

Derivation of the relationship between neutral mutation and fixation solely from the definition of selective neutrality

Jun-ichi Tomizawa*

National Institute of Genetics, Mishima, Shizuoka-ken 411-8540, Japan

Contributed by Jun-ichi Tomizawa, March 27, 2000

A mutation whose fixation is independent of natural selection is termed a neutral mutation. Therefore selective neutrality of a mutation can be defined by independence of its fixation from natural selection. By the population genetic approach, Kimura [Kimura, M. (1962) *Genetics* 47, 713–719] predicted that the probability of fixation of a neutral mutation (u) is equal to the frequency of the new allele at the start (p). The approach traced the temporal sequence of the fixation process, and the prediction was obtained by assuming the selective equality of neutral mutant and wild-type alleles during the fixation process. The prediction, however, has not been verified by observation. In the present study, I search for the mathematical equation that represents the definition of selective neutrality. Because the definition concerns only mutation and fixation, an ideal approach should deal only with these and not with the intervening process of fixation. The approach begins by analysis of the state of fixation of a neutral mutation, and its relation with the initial state is deduced logically from the definition. The approach shows that the equality of the alleles during the fixation process is equivalent to the equality of probability of their ultimate fixation in a steady state. Both are manifestations of the definition of selective neutrality. Then, solely from this dual nature of the definition, the equality between u and p is derived directly. Therefore, the definition of selective neutrality can be represented by the equation $u = p$.

The relationship between the frequency of mutation and the probability of formation of a mutant population has been the subject of intensive study in population genetics. In the long process of evolution, a mutant allele that exists in a finite population either propagates to form a homogeneous population of the mutant allele (fixation) or is lost from the population. Formation of a mutant population was thought to be the result of natural selection of an advantageous mutation. Kimura (1) was the first to propose that mutations could be fixed independently of natural selection. By assuming the random sampling of gametes for mating and propagation, he concluded that a mutation could be fixed by random fluctuation of gene frequencies in a natural population (random drift mechanism) even without natural selection. By extending the theoretical work, he predicted that the probability of fixation of a neutral mutation was equivalent to the frequency of the new neutral allele at the start (2).

The logical analysis of the process of fixation of a neutral mutation must be based on the definition of selective neutrality (“neutrality” hereafter). Because the definition concerns the relationship between mutation at the initial stage and fixation at the final stage, a system that links them directly is the most desirable for logical analysis. Such a system is particularly desirable, because the initial stage and the final stage can be defined, whereas the intervening process is largely unknown and could be unique for each evolutionary system. By studying a logical system that deals only with mutation and fixation, the relationship between the frequency of neutral mutation and the probability of fixation will be derived by logical deduction solely from the definition of neutrality. The

Hypothetico-deductive approach

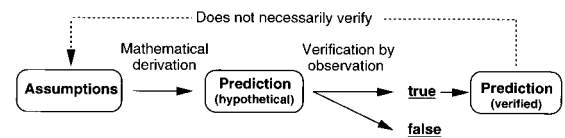


Fig. 1. The logical sequence of a hypothetico-deductive approach. Different natures of the predictions are indicated in parentheses. A hypothetical prediction is subjected to verification usually by observation. An appropriate way of verification is not always available, in which case the prediction remains hypothetical. Verification of the prediction does not necessarily mean verification of the assumptions and the mathematical derivation.

characteristics and achievements of the original hypothetico-deductive approach are described first to help understand the approach proposed herein, which involves a time-indifferent logic.

The Hypothetico-deductive Approach. A hypothetico-deductive approach (“deductive approach” hereafter, where appropriate) starts by proposing a mathematical relationship based on assumptions about the system and then derives a prediction in the form of an equation (Fig. 1). Kimura (2, 3) studied the process of fixation of mutation of a sexually reproducing diploid organism, assuming that a population of mutant individuals is derived from a single line of mutants originating from an ancestor. Random mating and nonoverlapping generations were also assumed. Taking account of random sampling of gametes for mating and of individuals for reproduction in a finite population, Kimura analyzed the stochastic process mathematically.

Fixation of a mutation was analyzed after the temporal sequence of the fixation process. Considering an idealized population of which the size N (the effective number of breeding individuals) is constant, he deduced the relationship between mutation and fixation. The relationship was predicted in which the new allele was under constant selective advantage (s) over the preexisting alleles, and Ns was much smaller than 1. Then, the following equation was obtained:

$$u = p + 2Nsp(1 - p), \quad [1]$$

where u is the probability of fixation of a mutant allele, defined as the value in the steady state approached at infinite time (2), and p is the frequency of the new mutant allele at the start. When s approaches 0, the condition in which the mutant and wild-type

*To whom reprint requests should be addressed. E-mail tomizawa@lab.nig.ac.jp.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

alleles are equivalent in terms of natural selection during the fixation process, he obtained

$$u = p. \quad [2]$$

When both sides of Eq. 2 are divided by the same time unit (physical time, generation time, or other constant time unit), the following equation for the rate of fixation (k) per unit of time and the corresponding rate of mutation (v) are obtained, namely,

$$k = v. \quad [3]$$

The same relationship was predicted for haploid or polyploid organisms, for organisms that propagated asexually (2), or for situations in which the population size fluctuated (4, 5).

Because a prediction made by a hypotheticodeductive approach is based on assumptions used in deriving the mathematical equation, such a prediction must be verified (Fig. 1). Although efforts to verify the prediction by comparing it with observations have not been successful, failure does not necessarily discredit the prediction. Problems of verification are discussed below.

In this hypotheticodeductive approach, neutrality is defined in terms of the assumed selective equality of alleles during the fixation process. However, this assumption, which is based on the equal behaviors of the mutant and wild-type alleles, is not exactly the same as the authentic definition of neutrality, which concerns only the initial and the final states of fixation. Therefore, the significance of the prediction based on the assumption needs to be examined by an approach based on the definition.

The Time-Indifferent Approach. In contrast to the hypotheticodeductive approach whose logic follows the temporal sequence of the fixation process, the present approach begins by examination of the state of fixation of a neutral mutation, and its relationship to the initial state of mutation is analyzed. Because the fixation of a mutation contains a concept of time independence, its existence must imply the presence of a steady state. Therefore, an event that would disturb formation of a steady state, such as further mutation at the site of interest, is ruled out. This approach that concerns only mutation and fixation, but not the intervening process, should be ideal for the present study, which aims to understand the relationship between the frequency of neutral mutation and the probability of fixation. This approach is called the time-indifferent approach, because it ignores the temporal sequence of the fixation process.

Neutrality of mutation is defined by independence of its fixation from natural selection. Therefore, the probabilities of fixation of a neutral mutant and a wild-type allele should be equal. Consequently, in the steady state, each allele, whether mutant or wild-type, must be equal in its potential for fixation (arrow a in Fig. 2Aa). Then, from the axiom that equivalent causes produce equally probable effects, each allele has an equal probability of fixation (arrow b in Fig. 2Aa). One can then conclude that the concepts of equal probability of ultimate fixation and the equality of alleles during the fixation process are equivalent in the steady state, as shown schematically in Fig. 2Ab. Both are manifestations of the authentic definition of neutrality.

Solely from the dual nature of the definition, the equality between the probability of fixation and the frequency of neutral mutation, $u = p$, can be derived directly (Fig. 2Ab). Conversely, the authentic definition can be derived from the above relationship in a steady state. Therefore, the relationship is equivalent to the definition and is applicable to neutral mutation fixed by any mechanism in any population. The equality of the potentials of both mutant and wild-type alleles during the fixation process is the assumption used in the hypotheticodeductive approach. The

present approach justifies the use of this assumption as an alternative form of the definition.

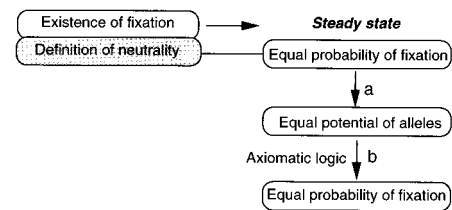
In this approach, the frequency of a neutral mutation is considered to be equivalent to the probability of the neutral mutation. However, because the probability is not obtainable by a determination of the frequency of a single actual distribution of mutants, the above statement is logically compromised. The hypotheticodeductive approach described above already uses these terms with similar meanings. However, in a logical analysis, one could imagine evolutionary systems that are equivalent to the system on the earth and consider an ensemble average of corresponding events. In this context, the stochastic concept can be applied to an evolutionary event.

An Alternative Hypotheticodeductive Approach and Its Comparison with the Time-Indifferent Approach. An alternative approach that involves fewer assumptions than the hypotheticodeductive approach has been proposed. In this approach (6, 7), which is based on the original proposal (8), a population of a diploid organism of size N is considered. When the mutation frequency at the site of interest is p , in each generation, $2Np$ alleles are produced. Of the $2N$ alleles in the population, only one allele will be fixed. Then, the probability (u) that the fixed allele is a mutant is calculated to be $2Np(1/2N)$, and therefore $u = p$.

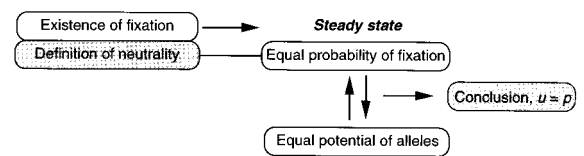
The logical structure of this approach will be compared with that of the time-indifferent approach, because they have been misunderstood frequently to be the same. The above approach

A Time-indifferent approach

(a) Logical sequence



(b) Logical structure



B Alternative hypotheticodeductive approach

(a) Logical sequence

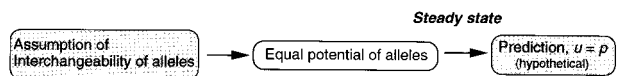


Fig. 2. The logical sequences of the time-indifferent approach and the alternative hypotheticodeductive approach. (Aa) This sequence is presented to facilitate understanding of the logical derivation of the time-indifferent approach described in the text. "Equal probability of fixation" means the equality of the probability of formation of a mutant population and that of a wild-type population for future fixation, and "Equal potential of alleles" means the equality of a mutant allele and of a wild-type allele during the fixation process. An arrow indicates the direction of a logical but not temporal sequence. (Ab) The logical structure of the time-indifferent approach. (Ba) The logical sequence of the alternative hypotheticodeductive approach. This approach assumes interchangeability of alleles in addition to formation of a steady state. This prediction needs verification.

proceeded forward along the process of fixation by assuming interchangeability between the neutral mutant and wild-type alleles during the fixation process. Then, the equality between the probability of fixation of neutral mutation and the frequency of the mutation in the steady state were predicted (Fig. 2*Ba*). Although this approach is essentially hypotheticodeductive after the temporal sequence, the method to obtain the probability is time-independent. Therefore, this approach shares a certain similarity with the time-indifferent approach.

However, the aims of these approaches are very different. The time-indifferent approach is intended to find a mathematical equation that implicitly contains the authentic definition of neutrality. On the other hand, the aim of the above approach is to investigate a relationship between the important parameters when neutral mutations are fixed. Although both approaches yield the same equation, each accomplishes a different task with a different aim.

The difference between the logical processes of these approaches is a direct reflection of the difference of their aims. The time-indifferent approach, whose logical process starts from the state of fixation, shows that the relationship, $u = p$, as well as the interchangeability of alleles are equivalent to the definition of neutrality. In contrast, the above deductive approach traces the temporal sequence and accomplishes its aim by predicting the same relationship. However, the logical meaning of the prediction, namely, its equivalence to the definition, cannot be deduced by the approach. Consequently, the tautological nature of the approach, the equivalence of the assumption and the prediction, was unnoticed until revealed by the time-indifferent approach. Furthermore, because the approach depends on its assumptions, the prediction needs verification. On the other hand, as described below, the time-indifferent approach is free from the problem of tautology, and its conclusion does not need verification. These arguments clearly show that the aim and the logical structure of the above approach are very different from those of the time-indifferent approach.

Significance of the Time-Indifferent Approach. The time-indifferent approach, which has the specific aim of finding an equation that implicitly represents the definition of neutrality, starts from the analysis of the state of fixation of a neutral mutation. In this way, the logical analysis rests on the definition, which is based on the probability of fixation. Formation of a steady state is not an assumption but an obligatory condition of the definition. Therefore, the argument by this approach is free from assumption, and the conclusion does not need to be verified by an independent method. On the other hand, the hypotheticodeductive approaches start their logical processes with assumptions including the equal potential of fixation of alleles or their interchangeability during the fixation process and formation of a steady state. Consequently, all subsequent arguments are hypothetical, and their conclusions need verification.

The time-indifferent approach directly links the frequency of neutral mutation with the probability of fixation through analysis of the definition of neutrality (Fig. 2*Ab*). The approach shows that the concept of the interchangeability of the alleles during the fixation process is equivalent to the definition. Because the relationship of $u = p$ is derived solely from the dual nature of the definition by logic, it must be equivalent to the definition and must cover all cases of fixation of neutral mutations.

The assumption of the equality of alleles during the fixation process used by the hypotheticodeductive approach is given a status equal to the definition of neutrality by the time-indifferent approach. However, because the assumption and the definition are two aspects of the same phenomenon, they might have been considered to be equivalent. Had this fact been realized, the relationship $u = p$ and its equivalence to the authentic definition

would have been known simply by logic without referring to the fixation process or to any mathematical manipulation. In fact, as described above, the equation can be derived solely from the authentic definition without any assumption.

Although the time-indifferent approach gives a conclusion that is generally applicable to all cases of fixation of neutral mutations, each hypotheticodeductive approach to a different evolutionary system is formulated on a different set of assumptions. Because each prediction by such an approach is a statement on a particular case, the generality of the prediction cannot be guaranteed. As a result, the deductive approaches cannot relate their predictions to the definition of neutrality.

In the preceding section, the tautological nature of the alternative hypotheticodeductive approach was described. The same argument is applicable to the approach proposed by Kimura (2, 3). The tautological nature of its logic was not realized, because the equivalence of the assumption and the prediction had not been recognized. The use of the assumption based on the future probability of fixation of alleles in characterizing their behavior in the process of fixation caused the tautology and therefore should have been avoided.

Verification is an essential part of the hypotheticodeductive approach (Fig. 1). Comparison of the prediction with the observed facts is the usual way to examine the applicability of a prediction in nature. For verification of a prediction by observation, universal and extensive agreement between the prediction and observation is required (see, for example, ref. 9). The primary purpose for verification of the prediction from the hypotheticodeductive approach should be to demonstrate that the prediction is valid for neutral mutation. Then, verification by observation should be possible only when all mutations would be neutral; otherwise, the presence of mutations that are not neutral disturbs the observation. In any event, efforts have been made to verify Eq. 3 by observation (ref. 10 and references therein). However, because determination of the stochastic parameters k and v from historical data is not conceivable and because of the difficulty of obtaining reliable values of rate constants of an alternative equation derived from Eq. 3, the prediction has not been verified by observation.

On the other hand, because the relationship of $u = p$ can also be derived without assumption by the time-indifferent approach and shown to be equivalent to the definition of neutrality, the relationship must be applicable to fixation of neutral mutations that occur in nature. Therefore, this prediction by the hypotheticodeductive approach is not an object for verification by observation. In other words, irrespective of observation, the prediction is always logically valid. The verification does not necessarily mean that the set of assumptions used in the deductive approach is the only acceptable one and that Eq. 1 from which the prediction is derived is valid.

The above argument shows that the relationship $u = p$ is applicable to fixation of neutral mutation in nature. Accordingly, the approaches aimed to prove the prediction could serve for determination of the fraction of neutral mutations. In this case, however, there is no particular value to expect, and the requirement for the accuracy of determination is less stringent. Even so, because of the complexity of evolutionary events, a reliable estimate is very difficult to obtain.

Since the relationship of $u = p$ (and $k = v$) was predicted nearly 40 years ago by the hypotheticodeductive approach, the validity of the relationship has been taken for granted without rigorous verification. Nonetheless, the relationship has been considered as fundamental for fixation of neutral mutations, without its equivalence to the definition of neutrality being recognized. The revelation of the nature of the predicted relationship by the time-indifferent approach now sets the field on firmer ground for further understanding of the evolutionary process.

The logical meaning of the equation $u = p$ deduced by the hypotheticodeductive approach is different from that of the same equation obtained by the time-indifferent approach. Because this difference reflects the general characteristics of hypotheticodeductive approaches, the present study gives an opportunity to understand the usefulness and limitation of

hypotheticodeductive approaches, which will continue to be widely applied for analyses of complex natural phenomena.

I thank many friends for useful comments, particularly Dr. Nobuo Shimamoto for reviewing physicomathematical descriptions and Dr. Richard S. Hayward for careful reading of the manuscript. I am also grateful for the encouragement of Dr. James F. Crow.

1. Kimura, M. (1954) *Genetics* **39**, 280–295.
2. Kimura, M. (1962) *Genetics* **47**, 713–719.
3. Kimura, M. (1968) *Nature (London)* **217**, 624–626.
4. Kimura, M. (1987) *J. Mol. Evol.* **26**, 24–33.
5. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, U.K.).
6. Crow, J. F. (1986) *Basic Concepts in Population, Quantitative, and Evolutionary Genetics* (Freeman, New York).
7. Gillespie, J. H. (1998) *Population Genetics, a Concise Guide* (Johns Hopkins Univ. Press, Baltimore).
8. King, J. L. & Jukes, T. H. (1969) *Science* **164**, 788–798.
9. Reichenbach, H. (1951) *The Rise of Scientific Philosophy* (Univ. California Press, Berkeley, CA).
10. Hartl, D. L. & Clark, A. G. (1989) *Principles of Population Genetics* (Sinauer, Sunderland, MA), 2nd Ed.