into rotary motion. Among the designs that have been developed for rotary motion is the double-acting Watt engine, in which steam is admitted into the cylinder on both sides of the piston in an alternating manner. This resulted in a more powerful action, but also in a much more regular movement of the piston.

Finally, in the second half of the 1790s, Richard Trevithick developed the first high-pressure engine (Watt engines used steam at a little more than atmospheric pressure). This type of engines did not use the separate condenser, but discharged exhaust steam directly into the atmosphere. For this reason, they were called "puffers". The main advantage of this type of engines was their compactness and their cheaper cost of installation due to elimination of the condenser, the air pump and the beam.[21]

As it is apparent from this narrative, such a historical depiction is akin to chronicling a sort of "glorious march of invention", where most of the emphasis is put on the creative contributions of a succession of individual inventors (the line Savery-Newcomen-Watt-Trevithick). Each inventor tackled the shortcomings of the technological "state of the art" devising improvements that made previous engine designs obsolete through a process of technological substitution.

Early steam engine data

The data we use are taken from an up-to-date version of the database collected by John Kanefsky.[22] The database contains a list of all steam engines (more precisely, those for which some historical evidence has been found) erected in Great Britain over the period 1700-1800. We have limited ourselves to the period 1760-1800 as the period before 1760 was entirely dominated by the Newcomen design and thus was characterised by absence of variety and differentiation.[23]

[20] A number of Newcomen engines were successfully used to raise water over a water wheel which, in turn, delivered rotary motion for factory machinery. These types of engines were usually called returning engines.

[21] Von Tunzelmann (1978, p. 263).

[22] For more details on the original data see Kanefsky (1979). For a more accessible reference, see Kanefsky and Robey (1980).

[23] To be more precise, apart from the Newcomen design a second engine design was available before 1760. This design is the Savery engine, which we have excluded altogether from the analysis as it did not meet the classification of our design space. We consider the Savery engine to be a steam pump rather than a steam engine as it lacks the characteristic piston-cylinder arrangement characteristic of all the other steam engines. The

The database contains 1370 engines for the period 1760-1800. Each of these engines is coded as a string of seven alleles that describes the engine design as a point in a seven-dimensional design space. Dimensions and alleles are given in Table 2. The design dimensions have been constructed in such a way that each design could be coded as a unique string, thus covering the most relevant dimensions of early steam engine technology. After having coded each engine in the database as a design string according to the classification of the design space in Table 2, we constructed yearly frequency distributions and computed the entropy and mutual information values.

Note that we have consider three-years moving averages of the yearly entropy and mutual information values in order to smooth short term fluctuations and obtain a "neater" pattern. The results in the figures below are shown per year, where each year stands for the in between year of a three-year period. The transformation of yearly values into three-year moving averages does not affect in any way our conclusions.

TABLE 2 AROUND HERE

Results on steam engines

From the results, it is immediately clear that variety (entropy) and differentiation (mutual information) have both increased very rapidly from 1774 onwards when the Watt engine became a popular design next to the older Newcomen design. The rise in variety and differentiation levelled off around ten years later (more or less from 1785). What is also clear is that, as both entropy and mutual information have been rising, the sum of marginal entropy values must have risen too following formula (3.5). This shows that the technological evolution of the steam engine has been characterised by the introduction of new alleles in several dimensions accounting for the rise in the sum of marginal entropy values. The introduction of new alleles has been such that both the variety in designs and the degree of differentiation in design families have risen. Put another way, the design variety has been made possible by the development of new alleles that are combined in a highly non-random ways.

FIGURE 5 AROUND HERE

Closer inspection of the graph also shows that during the 1770s and early 1780s the rise of entropy precedes increases in mutual information. We understand this as being probably due to the fact that, first, new combinations of alleles were tried, leading to an increase of variety. However, some of these new combinations did not reach adequate levels of fitness, and so we see that, with a delay, mutual information "catches up" with the entropy, which means that the product population is clustering around some specific points of the landscape. In other words, we have first a phase of exploration and discovery of new areas of the landscape, followed by concentration in some points that are likely to be local optima. The "levelling off phase" seems to suggest that from the late 1780s a stable pattern of differentiation finally emerged.

FIGURE 6 AROUND HERE

Results on two-dimensional mutual information values are depicted in Figure 6. The figure shows along which couples of dimensions differentiation has been most pronounced. Hence, these results are also informative on the nature of the technological interdependencies

---

exclusion of Savery engine should not affect our results since only 33 Savery engines are present in the original data. More details on the Savery engine can be found in Frenken and Nuvolari (2002).

(epistatic relations) existing among the constituting elements of our design space. The highest mutual information values are reached by the pair $T(X_2,X_6)$, which reflect to interdependency between condensation and the closed top cylinder. Separate condensation and the closed top cylinder are the two salient features distinguishing Watt type of engines (0100010) from the Newcomen atmospheric engine without condensation and open top (0000000). Importantly, the high values of $T(X_2, X_6)$ are not temporary but continue during the whole period considered. These results thus confirm the thesis of an emergence of a pattern of differentiation.

The other couples of dimensions with high mutual information values are $T(X_2, X_3)$, $T(X_2, X_5)$, $T(X_3, X_5)$, $T(X_3, X_6)$, and $T(X_5, X_6)$. What becomes clear from these results is that the high values are limited to four dimensions: X_2, X_3, X_5, and X_6 (respectively, with/without condenser, single/double action, reciprocating/rotary/water returning, and open/closed top). As explained above, dimensions X_2 and X_6 differentiate Newcomen and Watt engines. Dimensions X_3 and X_5 concern different type of solutions to deliver particular types of motion. Double action was a typical feature of Watt rotary engines (0110110), while Newcomen engines delivering rotary motion made use of either returning a stream over a waterwheel (0000200) or directly by using alternatively two cylinders (0000101).

Discussion

The pattern of growing variety and differentiation of early steam engine technology suggests that newly developed designs did not simply substitute older designs, but enlarged total design variety. Our analysis shows that the "linear" view of the early history of the steam engine is essentially untenable. Instead, technological evolution in this period it is better characterised by progressive differentiation into distinct design families.[24] In terms of the NK model, the clustering of the product population around some specific designs can be seen as a reflection of the various local optima in the fitness landscape.

The process of differentiation proceeded along four specific technical dimensions (with/without condenser, single/double action, reciprocating/rotary/water returning, and open/closed top). These dimensions may well be related to different user contexts, in particular to the cheapness of coal and to the desired properties of the rotary motion. The pattern of specialisation we find contradicts received histories of early steam engine evolution that point to a process of substitution between Watt engines and Newcomens. Though Watt's inventions are considered to have solved the main shortcomings of the Newcomen engine, it is misleading to assume that their led to the substitution of Newcomen engines.

Regarding the superiority of Watt's fuel efficiency, one can understand the limited substitution of Newcomen engines by Watt engines, taking into account the higher costs of erection and maintenance of the Watt engine. In this respect, von Tunzelmann (1978) has argued that in areas where coal was cheap enough, the Newcomen engine detain an important advantage due to its lower costs of installation and maintenance. Besides, whereas the Newcomen engine was well within the engineering capabilities of the time, the Watt engine imposed very compelling requirements on the degree of accuracy of the various components of the engine. This points to the existence of a fundamental trade-off concerning fuel-efficiency versus simplicity of construction and maintenance.[25]

[24] A similar conclusion based on historical grounds, stressing the role of variety, has been reached by Von Tunzelmann (1978, p. 24): "It is misleading to see the pattern of progress [in steam engine technology] as linear and inevitable: in explaining the direction and the chronology of 'technical progress' in the economist's sense, it is vital to keep this diversity in mind."

[25] Joseph Bramah stated that the Newcomen engine detained over Watt "an infinite superiority in terms of simplicity and expense". John Smeaton, one of the leading engineers of the time, considered that the Watt engine demanded too higher standards for construction and maintenance. See Harvey and Down Rose (1980, pp. 22-23).

Regarding the type of motion that Watt engine were capable of delivering, the significance of Watt's design modifications also requires further nuance. Although Watt's inventions for supplying rotary motion were highly celebrated (Dickinson and Jenkins, 1927), they should not by any means be considered as definitive, especially, given the accuracy of workmanship of the time. We are aware of many cases of unsatisfactory performance of Watt rotary engines in textile mills.[26] This explains why Watt engines only partially substituted alternative designs that delivered rotary motion.

FIGURE 6 AROUND HERE

Interestingly enough, there was an attempt of developing an "hybrid" engine combining the simplicity of Newcomen with the fuel-efficiency of Watt. This was the "improved atmospheric engine" patented by Symington in 1787 (0100000). Unfortunately, we have scant information on this engine (especially on its actual fuel efficiency compared to Watt). We know that about twenty of these type of engines were erected mainly in Scotland and that they generally proved rather successfully.[27] Some historians of technology (Dickinson and Jenkins, 1927) have dismissed Symington simply as "schemer" who tried to circumvent Watt's patent.[28] Our results, instead, suggest that his attempt of merging the two separate design trajectories of the Newcomen and Watt designs was genuinely aimed at solving a teething trade-off.

Summarizing, the existence of various user contexts implied that engine designs differentiated in order to provide adequate responses to the specific demands of the various users'sectors. In our case, this determined a divergence of design trajectories and the emergence of what we might call technological niches. So what we see is indeed a process akin to speciation in biology. In a companion paper (Frenken and Nuvolari, 2002), we study in greater detail the pattern of specialisation of different type of steam engines in the various users contexts.

4.2 Aircraft

Aircraft history

Both historians and economists have analysed the development of aircraft technology in considerable detail (e.g., Miller and Sawers, 1968; Constant, 1980; Bilstein, 1996). Although these studies differ in their perspectives and methodologies, there is a general consensus on the main stages of aircraft development, which can be divided in four periods.

The early history of aircraft from the turn of the century to roughly 1930 is characterised by a large variety of designs and limited demand. A large number of new, small firms experimented with various designs and materials. This period is commonly considered an explorative stage in the industry characterised by a great deal of trial-and-error. During this period, series production remained limited causing production costs and prices to be too high for mass consumption.

The second stage covering the thirties and early forties has been marked as the period of technological convergence towards what has been termed a 'dominant design' (Abernathy and

[26] See Hills (1970, pp. 179-186). Many contemporary engineers believed that the rotary drive produced by a water returning engine was much more regular and, in the end, "better" than the one obtained from rotary Watt engine. See Von Tunzelmann (1978, pp. 142-143).

[27] On the Symington engine see Harvey and Down Rose (1980, chapter 3).

[28] Dickinson and Jenkins (1927, p. 318). See also Farey (1827, p. 656).

Utterback, 1978). The Douglas DC3 developed in the mid-thirties is generally considered the exemplar of this dominant design. The DC3 is an all-metal, monocoque, piston propeller monoplane with two-engines placed under the wings. Production costs of this design rapidly fell due to its commercial success both in military and civil aviation. In the early forties, total production of the DC3 reached 10,000 models (Jane's 1978). The DC3 design also provided the basis of the development of a whole product family developed throughout the forties and the fifties including the DC4, DC5, DC6, and DC7. The main difference in the later models concerns the use of four engines to provide more engine power. At the time, many firms including Boeing copied the DC3 designs in their piston propeller product lines for passenger aircraft and bombers.

The third stage covering the period of forties and fifties is characterised by the introduction of jet engines. The first experiments with jet engines go back to W.W. II, but its successful application in both military and civil aircraft took place only in the fifties. The transition from piston propeller to jet engines has been widely recognised as a technological revolution, which has established a shift in the prevailing 'technological paradigm' (Constant, 1980; Dosi, 1982). The introduction of jet engines did not simply replace piston propeller engines in existing designs, but also led to the development of new technologies in other parts of the aircraft, notably the introduction of swept wings that were better fitted to cope with the increased engine power of jet engines. The revolutionary nature of jet engine technology can be further supported by the fact that the Douglas, as the most successful company in piston propeller aircraft, lost its leading position to Boeing that dominates the aircraft industry up till the present day.

The fourth stage of aircraft development has been characterised by the further diffusion of jet engine aircraft designs into the realm of short-distance aircraft and business aircraft. In the period after the fifties, no major change in aircraft design has taken place as innovative activities increasingly shifted from aircraft design to avionics.

Aircraft data

The data on aircraft designs concern the alleles of six design dimensions and covers the period 1913-1984. As aircraft development only took off in the early 1900s, the data can be considered to cover the larger part of the aircraft history. The choice of the six dimensions and its alleles is based on the limitation posed by the data source, which concerns photographs of aircraft designs. Admittedly, other dimensions that are known to have played an important role including the types of landing gear and the type of materials used, could not be coded due to the limitation of the source materials. The photographs were drawn from Jane's (1978, 1989) encyclopaedia on aviation, which is known to be among the most comprehensive encyclopaedia of aviation and aircraft designs from all countries. The data of the six dimensions have been compiled for a sample of 731 aircraft models, corresponding to a sample covering other aircraft variables previously assembled by Paolo Saviotti.[29]

TABLE 3 AROUND HERE

The frequency distributions of designs that are used to measure entropy and mutual information at particular moments in time, are not the yearly distributions of product designs. In this case, a year is a too short time-span as aircraft designs are typically products that remain on offer for many years after their introduction. We used ten-year distributions, but

[29] See, Saviotti and Bowman (1984) and Saviotti (1996).

calculations for five-year and fifteen year distributions yielded the same trends as discussed below.

The results in the figures below are shown using still a year basis, where each year covers a ten-year period. Thus, the distribution of designs associated with a specific year corresponds to a time period of ten years beginning in that year. In other words, the year 1913 stands for the distribution of designs introduced between 1913 and 1922, the year 1914 stands for the distribution of product designs introduced between 1914 and 1923, *et cetera*.

Results on aircraft

The results on entropy and mutual information for aircraft are given in Figure 7. Entropy increased in the early decades and decreased only slightly in the thirties. In the forties and early fifties entropy increased rapidly again to level off in the late-fifties. Mutual information shows less fluctuations, with the general trend being upwards. Notably, mutual information has risen substantially during the period of the forties and fifties and levelled off hereafter. The results suggests that the long term evolution of aircraft is characterised by both growing variety and growing differentiation into different design families.

FIGURE 7 AROUND HERE

The results for the pair-wise mutual information in figure 8 prove informative regarding the dimensions along which the differentiation process has taken place. It is clear that the rise in mutual information in the post-war period is primarily related to rising mutual information between the engine type and the wing type $T(X_1,X_4)$, between the engine type and the number of engines $T(X_1,X_2)$, and between the number of engines and the wing type $T(X_2,X_4)$. The values for these three pairs of design dimensions have increased very rapidly. The emergence of design families can thus be related to the interdependencies between these design dimensions. The functionality of a particular engine type depends heavily on the complementarities with the number of engines used $T(X_1,X_2)$ and with the type of wings $T(X_1,X_4)$. And, the functionality of the number of engine used depends heavily on the wing type of the aircraft $T(X_2,X_4)$. The local optima in fitness landscapes are thus primarily characterised by different alleles engine type, wing type, and the number of engines. Counting the various designs in the final period after 1960 lead us to distinguish between four design families (Frenken, 2001): one- and two-engine piston propeller aircraft with straight wings, two-engine turboprop monoplanes with straight wings, one- and two-engine jet aircraft with delta wings and two-, three and four-engine turbofan aircraft with swept wings.

FIGURE 8 AROUND HERE

Epistatic relations among other pairs of dimensions do not show high dependence suggesting that dimensions X_3, X_5 and X_6 have not been constitutive for the emergence of design families. All two-dimensional mutual information values including X_3, X_5 or X_6 remain low throughout the period with the exception of the value $T(X_2,X_5)$, This value shows some increase in the twenties and early thirties and indicates the common use of an uneven number of engines in two-tail aircraft design, with one engine place between the tails. After the thirties, however, two-tail aircraft designs were hardly being introduced, which suggests that this trajectory has proven a dead-end.

Discussion

From our results we conclude that the history of aircraft technology is characterised by a progressive development of designs into four distinct families: one- and two-engine piston propeller aircraft with straight wings, two-engine turboprop monoplanes with straight wings, one- and two-engine jet aircraft with delta wings and two-, three and four-engine turbofan aircraft with swept wings. Though not entirely disagreeing with the histories of the aircraft industry as sketched before, these results offer a number of new insights into its evolutionary dynamics.

First, the emergence of a dominant design in the thirties commonly associated with the Douglas DC3 had only a limited effect on the total design variety in the industry. The results on aircraft entropy show that the increase in variety was indeed halted during the thirties, but not decrease substantially. Second, the advent of jet engine aircraft in the forties and fifties contributed, as expected, to design variety with entropy values rapidly rising during this period. However, after the fifties entropy remained at a high level suggesting that jet engines design did not fully substituted propeller designs. Instead, a pattern of differentiation occurred as indicated by the rising values of mutual information, with piston propeller and turbo-propeller engine design co-existing alongside jet and turbofan engine designs.

We understand this stable pattern of differentiation as reflecting the different uses of different aircraft designs found in an earlier study that related engine types to market applications (Frenken, 2000). Piston propeller engine design has become dominant in low-cost, small-distance operations including trainer aircraft, business aircraft and agricultural aircraft. Turbo-propeller engine aircraft are used for small distance passenger aircraft and military transport, while turbofan engine aircraft are used for medium and long distance passenger aircraft. Finally, jet engines are predominantly used in high-speed fighter aircraft.

Note that the history of aircraft technology shows some interesting parallels with early steam engine technology in that both technologies have witnessed the introduction of a revolutionary design (the jet engine and Watt's engine, respectively). Yet, in both industries the introduction of the revolutionary design has not so much led to a substitution process as well as to a process of progressive differentiation into different design families.

4.3 Helicopters

Helicopter history

Though the concept of helicopters has a long history that goes back to China in about 400 B.C., the first successful helicopter dates from 1939 with the development of the VS-300 by Sikorsky (Taylor, 1995). The advent of helicopter technology quickly received interest from armies and navies, because of its capacity to evacuate people from areas that were not accessible by airplanes. The military demand for helicopter induced a great deal of explorative activity in the forties and fifties including variations in the type of engine, the number of rotors, and the number of blades. At the time, commercial expectations were high as evidenced by popular magazines predicting that American households would soon have a family helicopters in the garage.

In the late fifties, the explorative stage of technological development largely came to an end as design convergence took place with the apparent superior engine performance of turbines to piston engines. According to Bilstein (1996: 91), the single-rotor twin-turboshaft Kaman-model introduced in 1954 can be considered the "pioneering" design with hindsight. Hereafter, the twin-engine turboshaft design with one rotor became the 'dominant design'.

Commercially, however, helicopter technology never took off as a mass produced product. Compared to aircraft, the costs and limited range of helicopters impede its wider diffusion in segments currently dominated by aircraft (Taylor, 1995). Instead, most helicopters are used for transporting people in areas not accessible by aircraft (like military troops or off shore oil platform personnel), while niche applications exist for a variety of uses including ambulance operations and fighter operations.

Helicopter data

The data on helicopter concern the alleles of five design dimensions and covers the period 1940-1983. The data of the five dimensions have been complied for a sample of 144 helicopter models. As for the data on aircraft, the helicopter data have been compiled on the basis of observable characteristics on photographs and correspond to the sample previously compiled by Paolo Saviotti[30] from Jane's (1978, 1989) encyclopaedia on aviation.

TABLE 4 AROUND HERE

As for aircraft, the frequency distributions of designs that are used to measure entropy and mutual information at particular moments in time, are not the yearly distributions of designs. We used again ten-year distributions, but the calculations for five-year and fifteen year distributions yielded the same trends as in the results based on ten-year distributions discussed below. The results in the figures below are shown per year, where each year stands for the first year of a ten-year period.

Results on helicopters

The results on entropy and mutual information are given in figure 9. Interestingly, the results on helicopter variety and differentiation show patterns that are altogether different from the results on early steam engines and aircraft. After a short period of rising values, entropy has fallen from the 1955 onwards showing that product variety in product designs has fallen. The value for mutual information peaked earlier in 1949 and also shows a declining trend hereafter. Note that the decline in mutual information has been relatively greater than the decline in entropy values (mutual information halved during the period 1950-1980).

FIGURE 9 AROUND HERE

The two-dimensional mutual information values for helicopters also show decreasing trends. Only one pair of dimensions $T(X_2,X_3)$ shows the highest values over the whole period reflecting complementarities between the number of engines and the number of blades. This relationship points to the common use of more blades when more engines are incorporated in a helicopter design to carry the higher weight.

FIGURE 10 AROUND HERE

Discussion

The fall in mutual information accompanied by a fall in entropy suggest that after a brief period of differentiation, we have a prolonged phase in which design variety is decreasing. The results correspond to Bilstein's (1996, p. 91) historical account that identified the single-

[30] See, Saviotti and Trickett (1992) and Saviotti (1996).

rotor twin-turboshaft design as the dominant design emerging in the fifties. Our analysis is also in line with findings by Saviotti and Trickett (1992, p. 116) who found that the single-rotor turboshaft helicopters increased their share in the population from around 30 percent in the late fifties to around 80 percent in the early eighties. A small second family of design is constituted by two-rotor two-engine helicopters that are typically used for heavy transport operations.

In the case of helicopter technology, de-differentiation can not be attributed to absence of heterogeneity in demand. In fact, Saviotti and Trickett (1992) distinguish between up to 22 different uses of helicopters ranging from fighter operations to troops transport to ambulance to business transport. User heterogeneity may well be at least as high as in aircraft industry, even though sales of the helicopter industry is only a fraction of sales in the aircraft industry. Given the heterogeneity of helicopter demand and the process of de-differentiation of helicopter supply, Saviotti and Trickett conclude that heterogeneity in demand is met by modular designs capable of being used in a variety of contexts. In this context, one must think of helicopters in which the interior is easily adapted without changing the helicopter design itself.

The results still leaves open the question why heterogeneity in user contexts in early steam engine and aircraft have triggered differentiation to take place, while heterogeneity in helicopters users has not led to a sustained pattern of differentiation. Following Frenken *et al.* (1999b), one may attribute this evolutionary pattern to the existence of inter-technological competition between helicopter technology and aircraft technology. Within the market for air transport, helicopter technology itself operates within a relatively small niche, which is bounded by the presence if aircraft technology. Over the past few decades, single-rotor helicopter performance has been limited by a flight range of around 1000 km, a speed of 300 km/h and payload of around 10000 kilograms. The halt in improvements does not reflect technical difficulties, but competition with aircraft: further improvements in speed, range or payload are technically perfectly realizable, but would lead helicopters to compete with small aircraft covering the market segments of longer distances, higher speeds and higher payload. Put another way, the range of performance levels that helicopters are technically capable of applying is not fully explored due to the presence of cheaper aircraft technology.

5. Conclusion

We started our study by introducing the NK-model as a formal model of complex evolving systems that are characterised by interdependencies among their constituting components. We proposed a number of generalisations to the original NK-model to account for the specificities of technological evolution. By examining the properties of this generalized NK model , we concluded that technological development in complex technologies is likely to lead to a process of differentiation of designs into distinct families. This view contradicts models of technological substitution that depict competition among designs as uni-dimensional (cost-based) process that leave room for only one surviving technology.

To analyse the evolutionary pattern of technological development in terms of changes in variety and differentiation, we proposed the methodology of entropy statistics. Entropy provides us with a comprehensive measure of design variety, while mutual information indicates to what extent this variety is non-random, i.e. clustered in specific areas of the design space. The existence of multiple clusters indicates the presence of local optima in the technology's fitness landscape.

We applied the entropy statistics to data on design dimensions of three technologies. The results confirmed our hypothesis of increasing variety through differentiation for aircraft and steam engines, while the de-differentiation process of helicopter technology could be attributed to the presence of a competing aircraft models. Furthermore, the empirical results offered us insights on the (quantitative) evolution that differ from received histories of steam engines and aircraft. We found that the evolution of these two technologies is better described as a evolutionary process of differentiation rather than a linear substitution process. Obviously, a next step is to apply the same technology to other technologies as well. The proposed methodology can be applied to any technology given that sufficient empirical data are available on the relevant design dimensions of the technology in question.

REFERENCES

Abernathy, W.J., Utterback, J. (1978) 'Patterns of industrial innovation', *Technology Review* 50, pp. 41-47.

Altenberg, L. (1995) 'Genome growth and the evolution of the genotype-phenotype map', pp. 205-259, in: Banzhaf, W., Eckman, F.H. (eds.) *Evolution and Biocomputation* (Berlin & Heidelberg: Springer-Verlag).

Altenberg, L. (1997) 'NK fitness landscapes', in: Back, T., Fogel, D., Michalewicz, Z. (eds.) *The Handbook of Evolutionary Computation* (Oxford etc.: Oxford University Press).

Arthur, W.B. (1989) 'Competing technologies, increasing returns, and lock-in by historical events', *The Economic Journal* 99, pp. 116-131.

Auerswald P., Kauffman S., Lobo J., Shell K. (2000) 'The production recipes approach to modeling technological innovation: an application to learning-by-doing', *Journal of Economic Dynamics and Control* 24, pp. 389-450.

Basalla, G. (1988) *The Evolution of Technology* (Cambridge: Cambridge University Press).

Bilstein, R.E. (1996) *The American Aerospace Industry* (New York: Twayne Publishers / Prentice-Hall).

Bradshaw, G. (1992) 'The airplane and the logic of invention', pp. 239-250 in: Giere, R. (ed.) *Cognitive Models of Science* (Minneapolis: University of Minnesota Press).

Constant II, E. W. (1980) *The Origins of the Turbojet Revolution* (Baltimore & London: John Hopkins University Press).

David, P.A. (1975) *Technical Choice, Innovation and Economic Growth. Essays on American and British Experience in the Nineteenth Century* (London: Cambridge University Press).

Dickinson, H.W. (1938) *A Short History of the Steam Engine* (Cambridge: Cambridge University Press).

Dickinson, H.W., Jenkins, R. (1927) *James Watt and the Steam Engine* (London: Encore).

Dosi, G. (1982) 'Technological paradigms and technological trajectories. A suggested interpretation of the determinants and directions of technical change', *Research Policy* 11, pp. 147-162.

Farey, J. (1827) *A Treatise on the Steam Engine* (reprint, Newton Abbot, David & Charles, 1971).

Fleming, L. (2001) 'Recombinant uncertainty in technological search', *Management Science* 47, pp. 117-132.

Fleming, L., Sorenson, O. (2001) 'Technology as a complex adaptive system: evidence from patent data', *Research Policy* 30, pp, 1019-1039.

Freeman, C. (1991) 'Innovation, changes of techno-economic paradigm and biological analogies in economics', *Revue Economique* 42, pp. 211-232.

Frenken, K. (2000) 'A complexity approach to innovation networks', *Research Policy* 29, pp. 257-272.

Frenken, K. (2001) *Understanding Product Innovation using Complex Systems Theory*, PhD Thesis, University of Amsterdam and University of Grenoble.

Frenken, K., Marengo, L., Valente, M. (1999a) 'Interdependencies, near-decomposability and adaptation', pp. 145-165 in: Brenner, T. (ed.) *Computational Techniques for Modelling Learning in Economics* (Boston etc.: Kluwer).

Frenken, K., Nuvolari, A. (2002) 'The early development of the steam engine: an evolutionary interpretation using complexity theory', *mimeo,* 29 August.

Frenken, K., Saviotti, P.P., Trommetter, M. (1999b) ‘Variety and niche creation in aircraft, helicopters, motorcycles and microcomputers’, *Research Policy* 28, pp. 469-488.

Gavetti, G., Levinthal, D. (2000) ‘Looking forward and looking backward: cognitive and experiential search’, *Administrative Science Quarterly* 45, pp. 113-137.

Harvey, W., Downs-Rose, G. (1980) *William Symington: Inventor and Engine Builder* (London: Northgate Publishing).

Henderson, R., Clark, K. (1990) ‘Architectural innovation’, *Administrative Science Quarterly* 35, pp. 9-30.

Hills, R. (1970) *Power in the Industrial Revolution* (Manchester: Manchester University Press).

Jane’s (1978) *Jane’s Encyclopedia of Aviation* (London: Jane’s Publishing Company Ltd).

Jane’s (1989) *Jane’s Encyclopedia of Aviation* (London: Studio Editions).

Kanefsky, J.W. (1979) *The Diffusion of Power Technology in British Industry 1760-1870*, PhD Thesis, University of Essex.

Kanefsky, J.W., Robey, J. (1980) ‘Steam engines in 18th-century Britain: a quantitative assessment’, *Technology and Culture* 21, pp. 161-186.

Kauffman, S.A. (1993) *The Origins of Order. Self-Organization and Selection in Evolution* (New York & Oxford, Oxford University Press).

Kauffman, S.A., Levin, S. (1987) ‘Towards a general theory of adaptive walks on rugged landscapes’, *Journal of Theoretical Biology* 128, pp. 11-45.

Kauffman, S.A., Lobo, J., Macready, W.G. (2000) ‘Optimal search on a technology landscape’, *Journal of Economic Behavior and Organization* 43, pp. 141-166.

Kauffman, S.A., Macready, W.G. (1995) ‘Technological evolution and adaptive organizations’, *Complexity* 1, pp. 26-43.

Lancaster, K.J. (1966) ‘A new approach to consumer theory’, *Journal of Political Economy* 14, pp. 133-156.

Lancaster, K.J. (1979) *Variety, Equity and Efficiency* (New York: Columbia University Press).

Langton, C.G. (1990) ‘Computation at the edge of chaos’, *Physica D* 42, pp. 12-37.

Levinthal, D.A. (1997) ‘Adaptation on rugged landscapes’, *Management Science* 43, pp. 934-950.

Levinthal, D.A. (1998) ‘The slow pace of rapid technological change: Gradualism and punctuation in technological change’, *Industrial and Corporate Change* 7, pp. 217-247.

Marengo, L., Dosi, G., Legrenzi, P., Pasquali, C. (2000) ‘The structure of problem-solving knowledge and the structure of organizations’, *Industrial and Corporate Change* 9, pp. 757-788.

Miller, R., Sawers, D. (1968) *The Technical Development of Modern Aviation* (London: Routledge & Kegan Paul).

Mokyr, J. (1990) *The Lever of Riches* (Cambridge: Cambridge University Press).

Nelson, R.R. (1995) ‘Recent evolutionary theorizing on economic change’, *Journal of Economic Literature* 33, pp. 48-90.

Nelson, R.R., Winter, S.G. (1982) *An Evolutionary Theory of Economic Change* (Cambridge MA & London: Belknap Press of Harvard University Press).

Nijkamp, P. Rietveld, P., Voogd, H. (1990) *Multicriteria Evaluation and Physical Planning* (Amsterdam etc.: North-Holland).

Pinch, T.J., Bijker, W.E. (1984) 'The social construction of facts and artifacts: or how the sociology of science and the sociology of technology might benefit each other', *Social Studies of Science* 14, pp. 399-442.

Prigogine, I., Stengers, I. (1984) *Order out of Chaos* (New York: Bantam).

Rivkin, J.W. (2000) 'Imitation of complex strategies', *Management Science* 46, pp. 824-844.

Rosenberg, N. (1976) *Perspectives on Technology* (Cambridge: Cambridge University Press).

Rosenberg, N. (1982) *Inside the Black Box* (Cambridge: Cambridge University Press).

Sahal, D. (1985) 'Technological guideposts and innovation avenues', *Research Policy* 14, pp. 61-82.

Saviotti, P.P. (1996) *Technological Evolution, Variety and the Economy* (Cheltenham & Brookfield: Edward Elgar).

Saviotti, P.P., Bowman, A. (1984) 'Indicators of output of technology', in: Gibbons, M., Gummett, P., Udgaonkar, B.M. (eds.) *Science and Technology Policy in the 1980s and Beyond* (London: Longman).

Saviotti, P.P., Metcalfe, J.S. (1984) 'A theoretical approach to the construction of technological output indicators', *Research Policy* 13, pp. 141-151.

Saviotti, P.P., Trickett, A. (1992) 'The evolution of helicopter technology, 1940-1986', *Economics of Innovation and New Technology* 2, pp. 111-130.

Simon, H.A. (1969) *The Sciences of the Artificial* (Cambridge MA. & London: MIT Press, third edition, 1996).

Simon, H.A. (2002) 'Near decomposability and the speed of evolution', *Industrial and Corporate Change* 11, pp. 597-599.

Taylor, J.W.R. (1995) 'Constraints impeding the commercial use of helicopters', pp. 155-167 in: Leary, W.M. (ed.) *From Airships to Airbus. The History of Civil and Commercial Aviation. Volume 1: Infrastructure and Environment* (Washington & London: Smithsonian Institution Press).

Theil, H. (1967) *Economics and Information Theory* (Amsterdam: North-Holland).

Theil, H. (1972) *Statistical Decomposition Analysis* (Amsterdam: North-Holland).

Theil, H., Fiebig, D.G. (1984) *Exploiting Continuity. Maximum Entropy Estimation of Continuous Distributions* (Cambridge MA: Ballinger).

Valente, M. (2000) *Evolutionary Economics and Computer Simulations. A model for the evolution of markets*. PhD thesis, Aalborg University.

Vincenti, W.G. (1990) *What Engineers Know and How They Know It. Analytical Studies from Aeronautical History* (Baltimore & London: John Hopkins University Press).

Von Tunzelmann, G.N. (1978) *Steam Power and British Industrialization to 1860* (Oxford: Clarendon Press).

Ziman, J. (2000) (ed.) *Technological Innovation as an Evolutionary Process* (Cambridge: Cambridge University Press).

APPENDIX. Derivation of mutual information for zero entropy and maximum entropy

Entropy is zero when one design occurs with frequency one implying that the alleles incorporated in this design also occur with frequency one. Therefore, the sum of marginal entropy values equals zero, implying that mutual information equals zero:

$$T\,(X_1,\dots,X_N) = \left(\sum_{i=1}^{\mathrm{N}} H(X_i)\right) - \; H\,(X_1,\dots,X_N)$$

$$T\,(X_1,\dots,X_N) = \left(\sum_{i=1}^{\mathrm{N}} -1\cdot\ln 1\right) - -1\cdot\ln 1 = 0 + 0 = 0$$

Entropy is maximum when all possible designs in design space have an equal frequency 1/S. In that case, the alleles along each dimension also have an equal frequency with marginal frequencies equalling $1/A_i$. Mutual information becomes:

$$T\,(X_1,\dots,X_N) = \left(\sum_{i=1}^{\mathrm{N}} H(X_i)\right) - \; H\,(X_1,\dots,X_N)$$

$$T\,(X_1,\dots,X_N) = \left(\sum_{i=1}^{\mathrm{N}} \ln A_i\right) - \ln S$$

$$T\,(X_1,\dots,X_N) = \ln\left(\prod_{i=1}^{N} A_i\right) - \ln S$$

$$T\,(X_1,\dots,X_N) = \ln S - \ln S = 0$$

Distribution	$\mathbf{p_{000}}$	$\mathbf{p_{001}}$	$\mathbf{p_{010}}$	$\mathbf{p_{011}}$	$\mathbf{p_{100}}$	$\mathbf{p_{101}}$	$\mathbf{p_{110}}$	$\mathbf{p_{111}}$
Case 1	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
Case 2	0.250	0.000	0.000	0.250	0.000	0.250	0.250	0.000
Case 3	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.500

Entropy	$\mathbf{H(X_1,X_2,X_3)}$	$\mathbf{H(X_1, X_2)}$	$\mathbf{H(X_1, X_3)}$	$\mathbf{H(X_2, X_3)}$	$\mathbf{H(X_1)}$	$\mathbf{H(X_2)}$	$\mathbf{H(X_3)}$
Case 1	ln 8	ln 4	ln 4	ln 4	ln 2	ln 2	ln 2
Case 2	ln 4	ln 4	ln 4	ln 4	ln 2	ln 2	ln 2
Case 3	ln 2	ln 2	ln 2	ln 2	ln 2	ln 2	ln 2

Mutual information	$\mathbf{T(X_1, X_2, X_3)}$	$\mathbf{T(X_1, X_2)}$	$\mathbf{T(X_1, X_3)}$	$\mathbf{T(X_2, X_3)}$
Case 1	0	0	0	0
Case 2	ln 2	0	0	0
Case 3	ln 4	ln 2	ln 2	ln 2

Table 1: Three examples of distribution for binary strings of $N=3$

STEAM ENGINE

Number of observations:	*1370*
Time span:	*1760-1800*
Area:	*Great Britain*

$\mathbf{X_1}$	**Pressure**
$A_1 = 2$	**0** low, **1** high
$\mathbf{X_2}$	**Condenser**
$A_2 = 2$	**0** yes, **1** no
$\mathbf{X_3}$	**Action**
$A_3 = 2$	**0** single acting, **1** double acting
$\mathbf{X_4}$	**Compounding**
$A_4 = 2$	**0** yes, **1** no
$\mathbf{X_5}$	**Motion**
$A_5 = 3$	**0** reciprocating, **1** rotary, **2** water returning
$\mathbf{X_6}$	**Top**
$A_6 = 2$	**0** open, **1** closed
$\mathbf{X_7}$	**Cylinder**
$A_7 = 2$	**0** single, **1** double

Table 2: design space of steam engine technology

AIRCRAFT

Number of observations: *731*
Time span: *1913-1984*
Area: *World*

X_1 **Engine type**
$A_1 = 5$ **0** piston-propeller, **1** turboprop, **2** jet, **3** turbofan, **4** rocket

X_2 **Number of engines**
$A_2 = 7$ **0** one, **1** two, **2** three, **3** four, **4** six, **5** eight, **6** twelve

X_3 **Number of wings**
$A_3 = 3$ **0** monoplane, **1** biplane, **2** triplane

X_4 **Wing type**
$A_4 = 4$ **0** straight, **1** delta, **2** swept, **3** variable swept

X_5 **Number of tails**
$A_5 = 2$ **0** one, **1** two

X_6 **Number of booms**
$A_6 = 3$ **0** one, **1** two, **2** three

Table 3: design space of aircraft technology

HELICOPTERS

Number of observations: *144*
Time span: *1940-1983*
Area: *World*

X_1 **Engine type**
$A_1 = 5$ **0** piston, **1** piston turbo, **2** ramjet, **3** gas generator, **4** turboshaft

X_2 **Number of engines**
$A_2 = 3$ **0** one, **1** two, **2** three

X_3 **Number of blades**
$A_3 = 7$ **0** two, **1** three, **2** four, **3** five, **4** six, **5** seven, **6** eight

X_4 **Number of shafts**
$A_4 = 2$ **0** one, **1** two

X_5 **Number of rotors per shaft**
$A_5 = 2$ **0** one, **1** two

Table 4: design space of helicopter technology

	i=1	i=2	i=3
w_1	**x**	x	-
w_2	x	**x**	-
w_3	x	-	**x**

FIGURE 1. Example of an architecture of N=3-system with K=1

	w_1	w_2	w_3	W
000:	0.7	0.2	0.3	0.40
001:	0.7	0.2	0.6	0.50
010:	0.9	0.3	0.3	0.50
011:	0.9	0.3	0.6	0.60
100:	0.8	0.7	0.1	0.53
101:	0.8	0.7	0.4	0.63
110:	0.3	0.5	0.1	0.30
111:	0.3	0.5	0.4	0.40

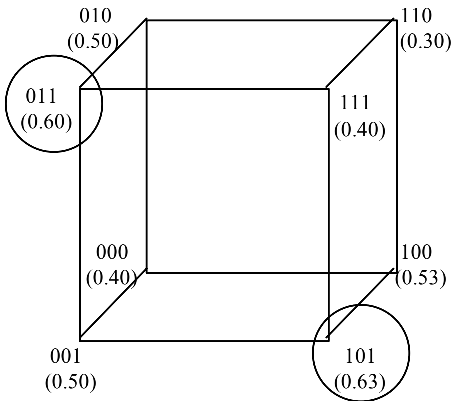

FIGURE 2. Simulation of a fitness landscape of a N=3-system with K=1

	i=1	i=2	i=3
f=1	x	x	-
f=2	-	x	x

FIGURE 3. Example of a generalised genotype-phenotype matrix

	w_1	w_2	W
000:	0.8	0.9	0.85
001:	0.8	0.6	0.70
010:	0.4	0.3	0.35
011:	0.4	0.2	0.30
100:	0.2	0.9	0.55
101:	0.2	0.6	0.40
110:	0.9	0.3	0.60
111:	0.9	0.2	0.55

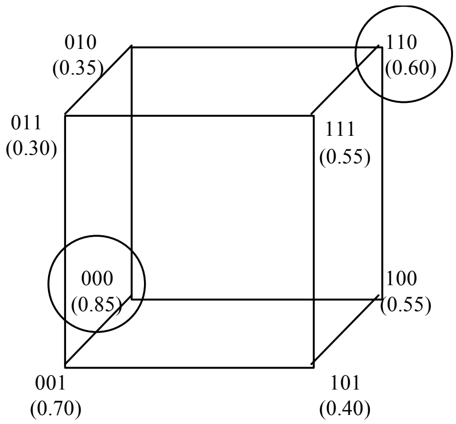

FIGURE 4. Simulation of fitness landscape of the matrix in figure 3

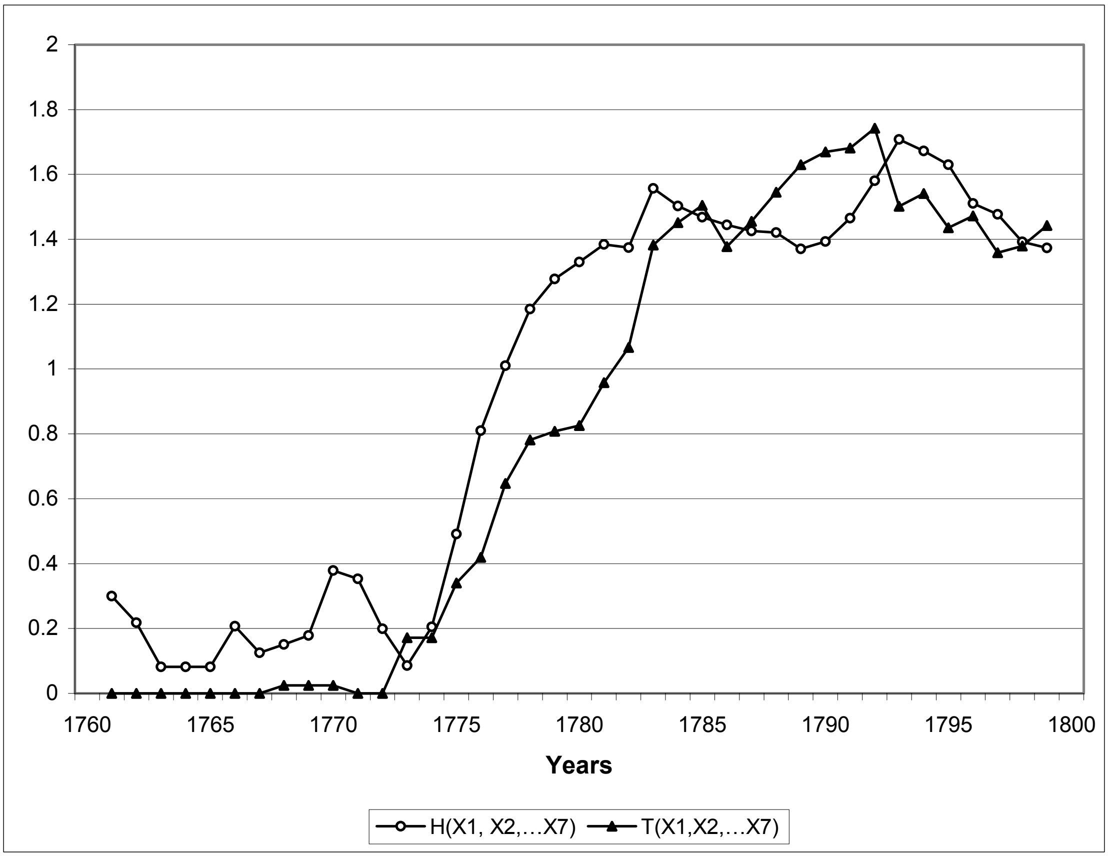

FIGURE 5. N-dimensional entropy (H) and mutual information (T) for steam engine designs

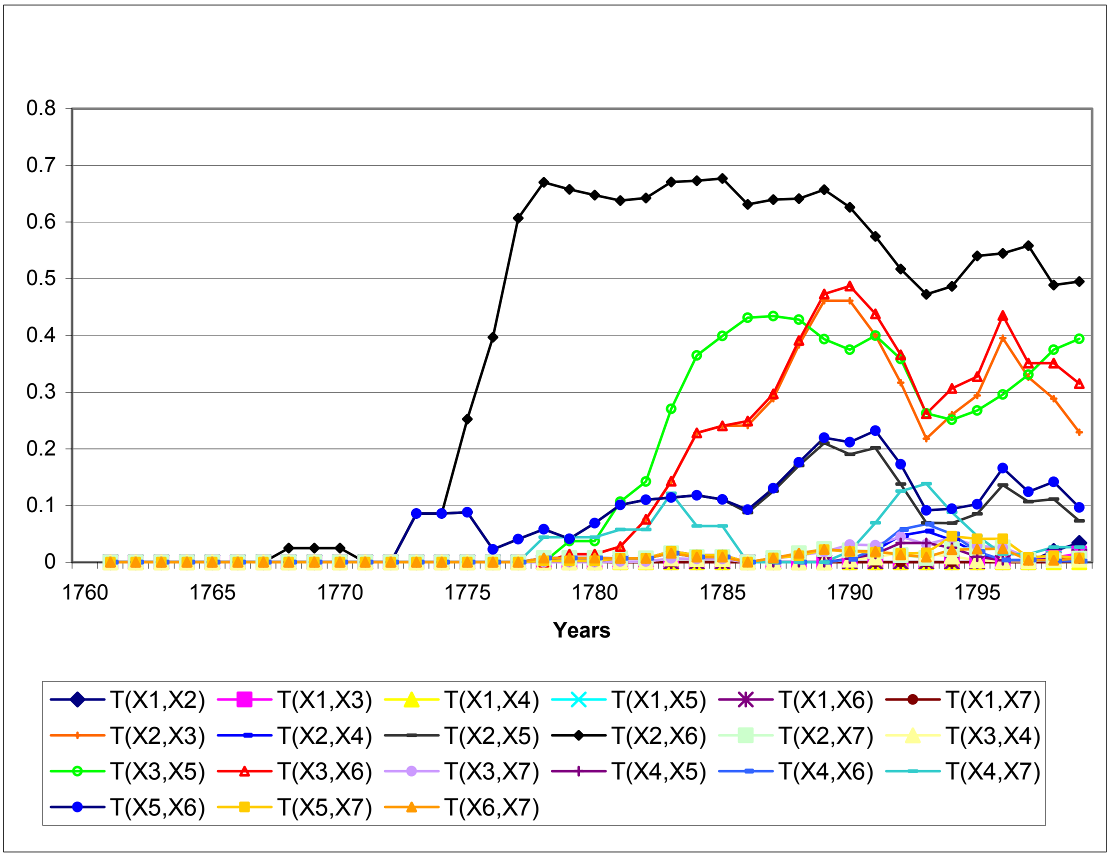

FIGURE 6. Two-dimensional mutual information (T) for steam engine designs

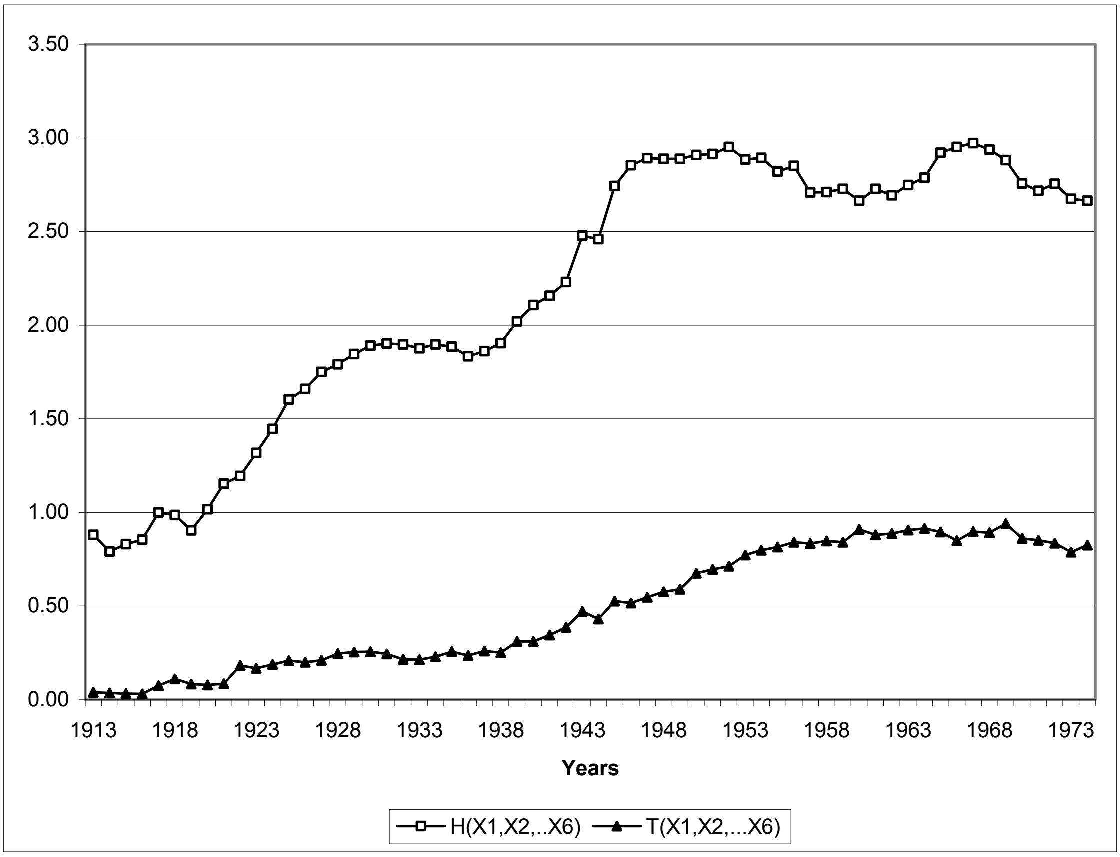

FIGURE 7. N-dimensional entropy (H) and mutual information (T) for aircraft designs

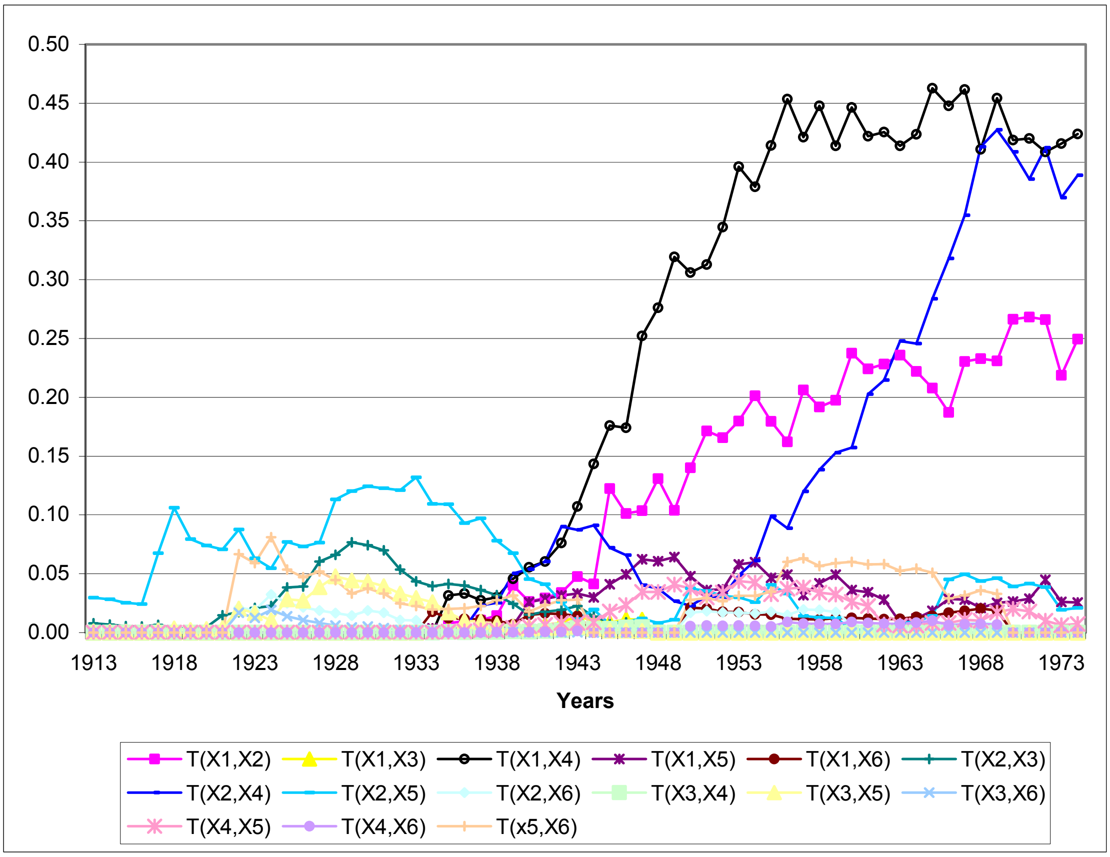

FIGURE 8. Two-dimensional mutual information (T) for aircraft designs

FIGURE 9. N-dimensional entropy (H) and mutual information (T) for helicopter designs

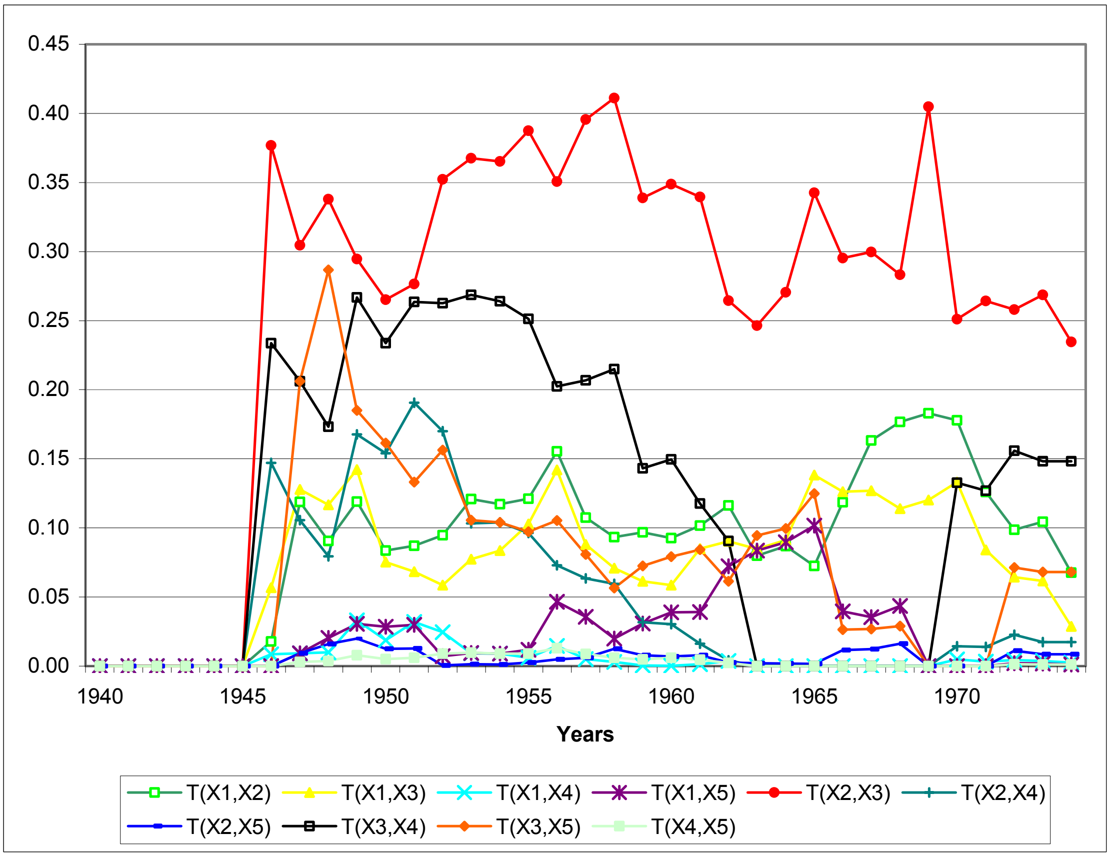

FIGURE 10. Two-dimensional mutual information (T) for helicopter designs

Eindhoven Centre for Innovation Studies

W O R K I N G P A P E R S

Ecis working papers 2001-2002 (September 2002):

01.01 H. Romijn & M. Albu
Explaining innovativeness in small high-technology firms in the United Kingdom

01.02 L.A.G. Oerlemans, A.J. Buys & M.W. Pretorius
Research Design for the South African Innovation Survey 2001

01.03 L.A.G. Oerlemans, M.T.H. Meeus & F.W.M. Boekema
Innovation, Organisational and Spatial Embeddedness: An Exploration of Determinants and Effects

01.04 A. Nuvolari
Collective Invention during the British Industrial Revolution: The Case of the Cornish Pumping Engine.

01.05 M. Caniëls and H. Romijn
Small-industry clusters, accumulation of technological capabilities, and development: A conceptual framework.

01.06 W. van Vuuren and J.I.M. Halman
Platform driven development of product families: Linking theory with practice.

01.07 M. Song, F. Zang, H. van der Bij, M.Weggeman
Information Technology, Knowledge Processes, and Innovation Success.

01.08 M. Song, H. van der Bij, M. Weggeman
Improving the level of knowledge generation.

01.09 M.Song, H. van der Bij, M. Weggeman
An empirical investigation into the antecedents of knowledge dissemination at the strategic business unit level.

01.10 A. Szirmai, B. Manyin, R. Ruoen
Labour Productivity Trends in Chinese Manufacturing, 1980-1999

01.11 J.E. van Aken
Management research based on the paradigm of the design sciences: the quest for tested and grounded technological rules

01.12 H. Berends, F.K. Boersma, M.P.Weggeman
The structuration of organizational learning

01.13 J.E. van Aken
Mode 2 Knowledge production in the field of management

01.14 A. Cappelen, F. Castellacci, J. Fagerberg and B. Verspagen
The impact of regional support on growth and convergence in the European Union

01.15 W. Vanhaverbeke, G. Duysters and B. Beerkens
Technological capability building through networking strategies within high-tech industries

01.16 M. van Birgelen, K. de Ruyter and M. Wetzels
The impact of attitude strength on the use of customer satisfaction information: An empirical investigation

01.17 M. van Birgelen, K. de Ruyter A. de Jong and M. Wetzels
Customer evaluations of after-sales service contact modes: An empirical analysis of national culture's consequences

01.18 C. Keen & M. Wetzels
E-tailers versus retailers: which factors determine consumer preferences

01.19 J.E. van Aken
Improving the relevance of management research by developing tested and grounded technological rules

02.01 M. van Dijk
The Determinants of Export Performance in Developing countries: The Case of Indonesian manufacturing

02.02 M. Caniëls & H. Romijn
Firm-level knowledge accumulation and regional dynamics

02.03 F. van Echtelt & F. Wynstra
Managing Supplier Integration into Product Development: A Literature Review and Conceptual Model

02.04 H. Romijn & J. Brenters
A sub-sector approach to cost-benefit analysis: Small-scale sisal processing in Tanzania

02.05 K. Heimeriks
Alliance Capability, Collaboration Quality, and Alliance Performance: An Integrated Framework.

02.06 G. Duysters, J. Hagedoorn & C. Lemmens
The Effect of Alliance Block Membership on Innovative Performance

02.07 G. Duysters & C. Lemmens
Cohesive subgroup formation: Enabling and constraining effects of social capital in strategic technology alliance networks

02.08 G. Duysters & K. Heimeriks
The influence of alliance capabilities on alliance performance: an empirical investigation.

02.09 J. Ulijn, D. Vogel & T. Bemelmans
ICT Study implications for human interaction and culture: Intro to a special issue

02.10 A. van Luxemburg, J. Ulijn & N. Amare
The Contribution of Electronic Communication Media to the Design Process: Communicative and Cultural Implications

02.11 B. Verspagen & W. Schoenmakers
The Spatial Dimension of Patenting by Multinational Firms in Europe

02.12 G. Silverberg & B. Verspagen
A Percolation Model of Innovation in Complex Technology Spaces

02.13 B. Verspagen
Structural Change and Technology. A Long View

02.14 A. Cappelen, F. Castellacci, J. Fagerberg and B. Verspagen
The Impact of Regional Support on Growth and Convergence in the European Union

02.15 K. Frenken & A. Nuvolari
Entropy Statistics as a Framework to Analyse Technological Evolution