A Speed Limit For Evolution

R. P. Worden†

Logica UK Ltd, Betjeman House, 104 *Hills Road, Cambridge CB2* 1*LQ, U.K.*

An upper bound on the speed of evolution is derived. The bound concerns the amount of genetic information which is expressed in observable ways in various aspects of the phenotype. The genetic information expressed in some part of the phenotype of a species cannot increase faster than a given rate, determined by the selection pressure on that part. This rate is typically a small fraction of a bit per generation. Total expressed genetic information cannot increase faster than a species-specific rate—typically a few bits per generation. These bounds apply to all aspects of the phenotype, but are particularly relevant to cognition. As brains are highly complex, we expect large amounts of expressed genetic information in the brain—of the order of 100 kilobytes—yet evolutionary changes in brain genetic information are only a fraction of a bit per generation. This has important consequences for cognitive evolution. The limit implies that the human brain differs from the chimpanzee brain by at most 5 kilobytes of genetic design information. This is not enough to define a Language Acquisition Device, unless it depends heavily on pre-existing primate symbolic cognition. Subject to the evolutionary speed limit, in changing environments a simple, modular brain architecture is fitter than more complex ones. This encourages us to look for simplicity in brain design, rather than expecting the brain to be a patchwork of *ad hoc* adaptations. The limit implies that pure species selection is not an important mechanism of evolutionary change.

© 1995 Academic Press Limited

1. Introduction

1.1. THE EVOLUTIONARY SPEED LIMIT

Since Darwin's time, the issue of the rates of evolution has been controversial. Darwin was a gradualist, and his belief that evolution is a slow continuous process has been the mainstream view ever since.

The theory of punctuated equilibrium (Eldridge & Gould, 1972) challenged this orthodoxy, suggesting that evolution proceeds through periods of comparatively rapid change separated by long periods of stasis. How fast can this rapid change be? What is the fastest evolutionary change? Are there any limits on the speed of evolution, which follow just from the nature of the selective process?

Two features of evolution are rarely questioned:

(i) The rate of evolution of any trait depends on the strength of selection on that trait—weak selection pressures (which give rise to only few deaths per thousand population) only cause slow changes in phenotype.

(ii) A species can only sustain a limited selection pressure; under too much pressure, it will die out.

We can combine these well-known facts to make a precise mathematical bound—a speed limit for evolution—which has important consequences for the evolution of many characteristics, particularly of cognition.

The speed limit bounds the rate of increase of the *genetic information in the phenotype* (GIP) which is a property of a population and is measured in bits. A population with some trait narrowly clustered around a central value has more GIP than another population, in which the same trait is more broadly spread. Halving the width of spread of some trait increases the GIP by one bit. Stabilizing selection increases the GIP, clustering individuals more closely around the optimum value of a trait for the species

† E-mail: wordenr@gled.logica.co.uk.

habitat. GIP is related (indirectly, through processes of development) to the raw information in the genotype, and is less than it. So GIP can be interpreted as that part of the genotype information which influences the phenotype.

The *total speed limit* states that (regardless of population size) the time-averaged increase in a species' GIP per generation is less than twice the log of the number of offspring per parent—typically 2–8 bits for mammals. The *partial speed limit* restricts the rate of increase of the GIP in any related group of traits to a smaller number (typically a fraction of a bit per generation) which depends on the strength of selection pressure on those traits.

An intuitive rationale for the total limit is as follows: if an animal has $2^3 =$ eight offspring, and only one survives to maturity, then this event conveys up to three bits of information from the environment into the surviving genotype and phenotype; animals which have eight offspring can accumulate up to three bits per generation of useful genetic information. These three bits will be spread across traits according to the selection pressure on them.

Showing that a bound like this holds even in the presence of sexual reproduction, mutations, large or small populations, and so on, is the main result of the paper. The bounds are stated precisely later. At first sight they may not seem restrictive; rarely is stabilizing selection severe enough to halve the spread of some trait in one generation, to add one bit of GIP. However, a phenotype is made up of many thousands of traits, and has a very large GIP. Selection may act on many traits at once, increasing the GIP in each trait by a small amount. The sum of all these small increases is limited by the bound; with very many traits, the bound is restrictive. This is particularly relevant to the evolution of intelligence. We expect the design of brains to be information-intensive; for higher animals, we expect a very large amount of GIP in the brain—but the bound severely limits the rate of change of this GIP. Therefore it has important consequences, both for general brain architecture and for particular facets of cognition.

1.2. SUMMARY OF RESULTS

Section 2 derives the speed limit, initially in a mathematical model of evolution that builds in some idealized assumptions of an infinite population, phenotypes which factor into independent groups of traits, discrete-valued traits, homogeneous initial population and non-overlapping generations. The model includes environmental influences on phenotype development, any possible form of selection pressure, and sexual reproduction with random mating.

GIP is defined. In the model, I prove the partial speed limit—that the average rate of increase in GIP for any facet of the phenotype is less than twice the selection pressure on the facet. The maximum rate is typically much less than one bit per generation. From this it follows that the (time-averaged) total rate of increase in GIP is less than a species-specific rate, typically a few bits per generation.

I then show, using more qualitative arguments, that the speed limit still holds under a range of biologically more realistic assumptions, including:

- mutations
- crossing
- finite populations
- inbreeding
- temporarily isolated sub-populations
- continuous-valued traits
- linkage between groups of traits
- non-homogeneous initial populations
- overlapping generations
- population saturation effects.

In most of these cases, the realistic complications tend (if anything) to produce slower rates of evolution, so that the speed limit as proved is very much an upper bound.

Section 3 discusses some implications of the speed limit for the evolution of cognition. The design of the brain is information-intensive: we expect large amounts of GIP (probably of the order of 100 Kbytes) in the brain. The speed limit restricts the increase of brain GIP to a small fraction of a bit per generation. Some consequences are:

- The amount of extra design information in the human brain, compared to the chimpanzee brain, is less than 5 Kbytes. This is not enough to define a language faculty *ab initio*; so language must be built largely on cognitive faculties we share with chimps.
- The speed limit enables us to predict which aspects of animal behaviour and cognition may be innate, and which are necessarily learned.
- For any functionality, a simple modular brain design can evolve faster, and have higher limiting fitness, than more redundant designs. We therefore expect brains to be the simplest and most modular possible for their functionality.

In general, the limit provides strong quantitative constraints on the rate of any change in cognition or behaviour, in response to other changes in a species or its environment.

Section 4 discusses other implications of the speed limit:

- I conjecture that the speed of evolution sometimes comes quite close to the limit, and that the genetic mechanisms enabling it to do so evolved long ago, before the evolution of eukaryotic cells.
- The theory of punctuated equilibrium is discussed in the light of the speed limit. I show that pure species selection has not been an important force in recent evolution. The anatomical changes seen in the fossil record may have occurred in a punctuated manner, since they need not involve large amounts of GIP; but equally, GIP may have accumulated in periods of apparent stasis.
- The speed limit applies not only to natural systems, but also to the computing technique of genetic algorithms.

2. The Speed Limit

In this section I derive the speed limit, initially in a mathematical model of evolution under the following conditions:

- infinite population
- phenotypes which factor into independent groups of traits
- discrete-valued traits
- homogeneous initial population
- non-overlapping generations
- environmental influences on phenotype development
- any possible form of selection pressure
- sexual reproduction with random mating

I then show how the result still holds under biologically more realistic conditions.

2.1. SUB-PHENOTYPES AND SELECTION PRESSURE

We intend to derive a limit not just for the total rate of evolution for a species, but also for the rate of evolution of a specific trait or group of traits, whose evolution can be considered (for some sound biological reasons) to be approximately independent of the evolution of other traits. Without some form of independence it would not be possible to prove anything about the evolution of a group of traits, as it would always be coupled to other factors involving other traits.

For the mathematical model of evolution, therefore, we need to make precise the idea of independence between groups of traits. For this purpose, we call each independent group of traits a *sub-phenotype*. We regard the phenotype as a product of several separate sub-phenotypes which are independent in a sense to be defined below, and are denoted by a sub-phenotype index $\mu = 1, \ldots Z$. Any animal has a value for each of its sub-phenotypes, denoted by an index i, which ranges from 1 to N_μ to denote all possible values of the subphenotype. For the moment we shall discuss only discrete-valued sub-phenotypes, for which N_μ is finite; although it may be very large. The extension to continuous values is discussed in Section 3.

A sub-phenotype is a group of related traits. For instance, a species might have distinct sub-phenotypes for its immune system, digestive system, circulatory system, cognitive system and so on. The different sub-phenotypes are assumed to be independent in the following senses:

(i) The sub-phenotypes are statistically independent in the population; the probability of any animal having a full phenotype—a set of values $(i, j, k, \ldots)$ for sub-phenotypes $\mu = (1, 2, 3, \ldots Z)$ —is a product of factors for the different sub-phenotypes:

$$P(i, j, k, \ldots) = \prod_\mu t_{i\mu}, \qquad (1)$$

where $t_{i\mu}$ is the probability that sub-phenotype μ has value i. This requires that there is no gene which is expressed in more than one sub-phenotype, and the linkage between genes for different sub-phenotypes is weak.

(ii) The sub-phenotypes have independent effects on fitness; for any individual the probability of survival to adulthood, S, is a product of factors for the different sub-phenotypes:

$$S(i, j, k, \ldots) = \prod_\mu s_{i\mu}, \qquad (2)$$

where all $s_{i\mu}$ are in the range $0 \leqslant s_{i\mu} \leqslant 1$. Each survival probability factor $s_{i\mu}$ for an individual depends on the value i which it has for the sub-phenotype μ.

The proportion A of animals which survive to adulthood is then:

$$A = \sum_{ijk\ldots} P(i, j, k, \ldots) S(i, j, k, \ldots) = \prod_\mu A_\mu \qquad (3)$$

$$A_\mu = \sum_i t_{i\mu} s_{i\mu}. \qquad (4)$$

Each sub-phenotype survival probability A_μ is in the range $0 \leqslant A_\mu \leqslant 1$.

We define the logarithmic selection pressure on the sub-phenotype μ as:

$$V_\mu = -\log_2 A_\mu. \qquad (5)$$

V_μ depends on how sharply the survival probabilities $s_{i\mu}$ peak in some narrow range of phenotypes i, relative to the range occupied by the species (to those i with largest $t_{i\mu}$). V_μ measures, on a logarithmic scale, how much the population is diminished per generation through unfitness in the sub-phenotype. For instance, if $s_{i\mu} = 1$ for all i, there is no selection pressure on the sub-phenotype μ, and $V_\mu = 0$. If the population is halved each generation through unfitness in that sub-phenotype, $V_\mu = 1$. V_μ is related to the "genetic load" defined by Crow (1958).

Suppose that each adult produces, on average, γ offspring. Over many generations, if $A < 1/\gamma$ (i.e. if fewer than one of these offspring survive to maturity), the species will dwindle to extinction. If we denote a time average over many generations by square brackets [ ], expressing A as a product of its factors A_μ then gives:

$$[V] = \sum_\mu [V_\mu] < \log_2 \gamma. \qquad (6)$$

Since $\log_2 \gamma$ is generally a small number, of the order of 1–3 for typical mammals, this is a tight limit on the total selection pressure, averaged over time.

We can now describe the process of evolution as it operates across the different sub-phenotypes. When a new environmental challenge arises in some sub-phenotype μ, the species is probably ill-adapted for it, os the selection pressure V_μ is strong. Action of this selection pressure tends to cluster the sub-phenotype on the favoured values i; so that average survival probabilities increase. As the species adapts, the selection pressure on that sub-phenotype effectively becomes weaker. So each selection pressure V_μ will show occasional peaks, followed by a long subsiding tail. This happens independently for each sub-phenotype μ; but subject to the constraint that the time-average of the total selection pressure $[V]$ over many generations must stay below the limit $\log_2 \gamma$. This process is illustrated in Fig. 1.

2.2. GENETIC INFORMATION IN THE PHENOTYPE

The concept of "useful genetic information" that we shall use is one of genetic information expressed in all the sub-phenotypes. Consider one sub-phenotype μ with discrete values i, with $1 \leqslant i \leqslant N_\mu$. The probability
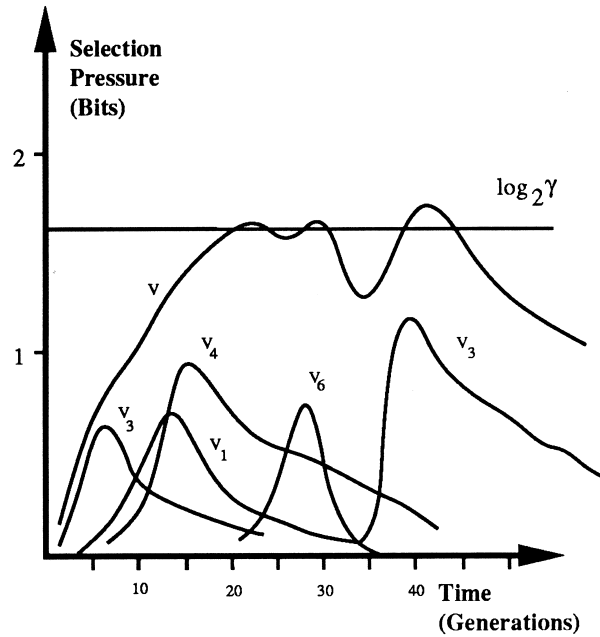

FIG. 1. Variation of selection pressures for different sub-phenotypes with time. The time average of the total selection pressure must not exceed $\log \gamma$, where γ is the average number of offspring per adult.

of finding an individual with value i for sub-phenotype μ is $t_{i\mu}$, so that:

$$\sum_i t_{i\mu} = 1. \qquad (7)$$

When the probabilities of all the different values i are equal, $t_{i\mu} = 1/N_\mu$ for all i, the genetic information expressed in the sub-phenotype μ (the GIP) is at its minimum, and we will define this minimum GIP to be zero. For other possible values of $t_{i\mu}$, the genetic information expressed in the sub-phenotype μ is defined by the usual information-theoretic measure, relative to the zero baseline:

$$G_\mu = \log_2 N_\mu + \sum_i t_{i\mu} \log_2(t_{i\mu}). \qquad (8)$$

So GIP is a measure of how much the observed values i in a large population tend to cluster on a few values; if there is no clustering, $G_\mu = 0$, and if there is complete clustering on one value, $G_\mu = \log_2 N_\mu$. It is a property of the population, not of an individual.

Particularly for sub-phenotypes whose values involve continuous parameters, there are some apparent problems in interpreting the GIP:

1. There might be dimensions of the sub-phenotype which we are not aware of. If so, there may be some positive contribution to the GIP which we fail to measure.

2. What if we thought two or more variables are independent dimensions of a sub-phenotype, when they were not, because animals in the population can never have totally uncorrelated values? For a centipede, 100 leg lengths are not independent variables; if we thought they were, we would over-estimate a part of the GIP by a factor 100.

Problem 1 causes no difficulty for the speed limit we shall derive, because it is a limit on the rate of increase of full GIP, observed and unobserved. If we observe less GIP, it applies a *fortiori* as a bound on the rate of increase of the GIP we can observe. Problem 2 would cause us to over-estimate the GIP, which is more serious for the speed limit; however it can be resolved by relating the GIP to information in the genotype, as we shall do in the next sub-section. In practice, neither problems 1 nor 2 are serious when we consider GIP in the brain, which we do not so much by looking at brains, as by looking at their performance, as we shall see in Section 3.

The total GIP is the sum over all sub-phenotypes:

$$G = \sum_\mu G_\mu. \qquad (9)$$

Intuitively, we expect the GIP to be less than the raw information content of the genotype; in practice it is usually much smaller.

From now on we shall consider one sub-phenotype μ at a time, and drop the index μ from some terms.

2.3. GENOTYPES AND DEVELOPMENT

The model assumes that sub-phenotypes are statistically independent in the population, which requires that no part of a sub-phenotype is controlled by a gene which also controls part of another sub-phenotype. Therefore, in effect, each sub-phenotype has its own dedicated set of genes—its own sub-genotype.

Suppose this sub-genotype consists of B_μ base pairs in each DNA strand of a chromosome. Since each base can be any one of four amino acids, the amount of information about the sub-phenotype in one germ cell (sperm or egg) is $(2B_\mu-1)$ bits, and the amount of information in one diploid cell is $(4B_\mu-2)$ bits. However, this definition of the amount of genetic information in one individual of the population, is not the one we will use in deriving the speed limit. We shall define genotype information in a population-dependent way that measures the spread or clustering of genotypes in the population, matching the definition of GIP.

We shall use two such population-dependent information measures: one for germ cells and the second for diploid cells.

In one germ cell, there are $M_\mu = 2^{(2B_\mu-1)}$ possible values of the haploid genotype; M_μ distinct DNA sequences possible for the sub-phenotype. Each possible sequence is denoted by an index j, which runs from 1 to M_μ. q_j is the probability that any random germ cell taken from the population will have the value j, so that:

$$\sum_j q_j = 1. \qquad (10)$$

As for the GIP, we define the *haploid genotype information* I_μ for the sub-phenotype μ to be zero when there is no clustering on any genotypes q_j, and all $P(q_j)$ are equal to $1/M_\mu$:

$$I_\mu = \log_2 M_\mu + \sum_j q_j \log_2(q_j). \qquad (11)$$

So the maximum possible value for I_μ is $2B_\mu - 1$, if all germ cells in the population have the same DNA sequence for the sub-genotype (if that sequence were fixed in the population).

An animal's sub-genotype for the sub-phenotype μ is made up of two parts, j and k, one from each parent. The probability of finding the pair (j, k), will be written as Q_{jk}. Because of the symmetry between parental genes, $Q_{jk} = Q_{kj}$.

We define the *diploid genotype information J_μ* for the sub-phenotype μ to zero when there is no clustering on any genotypes (j, k), and all $Q_{jk} = 1/M_\mu^2$:

$$J_\mu = 2 \log_2 M_\mu + \sum_{jk} Q_{jk} \log_2(Q_{jk}). \qquad (12)$$

So the maximum possible value for J_μ is $4B_\mu - 2$, for a pure-bred population where all animals in the population have the same sub-genotype (j, k).

If each fertile animal mates with another drawn at random from the whole population, then

$$Q_{jk} = q_j q_k, \qquad (13)$$

from which it follows that

$$J_\mu = 2I_\mu. \qquad (14)$$

When there is any degree of inbreeding (an above-average tendency for j and k to be the same, having come from the same ancestor), we have $J_\mu > 2I_\mu$; this case will be discussed in Section 3.

The value of the sub-phenotype i depends on the sub-genotype (j, k), but may also be influenced by environmental factors during development. The probability of any sub-phenotype value i depends on the distribution of genotypes in the population:

$$t_{i\mu} = \sum_{jk} D_{ijk} Q_{jk}. \qquad (15)$$

The development matrix D_{ijk} must obey several constraints. First, all $D_{ijk} \geq 0$. Second, to ensure that (7) is true for any possible distribution of Q_{jk}, we require

$$\sum_i D_{ijk} = 1 \quad \text{for all } j, k. \tag{16}$$

Third, we require that the minimum-information configuration of genotypes, $Q_{jk} = 1/M_\mu^2$ and $J_\mu = 0$, is also the minimum-information configuration of phenotypes, $t_{i\mu} = 1/N_\mu$ and $G_\mu = 0$. This requires

$$\sum_{jk} D_{ijk} = M_\mu^2/N_\mu \quad \text{for all } i. \tag{17}$$

(Without environmental influences, when all $D_{ijk} = 0$ or 1, eqn (16) is only possible when there are more possible genotypes than phenotypes, $M_\mu^2 > N_\mu$.)

Within these constraints, the coefficients D_{ijk} can represent any mixture of genetic and environmental influences on phenotype. The value of the phenotype i can be fully determined genetically, depending on any or all of the genes in (j, k); in this case, for every (j, k) $D_{ijk} = 1$ for just one i, and $D_{ijk} = 0$ for all other i. Alternatively, if for each (j, k) there are many nonzero D_{ijk}, the phenotype value $t_{i\mu}$ is mainly determined by random environmental influences. Any intermediate form of dependency is possible.

From the definitions given so far, we can prove† that

$$G_\mu \leq J_\mu = 2I_\mu. \tag{18}$$

Thus the genetic information G_μ expressed in the phenotype is less than the total genetic information $2I_\mu$. This result is intuitively not surprising. If there is some random environmental influence on development, we would expect it to "smear out" distinctions between genotypes and lose information: if there is little environmental influence, then at least one genotype must map onto each phenotype to make eqn (16) true, and a many-one mapping also loses information.

2.5. PROVING THE SPEED LIMIT

I shall first state two speed limits, using the above definitions. If a species has existed for n generations, the rate of evolution of the sub-phenotype μ is defined as the average increase of the GIP G_μ per generation, dG_μ/dn, over many generations; similarly for the total rate of evolution $dG/dn = \Sigma_\mu \, dG_\mu/dn$. The partial speed limit states that over many generations

† The proof uses the inequality $\ln(x) \leq (x-1)$, deployed so that in cases where (18) is an equality, $x = 1$ and $\ln(x) = (x-1) = 0$.

the average rates obey:

$$\left[\frac{dG_\mu}{dn} \right] = 2\eta_\mu [V_\mu]$$

$$\eta_\mu \leq 1 \quad \text{for all } \mu \tag{19}$$

and the total speed limit is:

$$\left[\frac{dG_\mu}{dn} \right] = 2\eta \log_2 \gamma$$

$$\eta \leq 1 \tag{20}$$

η_μ and η will be called evolutionary efficiencies. Equation (20) follows simply from eqns (19) and (6).

The strategy for deriving the limit (19) is as follows: first use the survival probabilities for sub-phenotypes, $s_{i\mu}$, to calculate expressions for the survival probabilities of genotypes, σ_j. Show that for any possible form of σ_j, there is a limit on the increase of haploid genotype information I_μ. Then the speed limit (19) for $[dG_\mu/dn]$ follows from this limit on I_μ and $G_\mu \leq 2I_\mu$, [eqn. (18)].

The survival probability of a diploid genotype Q_{jk} is given by:

$$\sigma_{jk} = \sum_i D_{ijk} s_{i\mu}, \tag{21}$$

so that

$$\sum_{jk} Q_{jk} \sigma_{jk} = \sum_{jk} Q_{jk} \sum_i D_{ijk} s_{i\mu}$$

$$= \sum_i t_{i\mu} s_{i\mu} = A_\mu. \tag{22}$$

Equation (22) says that the proportion which survive to adulthood is the same, whether you calculate it in terms of phenotypes of genotypes. All σ_{jk} obey $0 \leq \sigma_{jk} \leq 1$.

Next consider what happens to one of the haploid genotypes j in one generation. Through random mating, it gets paired with another haploid genotype k, with probability q_k; then the pair have a probability of surviving σ_{jk}. So the probability of j surviving to the adulthood of its bearer is:

$$\sigma_j = \sum_k q_k \sigma_{jk}. \tag{23}$$

All σ_j obey $0 \leq \sigma_j \leq 1$. With random mating [eqn (13)]:

$$\sum_j q_j \sigma_j = \sum_{jk} Q_{jk} \sigma_{jk} = A_\mu. \tag{24}$$

Consider what happens to a pool of haploid genotypes under this selective pressure for n generations, $g = 1, \ldots, n$, starting from a uniform distribution $q_j = 1/M_\mu$, with minimum information $I_\mu = 0$. The per-generation survival probability A_μ depends on the generation g. Define

$$F_\mu(n) = \prod_{g = 1, \ldots, n} A_\mu(g) \qquad (25)$$

$$W_\mu(n) = \sum_{g = 1, \ldots, n} V_\mu(g) = -\log_2[F_\mu(n)]. \qquad (26)$$

$W_\mu(n)$, being a sum of $V_\mu(g)$ over generations, is called the *cumulative logarithmic selection pressure*.

Usually, to calculate the information content I_μ by (11), we re-normalize the q_j after each generation of selection, so that $\Sigma q_j = 1$. Suppose instead of q_j, we use $\bar{q}_{j'} = Tq_j$, so that $\Sigma \bar{q}_j = T$. Equation (11) then becomes:

$$I_\mu = \log M_\mu + \frac{1}{T} \Sigma \bar{q}_j \log \bar{q}_j - \log T. \qquad (27)$$

If we simply let the survival probabilities σ_j act in each generation, and do not renormalize to $\Sigma q_j = 1$, then in each generation $\Sigma \bar{q}_j$ is multiplied by a factor $\Sigma_j q_j \sigma_j = A_\mu$. So at the end of n generations, $T = \Sigma \bar{q}_j = F_\mu(n)$.

Now set $\bar{q}_{max} = \text{Max}(\bar{q}_j)$ and $k_j = \bar{q}_j/\bar{q}_{max}$. From (26) and (27) we get:

$$I_\mu - W_\mu = \log M_\mu + \log(\bar{q}_{max}) + \frac{\bar{q}_{max}}{F_\mu} \Sigma k_j \log k_j. \quad (28)$$

Since $0 \leqslant k_j \leqslant 1$ for all j, the last term is negative. As we started from a homogeneous population $q_j = 1/M_\mu$, and all the survival probabilities obey $\sigma_j \leqslant 1$, this implies $\bar{q}_{max} \leqslant 1/M_\mu$ so the sum of the first two terms in (28) is negative. So $(I_\mu - W_\mu)$ is negative. Combining this with (18) gives:

$$G_\mu(n) \leqslant 2W_\mu(n). \qquad (29)$$

Over n generations of selection, starting from a homogeneous population, the average rate of increase of GIP is $[dG_\mu/dn] = (G_\mu(n) - G_\mu(0))/n = G_\mu(n)/n$. The average selection pressure is $[V_\mu] = W_\mu(n)/n$, so that:

$$\left[\frac{dG_\mu}{dn}\right] \leqslant 2[V_\mu]. \qquad (30)$$

This is the result we want, equivalent to (19), that the time-averaged rate of increase of GIP is less than twice the time-averaged logarithmic selection pressure. It holds for any form of survival probabilities s_i (or equivalently σ_j) which may even be time-dependent.

When survival probabilities depend on the numbers and behaviour of con-specifics, so that a species may evolve towards an evolutionary stable state (Maynard Smith & Price, 1973; Maynard Smith, 1982), the survival probabilities s_i are time-dependent, since they depend on the changing average properties of conspecifics. The proof of the speed limit includes time-dependent survival probabilities, so applies equally to evolution towards an ESS.

When all survival probabilities s_i are less than some s_{max}, we can similarly prove the stronger limit:

$$\left[\frac{dG_\mu}{dn}\right] \leqslant 2([V_\mu] + \log_2(s_{max})). \qquad (31)$$

An approximate form is useful when all the survival probabilities $s_{i\mu}$ are in a narrow range, with standard deviation $\pm d_\mu$ percent. Then, using $\ln(2) \cong 0.7$ gives approximately $V_\mu + \log_2(s_{max}) = d_\mu \sqrt{2}/70 = d_\mu/40$, so that

$$\left[\frac{dG_\mu}{dn}\right] \leqslant d_\mu/20. \qquad (32)$$

Recall that the total speed limit [eqn (20)] follows simply from the partial limit (19) and the limit on total selection pressure (6). Thus for mammals, the total increase in GIP per generation is typically not more than 5 bits per generation.

2.6. REALISTIC COMPLICATIONS

Having derived the speed limit in an idealized mathematical model of evolution, we need to show that the complications of real evolution do not somehow violate the limit. There are several complications to consider. The discussions of these will often not be as mathematically clear-cut as the original derivation; they are sometimes more in the spirit of plausibility arguments.

We can sometimes devise scenarios in which the speed limit is temporarily violated; however, to show a real violation, it is not enough just to create such a scenario; one must also show that it happens with reasonably high probability. A "Maxwell's Demon" scenario, where a large part of a population suddenly undergoes the same very favourable mutation, can in principle happen and will violate the limit; but it happens with such vanishingly small probability that we can ignore it.

We shall say that the speed limit is obeyed "on average" if the probability of exceeding the limit by G bits of the GIP over some interval is of order 2^{-G} or less. In this case the probability of any significant

gain over the speed limit (e.g. 50 bits or more of the GIP) is vanishingly small.

The speed limit holds in the presence of the following complications of actual evolution:

(*a*) *Mutations*: In a realistic probabilistic model, mutations always lead to a decrease of the haploid genotype information I_μ, so we can rigorously show that the limit is still obeyed in the presence of mutations.

(*b*) *Crossing*: Similarly, in a realistic model of crossing, we can show that crossing always decreases the diploid genotype information J_μ. This is not quite the same as proving that crossing always decreases I_μ, but is a powerful plausibility argument that it does so. In that case, crossing will not violate the limit.

(*c*) *Finite populations*: In a finite effective breeding population N_e, the evolutionary response to small selection pressures (which gives variations in fitness less than $1/N_e$) tend to get washed out by the effects of neutral molecular evolution (Kimura, 1983), so the speed limit is respected by these.

For larger selection pressures, the effect of finite population can be modelled by a random Poisson process on the genotype frequencies q_j; this process does not, on the average, increase I_μ, so does not lead to violations of the bound.

Finally we need to consider a fluke mutation, which leads to an improvement in the GIP (for one individual) of G bits, an increase in fitness αG, and spreads through the whole population to fix in $\log(N_e)/\alpha G$ generations. These occur with probability $N_e 2^{-G}$ per generation and, as long as α is small, do not violate the bound on average. In the present mature state of species, all the easy wins (high-α mutations) have been done long ago, and α is expected to be very small.

(*d*) *Temporarily isolated sub-populations*: We can construct scenarios where two isolated sub-populations are subject to different selection pressures (e.g. on different sub-phenotypes), and are then re-merged in an environment subject to both pressures. For large increases in the GIP in both sub-populations, this can lead to a violation of the bound by up to a factor 2 (or N, for N isolated sub-populations). However, these scenarios use highly contrived sequences of selection pressures, which arguably occur only very rarely in nature. The sequences depend on the environment, so we probably cannot argue that they occur with probability less than 2^{-G}, but we can still argue that they are rare enough not to violate the bound.

(*e*) *Non-homogeneous initial populations*: Inspection of eqn (28) shows that the limit can be violated if the initial population is already well-adapted to the selection pressure (so $\bar{q}_{max} > 1/M_\mu$). These cases arise either by a fluke (which occurs with probability $< 2^{-G}$, so the bound is still true on average), or through earlier selection under the same selection pressure. In the latter case, we argue that the time-average over the whole process still obeys the limit, as was proved above.

Even at the tail-end of a selection process, it is unlikely that there will be genotypes j with $\sigma_j \cong 1$; it is far more likely that even the fittest have $\sigma_j \cong \sigma_{max} < 1$. In this case, as $\bar{q}_{max}$ is multiplied σ_{max} every generation, after a fairly small number of generations, $\bar{q}_{max} < 1/M_\mu$ and again the bound is true.

We have discussed cases where (probably because of previous selection) the population is already well-adapted, so initially $q_{max} > 1/M_\mu$. Cases where the population is initially ill-adapted to a new selection pressure are more common, so the genotypes j which will survive best after many generations are initially very uncommon; $q_j \ll 1/M_\mu$. In these cases, where an inappropriate GIP must be removed, as well as an appropriate GIP gained, the rate is yet further below the speed limit.

(*f*) *Irregular mapping from genotype to phenotype*: One may question the assumption used in the proof that the minimum-information configuration of genotypes ($q_j = 1/M_\mu$, $I_\mu = 0$) is also the minimum-information configuration of phenotypes ($t_i = 1/N_\mu$, $G_\mu = 0$). This assumption is approximately true because we expect an ill-defined distribution of phenotypes in the absence of genetic information. What if the distribution of phenotypes were highly non-uniform when the genotype information was at a minimum? In this case, the state $I_\mu = 0$ corresponds to $G_\mu = G_{min}$ with some large G_{min}.

There are two possibilities; either the non-uniformity of phenotypes is favourable (high-survival phenotypes have high prior probability) or it is unfavourable. In the former case, the speed limit may be violated, just as in the example above. In the latter case, the average speed of evolution is much slower than the limit. The first case (that a natural "non-genetic" pre-disposition of the phenotype anticipates a selection pressure) is a fluke with probability of order $2^{-G_{min}}$, and is much less likely than the latter case if, we assumed, G_{min} is large; so we expect the bound to be respected on average.

(*g*) *Inbreeding*: To some extent this can be regarded as a variant of isolated sub-populations; if isolated sub-populations cannot break the speed limit, neither can inbreeding. However, it merits discussion in its own right.

In an inbred population, $J_\mu > 2I_\mu$, so the proof of the speed limit no longer applies. However, we can argue that this extra information in the diploid genotype has nothing to do with environmental pressures, so does nothing to ensure a faster response to them. This argument is supported by the example of ultimate inbreeding—parthenogenesis—where we can show rigorously that the speed limit holds, and is stronger by a factor 2.

(*h*) *Continuous-valued traits*: The general speed limit derived above applies to discrete values $i = 1, \ldots, N_\mu$ of the sub-phenotype; but nowhere have we made the assumption that N_μ is small. So the proof can be applied with arbitrarily large N_μ to mimic the effects of continuous sub-phenotypes to arbitrary accuracy, showing that the speed limit still applies in those cases. In this respect, the proof for continuous variables is similar to deriving integration as a limiting case of a sum. Nevertheless, we can analyse specific cases of continuous phenotypes as a check of the general argument, and some of these give interesting results. A sub-phenotype which is a continuous variable with Guassian initial distribution, under Gaussian selection pressure, with additive inheritance from both parents, obeys the speed limit at all times; the rate of increase in GIP is always at least a factor of 4 below the limit.

(*i*) *Linkage between different groups of traits*: The assumption that different sub-phenotypes are independent, both genetically and in their impact on survival probabilities, appears to be a strong one. Might sub-phenotypes be interdependent in such a way as to violate the bound? This is not possible, because the definition of sub-phenotypes is a choice which we make to analyse particular aspects of evolution. If two sub-phenotypes μ and σ turn out not to be independent of one another, we can combine them into one sub-phenotype μ', and by the arguments given above, the bound applies to the combined sub-phenotype.

A more concrete argument comes from considering specific cases of linkage between sub-phenotypes. For instance, take the case where two sub-phenotypes are fully independent (survival in each sub-phenotype depends on different aspects of the environment) except for one gene locus which determines a trait in both of them. Selection pressure on one sub-phenotype might pick out an allele at this locus which is beneficial for the other sub-phenotype, thus increasing the amount of useful information expressed in the two sub-phenotypes at a combined rate above the bound. While cases like this can happen, they only do so when independent selection pressures on two different sub-phenotypes happen to line up so as to favour a shared allele; so they can only be flukes, with the reverse (unlucky) case occurring much more often;

therefore they will not violate the bound in the long term.

(*j*) *Overlapping generations*: Several simple models of reproduction with overlapping generations can be constructed and do not violate the bound. There seems to be no good reason why they should.

(*k*) *Extending the genotype*: The proof of the speed limit assumes that the sub-geneotype μ consists of a fixed number B_μ of DNA base pairs. Since different species have different amounts of DNA, this assumption must sometimes be violated; we need to satisfy ourselves that such events cannot in some way break the speed limit.

This comes down to a qualitative argument that extension of a DNA sequence is a random process, just like mutation; for such a random process you win some and you lose some—probably you lose a lot more than you win. Selection is needed to filter out the few winners, and there is no reason to expect that the selection process is any more efficient for genotype extensions than for mutations.

2.7. SUMMARY OF THE SPEED LIMIT

We have shown that the speed limit holds even in the presence of many of the complexities of actual evolution; the average increase of GIP per generation, dG_μ/dn, is less than twice the logarithmic selection pressure V_μ.

From working through model examples, it seems very hard for the speed of evolution to come very close to the limit (say within 20%); in practically all examples it is at least a factor of 2 slower, and more typically is four to ten times slower. To come very close to the limit would require a largely dominant sub-genotype j with some survival probabilities σ_j very close to 1. So, as a working hypothesis, we shall assume that the practical speed limit is four times slower than the theoretical limit. Then, for weak selection pressures, the limit (32) becomes:

$$\left[\frac{dG_\mu}{dn} \right] \leqslant d_\mu / 80. \tag{33}$$

For instance, a selection pressure that gives variations in survival probability of $\pm 10\%$ can, in practice, increase the GIP by only 1/8 bit per generation.

3. Implications of the Speed Limit for Cognition

3.1. DESIGN INFORMATION IN THE BRAIN

The evolutionary speed limit is mainly useful for discussing the rate of change of certain characteristics; however, as a prelude, we shall estimate the total

amount of genetic design information (GIP) in the mammalian brain. This figure is interesting, first as an indication of the size of problem that we are tackling when we study the brain, and second, for comparison with the differences of brain GIP between species (see Section 3.2 below). To do this, we can calculate upper and lower bounds to the amount of GIP in the brain.

The upper bound follows from the speed limit. First consider the GIP which might have accumulated since the emergence of mammals some 200 million years ago, over approximately 10^8 generations, each with a "genetic improvement budget" $\log_2(\gamma)$ of 2–4 bits. Using the "practical limit" $dG/dn < 0.5 \log_2(\gamma)$ gives a total improvement budget (for the whole GIP) of 1.5 bits per generation, or around 20 Mbytes.

Most of this GIP has been related to other aspects of the phenotype—the immune system, body forms and so on. However, cognition is an important component of survival for mammals, and is always under some selection pressure. Suppose that over the generations, there have, on average, been cognition-dependent variations in survival probability of up to $\pm 8\%$ (i.e. in any generation, those with the fittest brains survive not more than 8% better than the average). By eqn (33), this gives up to 0.1 bits of selection pressure per generation on cognition, giving a maximum possible mammal-specific GIP in the brain of the order of 1 Mbyte.

Cognition did not start with mammals; its origins are some 300 million years earlier in the Cambrian era, and many of the cellular components of brain architecture, being shared with non-mammalian orders, were possibly evolved in that period. For pre-mammalian cognitive evolution, we conservatively include another factor of 2, leading to an upper bound of roughly 2 Mbytes.

There are two different ways to get a lower bound on brain GIP:

 (i) From the neuro-anatomical complexity of the brain, estimate the minimum information required to specify these structures;
(ii) From the complexity of innate behaviour and cognitive abilities of some species, estimate the minimum information required to specify this.

The two are not additive; they are different estimates of the same quantity. In practice, however, they are both rather hard to calculate. Anatomically, the brain is a highly complex organ, which leads us to believe that the total amount of design information needed to specify it is large. If, for instance, there are 500 different types of neuron in the brain, and for each one we need 100 bytes to specify its particular physiology, and the anatomical distribution of its input and output

synapses, this leads to an estimate of 50 Kbytes. However, common properties of many similar types of neuron might be specified more economically; for this we should reduce the lower bound, to say, 10 Kbytes.

Innate behaviour, and other forms of innate performance of the brain, require innate information; all this must be somehow encoded as GIP in the structure of the brain. Consider some examples:

● Many animals show complex innate sequences of behaviour, for instance for courting, nest-building and care-giving. To specify one these sequences, its releasing stimuli, its synchronization and its terminating stimuli all requires information, probably amounting to hundreds or thousands of bits.
● All innate reactions, such as a fear response to certain stimuli which indicate predators, require a specification of the releasing stimulus. The more precise the stimulus (e.g. a particular predator), the more information is required to specify it.
● Some causal relations in the environment are learnt faster than others. For instance, rats can learn a taste–nausea association very fast. GIP in the brain is required to define which pairs of stimuli should be learnt fast.
● Some aspects of cognition depend on universal constraints that follow from the laws of physics or mathematics. For instance, to interpret a visual scene our brains exploit the laws of illumination, optical flow and perspective (Marr, 1982). To predict the motion of objects around us, our brains use the laws of motion, rigidity, elasticity, etc. These laws are highly specific, and their information content must be encoded in the brain. This too requires hundreds or thousands of bits of information.

The GIP of the brain must somehow encode all this information about many facets of the outside world, as well as the "processing engines" to use all the innate information, together with sense data, effectively. From this viewpoint, again, 10 Kbytes seems a sensible lower bound on the required design information.

So we have an approximate lower bound of 10 Kbytes, and an upper bound of around 2 Mbytes, for the total GIP in the design of the mammalian brain. The lower bound is much less than the design information in a modern computer chip; one suspects it is so low simply because of our lack of understanding of the brain (and the crude methods of estimating GIP used above) and that the real figure is much higher. For a working estimate, we might use a figure of 100 Kbytes, remembering that this is only an order-of-magnitude estimate, and that the real figure could differ by a factor 10 in either direction.

3.2. HUMAN INTELLIGENCE AND LANGUAGE

Compared to our nearest living relatives—chimpanzees—we have a much-expanded cranial capacity, much higher general intelligence and the unique and powerful capacity for language.

Because of this, many believe that the human brain exceeds other primate brains not only in size and power, but also in complexity, and that recent human evolution has added fundamentally new capabilities to the brain. Chomsky (1975) has argued that the capacity for language is not foreshadowed in any other primate brain, but arose *de novo* in mankind. This view is supported by Pinker & Bloom (1990), who argue that, while the human genotype differs from the chimp genotype by only around 1%, this amounts to some 10 Mbytes of information which should be plenty to specify a complete language engine.

We can examine these claims in the light of the speed limit. Our antecedents and the antecedents of chimpanzees diverged some 5–7 Myr ago, so there have been of the order of 350000 generations since the split. Suppose that the average number of children per couple is three, giving $\gamma = 1.5$ The practical limit on total GIP growth implies that our phenotype differs from that of chimps by at most $175000 \log_2 \gamma \cong 100000$ bits of GIP—already much less than the 10 Mbytes quoted by Pinker.

What proportion of these 100000 bits have contributed to the evolution of intelligence? What proportion of the cumulative selection pressure has been selection for intelligence, rather than for other attributes? For a typical hominid group, it seems likely that the difference in fitness between the least and most intelligent—the difference in their probability of survival to adulthood and successful reproduction—was not larger than $\pm 10\%$. This is because differences in intelligence between contemporaries are usually not large, and many factors other than intelligence (size, strength, resistance to disease, and plain good luck) are needed to survive. This reduces the selection pressure for intelligence by a further factor.

From eqn (33), $\pm 10\%$ variations in survival probability can contribute GIP of up to 1/8 bit per generation. So the useful genetic information in the human brain, beyond that in the chimp brain, is at most 40000 bits, or 5 Kilobytes. We can compare the 5 Kbytes extra design information in the human brain with the previous estimate of 100 Kbytes total GIP in the mammalian brain. The design information which distinguishes our brains from those of chimps is around 5% of the total.

Five Kbytes is equivalent to a computer program of around 300 lines. Experience suggests that the functionality one can express in 300 lines of program code is very limited; certainly not enough to design a complete facility for language learning and use, unless the underlying computing engine is already very well-adapted for the language task. Five Kbytes is the practical upper bound on the amount of useful extra design information in the human brain, if the rate of evolution is as fast as possible. For development of any fundamentally new cognitive facility such as a language acquisition device, we would expect a much lower rate of evolution, giving a figure much lower than 5 Kbytes.

So we are led to the conclusion that the main difference between the human brain and other primate brains is one of capacity and power rather than of design complexity, and that our language ability is largely built on some pre-existing mental capacity which other primates have, although they do not use it for language.

3.3. INNATE AND LEARNED BEHAVIOUR

The speed limit can give insights into the mix of innate and learned behaviour which we observe in many animals. As an example, consider the alarm calls of vervet monkeys observed by Cheney & Seyfarth (1990) in Amboseli National Park. These monkeys have an innate fear of avian predators; young vervets make alarm calls indiscriminately when they see any large bird. However, only about three species of eagle prey on vervets, and a young monkey soon learns (possibly by observing the reactions of older monkeys to his calls, not by observing predation events) which kinds are predators and which are not; adults only give the "eagle alarm" call when one of the predator species is seen.

Why does the divide between innate and learned behaviour occur where it does? We can analyse this question by considering the difference between three possible monkey species: species A with no innate fear of birds, species B with a general fear of birds, and species C with an innate fear of only certain species. Vervets are a type-B species.

Different amounts of GIP in the brain are required for the three; the least for A, more for B, and yet more for C. We can estimate the extra GIP $(G_C - G_B)$ required to have an innate fear reaction for just the crowned eagle, the martial eagle and the tawny eagle, based on the plumage, movement and so on. This extra information in the design of a monkey brain is possibly of the order of 100 bits, to specify the distinctive visual appearance of these species.

If vervets did not have a fairly efficient way of learning and transmitting this distinction socially, the gain in fitness from knowing the difference innately

might be significant. Fifty per cent of vervets die through predation, and the species would have to have either an indiscriminate fear of birds (which would often interfere with feeding and other activities), or an even higher risk of predation. The cost in fitness, V_μ, might well be of the order 0.1; in which case an innate fear of the correct avian species might evolve within 500 generations, consistent with the speed limit.

However, vervets have an efficient social learning system: infants follow the adults' example in predator precautions (in the case of eagles, not staying in trees where the eagles can easily take them), and soon learn from the adults to give alarm calls only when appropriate. The cost in fitness of learning the distinction, rather than knowing it innately, is possibly of the order of 0.01 (i.e. one vervet in a 100 dies before learning). So by the speed limit, a specific innate fear reaction cannot evolve in less than 5000 generations. Primates are mobile, adaptable species, whose success often depends on an ability to colonize new habitats. In particular, vervets live in several different habitats, with different avian predators in different areas. The learning mechanism helps them to do this, whereas an evolved innate fear, which takes 5000 generations to evolve, would not be much help.

We have considered the difference between type B species (innate fear of any bird, refined by learning) and a type C species (innate fear of specific birds). We can also consider the difference between type A species (no innate fear, with learned fear of specific birds) and type B. The required GIP difference ($G_B - G_A$) is smaller—say 40 bits—since specific species are not innately identified by a type-B monkey. While the social learning mechanism might still mean that the difference in fitness between type A and type B is small (say of the order of 0.01), avian predators are a universal feature of life for a small primate. So there have been many more than 2000 generations (required by the speed limit) for a general fear of birds to have become innate.

The evolutionary speed limit leads us to expect vervet monkeys to be a type-B species, as is observed. This kind of analysis can be applied to many kinds of innate and learned behaviour, in many species.

3.4. OPTIMALITY, SIMPLICITY AND MODULARITY

We have seen how, if the behaviour of a species is to be innately well-adapted to its habitat, information about the habitat must be encoded as GIP in the design of the brain. The more information is encoded, the better adapted a species can be; yet the speed limit provides a restrictive bound on the rate at which such GIP can be encoded in the brain by natural selection.

Habitats are continually changing, and species continually presenting new challenges to one another in evolutionary arms races. Therefore "perfect adaptation to the habitat" is a perpetually moving target, and evolution can only chase this target at limited speed. The speed limit gives us a good reason to reject any over-simplified optimality arguments, for any facet of the phenotype, including cognition. The optimum is never precisely attained because it always changes. The speed limit gives some indications as to how close to any optimum evolution can actually come.

Some aspects of the environment, which are important for the design of brains, do not change at all with time. These are the laws of physics and mathematics that underly many facets of cognition: the laws of illumination, perspective and optical flow, which can be exploited in vision, the laws of motion which can be used to predict the movements of objects, the laws of two-dimensional geometry, which can be used for navigation. These constraints are not moving targets for evolution; even within the speed limit, we might expect brains to have evolved to exploit these constraints very precisely, if it gives any advantage in fitness (as it undoubtedly does). For these, we might expect very near-optimality.

What about those aspects of the habitat which are perpetually changing? Many biologists and cognitive scientists believe, along with Crick (1994) and Monod (1971) that evolution proceeds by tinkering, modifying some existing structure to new purposes, in ways that suit the needs of the evolutionary moment rather than any grand plan. So for many (e.g. Sejnowski & Churchland, 1991), the brain is seen as a patchwork of neural nets, each evolved for some original purpose and then possibly retro-fitted to others.

If this is the whole picture, we would expect many aspects of the environment to be encoded in the design of the brain redundantly; the same environmental features giving rise to several different brain structures or facets of brain structures, controlled by different genetic parameters that evolved at different times.

However, as features of the environment change over time, a simple non-redundant design of brain would be able to evolve faster, to track these changes more closely than a complex, redundant design; so it would be fitter. In the long term, this gives a selection pressure towards simpler, less redundant designs of brain.

The same argument also leads us to expect a kind of modularity of the brain (Fodor, 1983). If one aspect of the environment impacts many different behaviours

(and so must be reflected in brain structures for the control of the behaviours) then that aspect can either be embodied redundantly, in separate behavioral systems, or non-redundantly in one module of the brain which is used by the different systems. The non-redundant design needs less GIP to encode changing aspects of the environment; so it can evolve faster when they change, and so is expected to predominate.

Higher animals have many distinct behavioral systems—for feeding, reproduction, predation and so on. These systems all use common sensory and motor facilities of the brain. At this level, the existence of common modules of the brain is confirmed.

So the general implications of the speed limit for the design of brains are quite subtle. We do not expect simple optimality in brain design, yet in certain respects (exploiting the laws of physics and mathematics) brains may be near-optimal; and in spite of the piecemeal, *ad hoc* progress of evolution, simple, non-redundant and modular designs of brain are favoured in the long run. This encourages us to look for simplicity in the 100 Kbytes or so of the design of the brain.

4. Other Implications

This section raises some questions that follow from the speed limit and suggests tentative answers. Some of the ideas are more speculative than those in other sections.

4.1. ACTUAL SPEEDS OF EVOLUTION

How close do actual speeds of evolution come to the speed limit? There are two main possibilities:

- *either* species have now evolved in such a way as to come very close to the practical limit (because it increases their fitness to do so, to respond rapidly to environmental changes)
- *or* because of other limitations, most species are very far below the limit.

The answers, of course, may vary greatly across species and across traits. We can only know them reliably by observation. meanwhile, we might conjecture, as a working hypothesis, that:

 (i) If some feature of the environment fluctuates, and has been fluctuating for a very long time, we might expect species to have developed evolutionary mechanisms to rapidly track changes of the feature, so the rate of evolution of the phenotypic response is now close to the limit [eqn (19)].

(ii) For some completely new environmental challenge, which a species has never faced before, we might expect the rate of evolutionary response to be initially very slow (far below the limit) until the first effective response evolves by chance.

For example, species often need to evolve to change their overall body size, in order to move into new habitats, cope with climatic changes and so on. So we might expect the genetic mechanisms that control body size to be capable of rapid evolutionary response, changing body size as fast as the limit [eqn (19)] will allow. Experience of breeding domestic animals shows that body size can be rapidly controlled by selection.

4.2. EVOLVING THE ABILITY TO EVOLVE

The genetic code specifies the phenotype, just as a computer program specifies a computation. This analogy can suggest some ideas about the constraints on the speed of evolution.

Early computer programs were written in a machine code, which defines directly what the hardware does. One can program complex functions in machine code, but soon hits a complexity barrier; the programs become very long, and small errors are fatal. They are very hard to modify, to alter their function.

The answer was to develop high-level languages, which indirectly (through a compiler) create a sequence of low-level machine instructions; these can express complex functionality much more concisely. Complex programs can be written without intolerable levels of errors, and can be easily modified to do new functions. The history of programming has largely been the history of improving high-level languages.

If we make a random change (or mutation) in a high-level language program, there is a chance that the altered program will still perform some recognizable, perhaps useful, function. For a random change in a low-level language program, the chance of this is much smaller. So a high-level program can evolve much faster than a low-level program.

Similar ideas may apply to the genetic code. If the only genes were those which directly specify the proteins used for metabolism and structure, and there were not other genes which control the expression of the former, then the functionality which can be expressed might be very limited, and the amount of genetic code needed to express it might be very large. This would be the equivalent of machine code in the genotype.

However, as there are genes whose purpose is to control the expression of other genes, these can be in effect a high-level genetic language, allowing

function to be expressed much more concisely. The idea of genes controlling other genes is well known and is experimentally confirmed. It is clear that they are needed for many functions of eukaryotic cells: for instance, to allow differentiation into different cell types from the same genotype during development.

These high-level genes may also have evolved for another purpose: not just to make new functions possible, but also to make rapid evolution possible. Possibly, the design of high-level genes is such that their mutations and crossing tend to have useful consequences in some not-absurdly-low proportion of cases, so that evolution can proceed at a speed not far below the speed limit.

This second-order evolution (of the ability to evolve fast) might seem contrary to our intuition that evolution is a short-term, local optimization process. Surely any evolutionary change must improve fitness immediately; that criterion, rather than an improved ability to evolve later, will decide between strains. "Ability to evolve faster" does improve fitness, but in a way which only becomes evident over many generations, as environments change and competitors fail to respond. There is selection pressure to increase it; but this pressure acts very slowly, and could be swamped by more short-term pressures. Only changes which improve both short-term fitness and long-term ability to evolve will improve the later speed of evolution.

So the evolution of the ability to evolve fast must itself be very slow. If it has taken place, when did it? Life began over 3.5 billion years ago, and eukaryotic cells, which then led to multicellular life forms, did not appear for almost another 2 billion years. Thus there were (of the order of) 10^{13} generations of prokaryotic life before the complexity barrier to eukaryotic cells was breached.

I conjecture that it took this long to evolve a genetic high-level language which not only allows the expression of complex functionality—to build eukaryotic cells—but also allows "rapid" evolution, close to the speed limit. Earlier codes might have evolved which allowed the expression of complex function, but which trapped their owners in the slow lane of evolution; later to be overtaken by a genetic code which could function and evolve fast at the same time. Because this problem was so hard, it took 10^{13} generations—much more than have been available for evolution since then.

The legacy of those 10^{13} generations is an agile genetic code which allows rapid evolution of many traits, such as the body structures and cognition of higher animals, and supports the constant evolutionary arms races and Red Queen races which have taken place ever since.

4.3. PUNCTUATED EQUILIBRIUM AND SPECIES SELECTION

The theory of punctuated equilibrium (Eldridge & Gould, 1972) holds that evolution does not proceed by a steady accumulation of design information in the phenotype (i.e. of GIP), but consists of long periods of stasis, punctuated by comparatively short periods of change. It is motivated by the fact that the fossil record appears not to support a picture of steady change, but shows long periods of stasis in anatomical forms.

Two ideas are sometimes adduced in support of punctuated equilibrium:

(a) *Change in small populations*: It is proposed that evolution proceeds faster in a small population, because a favourable mutation can spread faster over the whole population; so evolutionary changes usually originate from small populations, particularly during speciation events.

(b) *Species selection*: The real units of natural selection are not individuals, but species (Stanley, 1975); evolution proceeds by whole species competing and dying.

The speed limit can give some insight into both these ideas.

Small populations: The same speed limit applies to large and small populations; given the same selection pressure, a large population can respond just as fast as a small one. In potential speed of evolution, there is nothing to favour small populations. However, other effects may come into play:

- A small population, living in a particular isolated habitat, is more likely to come under some special localized selection pressure (which leads to rapid evolutionary response) than a large one.
- Random genetic changes, when they occur, can become fixed in a small population more rapidly than in a large one.
- Speciation, which may precede a period of rapid evolutionary change, may be more likely in small isolated populations.

Species selection: a process of species selection is shown in Fig. 2. The circles represent speciation events, and the lines between represent the continuation of species. The cross-hatching represents the births and deaths of individuals within a species.

A hypothesis of pure species selection would hold that:

- No evolutionary change takes place along the lines which represent continuation of a species.
- When species A splits into species A1 and A2, the genotypes of A1 and A2 differ in a random manner.
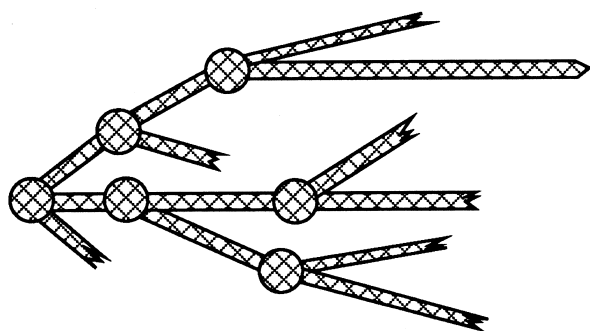
FIG. 2. Illustration of species selection.

● Evolutionary change occurs because only a few species survive, out of the many that are formed.

If there is such a pure species selection process, then the speed limit of this paper applies to it, just as to individual selection; the total increase in GIP per generation from species selection is not more than† $\log_2 \gamma$. But now one "generation" is the interval between the origin of a species and the origin of its "offspring" species, and γ is the number of offspring species from one species.

For instance, in Fig. 2, a species gives rise to eight daughter species, seven of which fail; so the new environmental information conveyed to the genotype is not more than $\log_2(8) = 3$ bits. Putting in a typical figure for the "species generation time" of 10^5 individual generations, even with $\gamma = 100$ (many failed "offspring" species from one parent species), we find that species selection is a very slow process—about 10^5 times slower than individual selection. Species selection can only provide a GIP 10^5 times smaller than individual selection—perhaps a few tens of bits per million years, compared to the megabits of total GIP. At this speed, it is highly unlikely that pure species selection has been an important force in recent evolution.

Whatever the merits of species selection as a descriptive framework for comparing the evolution of generalist and narrow-niche species (Eldridge, 1987; Stanley, 1991), it cannot (in pure form) give a causal account of evolutionary change; it is not capable

of accumulating enough information. Rejecting pure species selection for this role, we come back to individual selection as the force which drives evolutionary change; punctuated equilibrium holds that this occurs predominantly during speciation events. Does the evidence from the fossil record bear on this issue?

The anatomical changes seen in the fossil record are not information-intensive. These changes (typically of body size and shape) might be brought about with comparatively small changes in GIP; say tens or hundreds of bits.‡ Such small changes may well have occurred over intervals so short that we cannot resolve them in the fossil record. Equally, the long periods over which we see no anatomical change may have been the time when the bulk of useful genetic information—not visible in anatomy—accumulated.

The speed limit is inconsistent with some versions on the punctuated equilibrium theory (such as pure species selection) but is compatible with others.

4.4. GENETIC ALGORITHMS

The discussion of mutations in computer programs in Section 4.2 is not hypothetical—mutations and crossing of data structures analogous to genes are regularly used to search very large spaces (for instance, spaces of algorithms or physical designs) by the computing technique of genetic algorithms (Holland, 1975), which mimics natural selection.

Genetic algorithms can solve difficult search and optimization problems which may resist other techniques, and are now in use in industry. It has been argued (Holland, 1975) that, by simultaneously probing many vector subspaces of the total search space, genetic algorithms achieve extra efficiency compared with more conventional search techniques.

However much better they are than other techniques, genetic algorithms are still subject to the speed limit derived in this paper. If the proportion of genotypes which survive each generation is $1/\gamma$, then the total gain in information about the solution per generation cannot, in the long term, be greater than $2 \log_2 \gamma$ bits. Knowledge of this constraint may help to design better genetic algorithms, and to understand their performance.

Genetic algorithms may also serve as a useful theoretical laboratory to investigate the issues of speed of evolution raised in this paper. They can be used in simple "artificial life" environments to investigate what conditions can give rates of evolution close to the speed limit.

---

† Since species "reproduction" is usually asexual—a new species has only one parent species—the speed limit is lower by a factor of 2.

‡ The gross anatomical differences between, say, a horse and a monkey, are largely defined by differences in the length of around 100 bones; if each bone length is defined to a precision of 6 bits (1 in 64) this requires only 600 bits of GIP. Differences in behaviour, as discussed in Section 3, require much more GIP than differences in shape.

5. Conclusions

In the extensive discussions of rates of evolution, rates have generally been defined in terms of metric quantities (such as limb sizes and cranial capacities) which are measurable from the fossil record. Unfortunately there is little theoretical basis for understanding these rates; they give us no common scale on which to define fast or slow evolution.

In this paper I have proposed a measure of the rate of evolution not in terms of metric quantities, but in terms of information—the amount of genetic design information which is expressed in some aspect of the phenotype. The measure of this design information is the genetic information in the phenotype (GIP), measured on a common scale (bits or bytes) for all facets of the phenotype. The total GIP of a mammalian species is probably of the order of megabytes, much less than the 1000 megabytes of raw information in the human genotype.

While there may be megabytes of GIP in the phenotype, the rate of increase in GIP through evolution is severely limited by the speed limit derived in this paper. The total increase in GIP cannot be more than a few bits per generation, and the increase in GIP for some facet of the phenotype is bounded by the selection pressure on that facet—typically to below a fraction of a bit per generation.

The speed limit can be qualitatively understood as follows: it is not possible to refine a population, to have a more narrowly defined range of values for some trait, without imposing a selection load on the population. The rate of refinement is limited by the selection load.

For physical aspects of the phenotype, such as body forms and sizes, it is hard to relate the GIP measure to observations; but for cognition we can do so. This is because both cognition and GIP are about information. Any form of cognition or innate behaviour requires some amount of innate information, which can be estimated and which must exist as GIP in the brain. Therefore we can place lower bounds on the number of generations needed to evolve these forms of cognition.

Two specific consequences of the speed limit were discussed. The first concerns the evolution of human intelligence and language: I showed that the human brain differs from the chimpanzee brain by at most 5 kilobytes of design information, out of a 100 kilobyte total. So it must differ mainly in capacity and power, rather than in design complexity. This makes the evolution of a *de novo* language acquisition device in mankind seem highly unlikely. The second example concerns the evolution of innate fear reactions in vervet monkeys; we can use the speed limit to understand which aspects are innate and which are learned. I believe many other specific conclusions like these can be drawn, concerning many aspects of cognition.

We can also draw general conclusions about the form of cognition in higher animals; we expect it to be near-optimal in some respects, non-optimal in others, and we expect to find economy and modularity in brain design. Thus the evolutionary speed limit provides powerful new constraints on theories of cognition.

Finally, the speed limit has other consequences for other aspects of evolution. It implies that pure species selection has not been an important force in recent evolution.

REFERENCES

CHENEY, D. L. & SEYFARTH, R. M. (1990). *How Monkeys See the World*. Chicago: University of Chicago Press.
CHOMSKY, N. (1975). *Reflections on Language*. New York: Random House.
CRICK, F. (1994). *The Astonishing Hypothesis*. London: Simon & Schuster.
CROW, J. F. (1958). Some possibilities for measuring selection intensities in man. *Hum. Biol*. **30,** 1–13.
ELDRIDGE, N. & GOULD, S. J. (1972). Punctuated equilibria: an alternative to phyletic gradualism. In: *Models in Paleobiology* (Schopf, T. J. M., ed.). San Francisco: Freeman Cooper and Co.
ELDRIDGE, N. (1986). *Timeframes*. London: Heinemann.
FODOR, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.
HOLLAND, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.
KIMURA, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
MAYNARD SMITH, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.
MAYNARD SMITH, J. & PRICE, G. R. (1973). The logic of animal conflict, *Nature, Lond*. **246,** 15–18.
MARR, D. (1982). *Vision*. San Francisco: W. H. Freeman.
MONOD, J. (1971). *Chance and Necessity*. New York: Alfred A. Knopf.
PINKER, S. & BLOOM, P. (1990). Natural language and natural selection. *Behav. Brain Sci*. **13,** 707–784.
SEJNOWSKI, T. & CHURCHLAND, P. S. (1991). Brain and cognition. In: *Foundations of Cognitive Science* (Posner, M. I., ed.). Cambridge, MA: MIT Press.
STANLEY, S. M. (1975). A theory of evolution above the species level. *Proc. natn. Acad. Sci. U.S.A*. **72,** 646.
STANLEY, S. M. (1991). The species as a unit of large-scale evolution. In: *New Perspectives on Evolution* (Warren, L. & Koprowski, H., eds). New York: Wiley.