# Testing evolution

## Lee M Spetner

Redoxia Israel, Ltd. 27 Hakablan St., Jerusalem, Israel

**Abstract:** Although the problem of the probability of the evolutionary process is a possible test of evolutionary theory it has not been adequately addressed. A recent attempt was made to resolve it by estimating the time for evolution, and it concluded that there was plenty of time. This would have been a very significant result had it been correct. It turns out, however, that the assumptions made in formulating the model of evolution were faulty and the conclusion of that attempt is therefore unsubstantiated. The issue remains unresolved.

A major complaint often made against the theory of evolution is that it is not testable. The standard neo-Darwinian theory accounts for evolution as the result of long sequences of random mutations each filtered by natural selection. The mutations that drive evolution are assumed to be random. The random nature of this basic mechanism makes evolutionary events random. The theory can therefore, in principle, be tested by calculating the probabilities of those events.

Several studies have been made tangential to the probability of evolution, but none have actually dealt with the probability of the evolutionary process itself. Sewall Wright [1942] in his statistical study of genetics and evolution dealt only with the statistics of the genetic composition of populations and did not deal at all with the probability of evolution occurring in the first place.  Motoo Kimura [1969] studied the rate of evolution, but his results do not relate to evolution in the usual sense. The evolution he considered was the changes in the genome without regard to their phenotypic effects. He concluded that most genomic changes resulted from random drift and were not determined by natural selection. This type of evolution is not the kind that could be expected to build biological complexity or lead to common descent. He did not address the problem of the probability of what is commonly considered the evolutionary process.

Russell Lande [1976] suggested mathematical models for describing evolution and in particular for distinguishing between natural selection and random drift. Here, again, he did not relate to the probability of the evolutionary process. He did not even mention the probability of getting a beneficial mutation. Instead he made the tacit assumption that the beneficial mutations were already in the population. Barton and Whitlock [1997], in a more up-to-date analysis also dealt only with the issue of gene frequencies in a population and did not consider the probability of the evolutionary process. H. Allen Orr [2005] was concerned with parallel evolution. He did not relate at all to the probability of evolution as such. He considered only the case in which the appropriate mutations had already occurred and on that basis calculated the probability of the parallel evolution of two identical populations.

There is good reason why the probability of the evolutionary process has not been addressed, and that is because its parameters are to a large extent unknown and difficult to estimate with confidence. If one could show that the evolutionary process is sufficiently probable, the charge of not being testable would fall away.

A recent attempt, however, was made to address this very issue, and had it been successful, it would have been an important support for evolutionary theory. Wilf &

Ewens [2011] (W&E) recently attempted to calculate the waiting time for a significant portion of the evolutionary process. They found that the waiting time for the evolutionary process is proportional to the logarithm of the number of beneficial mutations. Because the logarithm is a weak function of the number of mutations, their result would have the advantage of almost bypassing the necessity of knowing the evolutionary parameters with any great precision. Unfortunately their attempt was unsuccessful. Their model of the evolutionary process omitted important features of evolution invalidating their conclusions. They considered a genome consisting of $L$ loci (genes), and an evolutionary process in which each allele at these loci would eventually mutate so that the final genome would be of a more "superior" or "advanced" type. They let $K^{-1}$ be the fraction of potential alleles at each gene locus that would contribute to the "superior" genome. They modeled the evolutionary process as a random guessing of the letters of a word. The word has $L$ letters in an alphabet of $K$ letters. In each round of guessing, each letter can be changed and could be converted to a "superior" letter with probability $K^{-1}$.

At the outset they stated the two goals of their study, neither of which they achieved. Their first goal was to "to indicate why an evolutionary model often used to 'discredit' Darwin, leading to the 'not enough time' claim, is inappropriate." Their second goal was "to find the mathematical properties of a more appropriate model." They described what they called the "inappropriate model" as follows:

> "The paradigm used in the incorrect argument is often formalized as follows: Suppose that we are trying to find a specific unknown word of $L$ letters, each of the letters having been chosen from an alphabet of $K$ letters. We want to find the word by means of a sequence of rounds of guessing letters. A single round consists in guessing all of the letters of the word by choosing, for each letter, a randomly chosen letter from the alphabet. If the correct word is not found, a new sequence is guessed, and the procedure is continued until the correct sequence is found. Under this paradigm the mean number of rounds of guessing until the correct sequence is found is indeed $K^L$."

They gave no reference for such a model and, to my knowledge, no responsible person has ever proposed such a model for the evolutionary process to "discredit" Darwin. Such a model had indeed been suggested by many, not for the evolutionary process, but for abiogenesis (e.g., [Hoyle & Wickramasinghe 1981]) where it is indeed appropriate. Their first goal was not achieved.

They then described their own model, which they called "a more appropriate model." On the basis of their model, they concluded that the mean time for evolution increases as $K \log L$, in contrast to $K^L$ of the "inappropriate" model. They called the first model "serial" and said that their "more correct" model of evolution was "parallel". Their characterization of "serial" and "parallel" for the above two models is mistaken. Evolution is a serial process, not a parallel one, and their model of the first, or "inappropriate", process is better characterized as "simultaneous" than "serial" because the choosing of the sequence (either nucleotides or amino acids) is simultaneous. What they called their "more appropriate" model is the following:

> "After guessing each of the letters, we are told which (if any) of the guessed letters are correct, and then those letters are retained. The second round of guessing is applied only for the incorrect letters that remain after this first round, and so forth. This procedure mimics the 'in parallel' evolutionary process."

W&E were mistaken in thinking the evolutionary process to be an in-parallel one — it is an in-series one. A rare adaptive mutation may occur in one locus of the genome of a gamete of some individual, will become manifest in the genome of a single individual of the next generation, and will be heritable to future generations. If this mutation grants the individual an advantage leading to it having more progeny than its nonmutated contemporaries, the new genome's representation in the population will tend to increase exponentially and eventually it may take over the population.

Let $p$ be the probability that in a particular generation, (1) an adaptive mutation will occur in some individual in the population, and (2) the mutated genome will eventually take over the population. If both these should happen, then we could say that one evolutionary step has occurred. The mean number of generations (waiting time) for the appearance of such a mutation and its subsequent population takeover is $1/p$. [*] After the successful adaptive mutation has taken over the population, the appearance of another adaptive mutation can start another step.

In $L$ steps of this kind, $L$ new alleles will be incorporated into the mean genome of the population. These steps occur in series and the mean waiting time for $L$ such steps is just $L$ times the waiting time for one of them, or $L/p$. Thus the number of generations needed to modify $L$ alleles is linear in $L$ and not logarithmic as concluded from the flawed analysis of W&E.

The flaws in the analysis of W&E lie in the faulty assumptions on which their model is based. The "word" that is the target of the guessing game is meant to play the role of the set of genes in the genome and the "letters" are meant to play the role of the genes. A round of guessing represents a generation. Guessing a correct letter represents the occurrence of a potentially adaptive mutation in a particular gene in some individual in the population. There are $K$ letters in their alphabet, so that the probability of guessing the correct letter is $K^{-1}$. They wrote that

$$1 - (1 - 1/K)^r$$

is the probability that the first letter of the word will be correctly guessed in no more than $r$ rounds of guessing. It is also, of course, the probability that any other specific letter would be guessed. Then they wrote that

$$[1 - (1 - 1/K)^r]^L$$

is the probability that all $L$ letters will be guessed in no more than $r$ rounds. The event whose probability is the first of the above two expressions is the occurrence in $r$ rounds of at least one correct guess of a letter. This corresponds to the appearance of an adaptive mutation in some individual in the population. That of the second expression is the occurrence of $L$ of them. From these probability expressions we see that according to W&E each round of guessing yields as many correct letters as are lucky enough to be guessed. The correct guesses in a round remain thereafter unchanged, and guessing proceeds in successive rounds only on the remaining letters.

---

[*] I am ignoring the generations needed for a successful adaptive mutation to take over the population. These generations must be added to the waiting time for a successful adaptive mutation to occur.

Their model does not mimic natural selection at all. In one generation, according to the model, some number of potentially adaptive mutations may occur, each most likely in a different individual. W&E postulate that these mutations remain in the population and are not changed. Contrary to their intention, this event is not yet evolution, because the mutations have occurred only in single individuals and have not become characteristic of the population. Moreover, W&E have ignored the important fact that a single mutation, even if it has a large selection coefficient, has a high probability of disappearing through random effects [Fisher 1958]. They allow further mutations only in those loci that have not mutated into the "superior" form. It is not clear if they intended that mutations be forbidden in those mutated loci only in those individuals that have the mutation or in other individuals as well. They have ignored the fact that evolution does not occur until an adaptive mutation has taken over the population and thereby becomes a characteristic of the population. Their letter-guessing game is more a parody of the evolutionary process than a model of it. They have not achieved their second goal either.

The probability test of evolutionary theory still has to be made. It can be made only when we are able to calculate, for example, the probability of a known example of evolution of for which data are available. Until that time, the probabilistic aspect of evolutionary theory remains untested.

**References**

Barton, N. & Whitlock, M. C. (1997) The evolution of metapopulations. In *Metapopulation dynamics: ecology, genetics and evolution* (ed. I. Hanski & M. Gilpin), pp. 183-210. San Diego CA

Fisher, R. A. (1958) *The Genetical Theory of Natural Selection*, Oxford. Second revised edition, New York: Dover. [First published in 1929]

Hoyle, F. and N. C. Wickramasinghe (1981) *Evolution from Space*, London: Dent.

Kimura, M. (1969) The rate of molecular evolution considered from the standpoint of population genetics. *Proceedings National Academy of Sciences USA* **63**: 1181-1188

Lande, R. (1976) Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30 (2) 314-334**

Orr, H. A. (2005) The probability of parallel evolution. *Evolution* **59**: 216-220.

Wilf, H. S. & Ewens, W. J. (2010) There's plenty of time for evolution. *Proceedings National Academy of Sciences USA* **107 (52)**: 22454-22456.

Wright, S. (1942). "Statistical genetics and evolution," *Bulletin American Mathematical Society*, vol. 48, pp. 223-246.