

Natural Selection and Genetic Drift: Their Probabilistic Characterization

Yosaku Nishiwaki¹

Abstract

The modern theory of heredity is deeply based on the fact that genetic information transmitted from generation to generation supervenes on the chemical material, called DNA sequences. The evolutionary nature of genetic transmission has been studied through the population genetic models, originated by Mendel, and the collection of these models is the theory of population genetics. In these models, natural selection and genetic drift are formalized differently and from this formulation they are thought to be fundamentally different. This is what I want to discuss in this paper. First I will show how the evolutionary theory is combined with probability. And then although natural selection and drift may be thought to be different and certainly have different aspects each other, I think that they are tightly connected. I will show how they are connected in a probabilistic framework. A conclusion of the study shows that if there is selection, then there has to be drift. This does not mean that I do not admit any neutrality at all. To say more precisely, the concept of drift presupposes the concept of selection.

1. Introduction or preliminary remarks

We will briefly discuss general concepts that are necessary for our following arguments. These concepts are symmetry, determinism, causality and explanation. A coherent physical characterization of nature by symmetry is known as follows:

- (1) The universality of the laws of nature is physicalized by the concept of symmetry. If we are asked to confirm the universal proposition where any natural law is expressed, then we cannot confirm it directly with using only logic. But by using symmetry and its mathematical twin, group theory, we can reach a coherent confirmation in a mathematical sense. This is a kind of the inference to the best explanation.
- (2) Epistemological characterization of the physical systems can be done by the introduction of the coordinate system and of the concept of symmetry. This is also a physicalization of the epistemological relationship between the observer and the physical system.²

¹ Department of Philosophy, Keio University, Tokyo.

² If we want to find the origin of the concept of symmetry, we have to go back to the variational principle, which was made for refining the Newtonian mechanics. The first indication of the power of symmetry to describe physical laws was done by Einstein in 1905, who rewrote the laws of physics on the strength of a symmetry principle, *the principle of relativity*. Reference of frames is the description of events in space and time from the point of view of different observers. Einstein's vision of the deep equivalence of different points of view led him to pursue even further the idea that the laws of physics should appear exactly the same to all possible observers. *The principle of equivalence* says that the laws of gravity must be such that the apparent forces due to any possible kind of motion are indistinguishable from gravitational forces. The next step was Noether's theorem (1918). It says that for any symmetry of the laws of physics there is a corresponding conservation law. Though the converse is not part of her original theorem, it is known to be true: Every conservation law must be associated with a corresponding symmetry. It is the Lagrangian formalism that allows to choose coordinates that best reflect the symmetry of the potential, without altering the simple form of the equations. Conservation laws are easily obtained in this formalism.

(1) Spatial symmetry implies the conservation of linear momentum.

(2) Rotational symmetry implies the conservation of angular momentum about the axis of symmetry.

(3) Symmetry under temporal transition implies the conservation of energy.

The next stage was the discovery during the last 35 years that many, and maybe all of the laws of physics seem to be caused by a special kind of field called a gauge field, whose structure and behavior are completely dictated by a new symmetry requirement: the

The existence of asymmetric laws, if any, will terrify the subvenience of the basic physical laws below the phenomena. A good example, only the example in physics, is the 2nd law of thermodynamics.³ But historically we have not noticed this so clearly. For we have used implicitly the causality principle instead of the direction in time. Unfortunately this is only a metaphysical substitute. Even the classical mechanics have not been interpreted naturally without using the causality principle. If there is no such a principle, we cannot even imagine the classical causal world of Newtonian mechanics.⁴ However, the causality concept itself has its own weakness, for we have to define or to characterize the concept of cause and effect independent of the principle.

A cause doesn't have to be a sufficient condition for its effect. Causes often aren't necessary conditions for their effects.⁵ Causal determinism says only that if all the causally relevant facts are set out, these will leave open only one possible future. If all matter is deterministic and if a person's mind is a material thing, then human behavior is physically determined. Whether determinism is true affects how we can use the concept of probability. If determinism is true, probability is a lack of our information or knowledge about our world. If determinism is false, probability describes an objective fact about our world. Suppose that your beliefs and desires were due to genes, environments, and chance. If determinism robs you of freedom, chance seems to rob you of freedom as well. If we assume that our world is asymmetric, then it is possible to have causation without determinism and indeterminism. Causality can exist in an indeterministic universe just as much as it can in a deterministic one. Determinism doesn't rule out the idea that if the past had been different, the present would be different. Determinism doesn't rule out the idea that I can affect what the future will be like by now acting one way rather than another. Fatalism denies this; it says that the future is independent of what you do in the present. Hence, determinism and fatalism are almost contrary claims.

That a mechanical system is symmetric with respect to time and space implies many important results. Time asymmetry shows the existence of the origin of a system. If a system is time symmetric,

requirement of *local symmetry*. The laws show a local symmetry if the equivalence persists even when you choose a different point of view at every point in space and at every possible time.

³ A thermodynamic system has, at any given time, a state characterized by a density $f(x)$, not necessarily independent of time. Given a density f , $\mu(A)$ of the set A in the phase space X is defined by $\mu(A) = \int_A f(x) dx$. We define an observable O to be a real number characterizing some aspect of a thermodynamic system, for example the energy, pressure, or temperature. An observable is a map $O: X \rightarrow R$. The expected value of the observable O is given by: $E(O) = \langle O \rangle = \int X O(x) f(x) dx$. Then the Boltzmann-Gibbs entropy of a density f is given by: $H(f) = - \int X f(x) \log f(x) dx$. It is easily shown that the only observable which is a function of a thermodynamic state that gives the requisite additive property to make the entropy an extensive quantity is the logarithmic function. A typical presupposition of these definitions is Tolman's position as follows. 'For the extensions to statistical mechanics we take the assumptions of equal *a priori* probabilities for classical states, or of equal *a priori* probabilities and random *a priori* phases for quantum mechanical states, as the necessary additional hypotheses.' (Tolman, p.11)

⁴ The relation between causality and symmetry can be reconciled as follows. We define the symmetry group of the cause as the symmetry group of the whole system for cause-equivalence in its phase space. Similarly, the symmetry group of the effect is defined as the symmetry group of the whole system for effect-equivalence. Since cause-equivalence implies effect-equivalence, it follows that every element of the symmetry group of the cause must necessarily also be an element of the symmetry group of the effect. (Rosen)

⁵ Determinism says that a complete description of the causal facts guarantees what will happen next. Indeterminism holds that even a complete description of the present will leave open more than one possible future.

there is no origin of the system and we can determine either its past state or its future state in the same manner when we know its present state. But if the system is time asymmetric, we have to think of its past and future differently. The question why the system begins with the low entropy is significant if the system is time asymmetric. In evolution too, the question of the origin of species is important because of the time asymmetry of evolution. So variation as an initial condition has the crucial role. The chance-like character of the origin is one of the reasons that adaptive explanation is not deterministic. Adaptive explanation is only partially deterministic. We can not eliminate the chance-like existence of variations and so we can not have complete determinism.

Symmetry implies that a system has to be closed. And so if a system is open, it may be asymmetric. If a system is asymmetric, we can think of causality in general and history in particular. Also such a system introduce the probabilistic treatment of the initial conditions of the system. This causality or history implies that mechanistic determinism or fatalism is quite different from usual determinism. This conclusion is enough to set selection and drift in the probabilistic framework. Our idea of treating drift is roughly like this: drift is formalized in which selection works between genotypes of organisms of a certain species and within organisms of a certain genotype selection does not work. Selection is explained causally and discretely. Causality and discreteness necessarily induce the coarse graining in the biological world (and also the information concept). Within a coarse causal graining drift is explained probabilistically.

System's asymmetry also implies that description, explanation and prediction are different each other. Probabilistic concepts in the synthetic theory of evolution do not permit to describe things individually. Here we have the contrast between individual description (explanation, prediction) and population description (explanation, description).

What we can learn from these things is that we should face positively to the asymmetry of a system.

2. Too Contrastive Comparisons

Too contrastive characterization of selection and drift has made things unclear. It has been thought that selection and drift are contrastively different in many respects. This has made the general view that selection and drift are fundamentally different evolutionary processes and that they are accordingly different concepts. Sometimes this distinction is exaggerated as the ideological standing point of Darwinism. One recent example is the controversy between the selectionist view of molecular evolution and the neutralist one.⁶

Before discussing the relation between selection and drift, it may be better to make clear the contrastive aspects of selection and drift. To confirm this strong tendency of the contrastive comparison between selection and drift, I will show the important points of almost commonly consented facts.

2.1 Historical contrasts

(1) Darwin and teleology

⁶ See the note 13 below.

Darwin showed that evolution has no intrinsic purpose and that it is the sequence of the changes caused by the outer forces working on the organisms. Chance means the lack of purpose in nature and in this meaning chance does not make any conflict with the causal explanation of evolution. This was the starting point of causal understanding of the living world by excluding the teleological interpretation of the world. Here we can find the corroboration of chance and causation. But if chance means not only the lack of purpose but also a positive randomness, then the corroboration will easily break. Deterministic causation and indeterministic chance can not corroborate together.

(2) Fisher-Wright controversy

Fisher constructed population genetics by synthesizing Mendelian genetics and the theory of natural selection in the light of statistical mechanics. Wright also systematized population genetics. But their ideas are different. Their difference depends on the main cause of evolution. Fisher thought that natural selection is the only main cause of evolution and that drift is just the noise or the error in the evolutionary processes. On the other hand, Wright thought that drift, particularly in a small population, has the strong effect in the evolutionary change. They both noticed well that probabilistic treatment has to be necessary to think of evolution. So their controversy was neither about probability, nor about determinism or indeterminism. It was about the status of drift in evolution. There is no difference between them how to formalize natural selection. It is formalized probabilistically. But the difference is how to think the drift concept. Is drift just an error or one of the main causes of evolution?

The past controversy appears again in a slightly different way as the new controversy, which many people have discussed after Williams and Dawkins.⁷ This issue is about the unit of selection. We may call this controversy reductionist-pluralist controversy.

2.2 Biological contrasts

(1) Darwin's explanation of adaptation

Darwin thought that adaptation is the result of natural selection. This may be one of the reasons that adaptationist explanation is right and necessary to explain the forms and the behaviors of living organisms. If we suppose that this type of explanations is always required, then we are forced to think that the existence of chance is an obstacle to complete the explanation. As a matter of fact, the explanation as a result of drift seems to be no explanation at all. When someone says that the existence of *A* is the result of chance, no one would accept it as the explanation of *A*'s existence. Does this mean that chance does not have any role in explanation?

Let's imagine a horse race. In this case, we use the effect of chance positively and sometimes bet a lot of money. When your friend says that she won by chance, we do not feel that her explanation was pointless because of using the chance concept. This means that the following proposition is not correct: drift is a result of chance and hence drift can not be used as one of the factors of the evolutionary changes.

Drift may not be a cause of evolutionary processes, but it may explain these processes. For, as we

⁷ We should notice that the controversy reflects the difference between the English version of the evolutionary theory and the American one.

already showed, explanation is not just the causal description of the evolutionary changes.

(2) Deterministic models and stochastic models

Selection models and drift models are differently characterized in the usual population genetic texts. Selection models are deterministic and drift models are stochastic. There artificial selection is the hint to consider natural selection and sampling error is the hint of considering drift.

Man and woman are different and this difference is important in many respects. But we cannot say that man and woman are essentially different. There are as many common points as different points. This is true in the case of selection and drift. Soon we will see the common characters of natural selection and drift.

2.3 Conceptual contrasts

(1) Darwin's characterization of selection

Selection works on the phenotypes of organisms in a population and through the selective force the evolution of the population occurs. The fundamental unit of selection, according to Darwin, is the organisms in a population. This is the direct consequence of Darwin's argument of the existence of natural selection. This may be called a *Darwinian process*. In this characterization, the most important concept is *Darwinian fitness*. The Darwinian process is just one of the usual causal processes. And then it follows that the Darwinian processes are the deterministic processes. Certainly, each selective process as a token is deterministic. However, the average or the distribution of a certain property of the members of a population is the property of the population, not reducible to each member. If selection works on the plural levels of the living world, the reduction of properties of a population into its members and their properties would be almost impossible.

This means that the selection, working on a single level like organism, is presupposed in order to think of selective processes as deterministic processes. If there is no such presupposition, the statement that selection is deterministic is not always clear.

(2) Sampling, biased or unbiased

We sometimes say that selection is a biased sampling, and drift is a sampling error. This characterization shows us that they are conceptually different. In particular, in the case of unbiased sampling, the difference among each result is usually considered to be caused not by selection but by drift. Moreover, the changes caused by a series of drift are not the causally deterministic processes, and hence they are stochastic processes.

By the way, if a sampling is done with the completely deterministic procedures, do we call such procedures sampling? We will not call them sampling. This intuition suggests that if we take out the sampling character from selection, we have to change the meaning of selection. And this implies that selection essentially include probabilistic factors.

What I have said through 2.1- 2.3 may be summarized in the following table.

	No purpose	Probabilistic	causally effective
selection		and ×	
drift			×

3. Hierarchy, supervenience, equivalence class and probability

3.1 Hierarchy

Modern scientific view of our world is hierarchical. Scientific knowledge supports that our world is hierarchical and that each level of the world is described by different scientific theories. Scientific division of labor is based on this hierarchical view. But this scientific division of labor does not prove that our world is hierarchical. If sciences are only instrumental, the hierarchical view is only our device of looking at the world. Here we suppose that our world is hierarchical, and think of the hierarchical world as a paradigm.

One crucial thing in biology is that we have a hierarchical classification of biological entities. Biological entities like species are classified hierarchically. In many cases we explain and justify this classification as the reflection of evolutionary history of these entities. If we assume this fact, then biological classification affords us a lot of concrete examples of supervenience and equivalence class.

3.2 Supervenience

Sometimes physicalism seems to exclude biological properties. Biological properties are not the same as physical properties. People, who think like this, criticize physicalism. And its typical criticism is toward the following physicalist claim.

Biological properties can be explained by explaining physical properties.

This claim entails that biological properties are physical properties. All biological properties are finally explained or described physically. This situation is the same as mental properties.⁸ Many philosophers have thought that biological studies are already reduced to physics. However, we can not find how to reduce biological properties to physical properties. It seems to be a long distance between biological properties and physical ones. One reason why it is not easy to reduce biological properties to physical ones is that physicalism is deeply depends on physics itself, not on any biological things.

To save this situation, we may use the concept of supervenience. Let's think two examples. One is fitness. Fitness represents how biological organisms fit to the environment. This property does not look to be physical. The other is any usual mental property. They look to be non-physical. But at the same time they seem not to exist unless there is an organism or a brain.

Supervenience is defined as follows. If the set of properties of object Q determines the another set of properties of it P and the converse does not hold, P supervenes on Q . If P supervenes on Q , then there is one-many relation between P and Q . This is easily understood if we think of information.

When there are many thermometers in a room, designating the same temperature, they have the same information about the room's temperature. But physical thermometers are all different. In a similar way, fitness supervenes on the properties of an organism and its environment.

To generalize these relations, all the properties in non-physical sciences supervene on the properties in physics. Now what does "all objects are physical objects" mean? The expression "an object is

⁸ There are a lot of discussions about the mental properties. For instance, see Kim.

physical” does not mean that a property of the object is neither physical nor that all properties of the object are physical. An object having a mental property can have a physical property. Fitness consists of physical properties but not physical itself.

Now define “an object is physical” as “as to all properties of the object, if they are not physical, then they supervene on the physical properties of the object.” With this definition, sciences investigate the supervenient properties. Hence, sciences can study non-physical properties. There has been a lot of discussion about this claim. One trouble is that supervenient properties can not be causes. For example, fitness is not any cause. So some say that the property, having no causally effective power, can not have any role in sciences. This is wrong. Causal ineffectiveness is different from explanatory ineffectiveness. Information, fitness, and mental properties are supervenient properties. And they are all causally ineffective, but they can explain many things. Scientific explanation is more than causal description.⁹

3.3 Equivalence classes

Let’s think here that there is no mutation of genes. Then such genes share the same character as the elementary particles in physics. This invariant character does not come from gene’s chemical properties, but comes from its functional role in life. Gene is the functional unit.

The importance of equivalent class is fully understood in physics and it also plays important roles, for instance, in logical and algebraic treatment of formal systems. We can see the same thing in population genetic models. With two alleles A and a , the equivalent class of AA , Aa or aa is just an object of genetic models. The organisms having the same genotype construct an equivalence class. And among such equivalence classes, we can find the Mendelian laws.¹⁰

If we assume the equivalence class like this, the individual organisms having the same genotype should be treated as the same object. This corresponds to the coarse graining in statistical mechanics.¹¹

Such a model cannot tell us the particular initial state of a population. We can only know an equivalence class of initial states having the same genotype. This means that the determination of a

⁹ This characterization of supervenience depends on Sober (1993), pp. 73-77.

¹⁰ For instance, we can define $x \sim y$ as follows.

$x \sim y$ iff x and y have the gene A or a .

Then, for some individual x , $x \sim = \{y \mid y \sim x\}$ is an equivalence class. If we define $x \sim_A = \{y \mid y \sim x \& x \text{ has } A.\}$, $x \sim_a = \{y \mid y \sim x \& x \text{ has } a.\}$, $x \sim_A \cup x \sim_a = x \sim$. This equivalence class is the whole space. If we consider the transmission of generations, $T(A) = A$, $T(a) = a$. As to the transmission, A and a make a group. Combining with the equivalence classes, we say that the set of transmissions of genes in a population makes a symmetric group with respect to the equivalence classes. This is a trivial fact. We can also define differently as follows.

$x \sim y$ iff x and y have AA , Aa or aa . $T(X) = X$ (Here $X = AA$, $X = Aa$ or $X = aa$)

¹¹ Leibniz’s principle of Identity of Indiscernibles also needs the equivalence class. We can formalize the principle as follows. For any property F , $x = y \iff (F)(Fx \iff Fy)$. The axiom of extensionality is formalized as $(F)(Fx \iff Fy) \iff x = y$. Putting them together, we have $x = y \iff (F)(Fx \iff Fy)$. This is also called Leibniz’s principle. For two things x and y , we say they are identical even though they are in different positions or they have different velocities. To save the principle from this kind of differences, we can use equivalence class as follows.

$x \sim y$ iff x and y are identical except their position and velocity.

From this definition, we can construct an equivalence class $x \sim = \{y \mid y \sim x\}$, and define $F(x \sim) = Fy \& y \sim x$. Then we have a modified principle $x \sim = y \iff (F)(F(x \sim) \iff F(y \sim))$.

particular initial state does not have any meaning for Mendelian laws. This is the same as in the case of coin tosses. Each toss has no physical difference. “Coin toss” is not a token, but a type. To say exactly, each coin toss is a realization of a type. This implies something stronger than the ignorance interpretation of probability.

The ignorance interpretation is a subjective interpretation of probability. The degree of ignorance is a probability value. If we classify ignorance *in principle* as a kind of ignorance, then all of the things that we cannot know is included. This seems to be irrational. Looking at the genetic models, we have to assume that the difference of members within an equivalence class is not known in principle. This difference is theoretically unknown.

There is one difference between physical equivalence and genetic equivalence. We use directly the internal properties of organisms in a population to construct the equivalence classes. On the other hand, in physics external properties, like a position, are used to construct equivalence classes. These internal properties are not local symmetries, but they have similar effects. We have to homogenize organisms in order to transform the internal properties of organisms. And this homogenization calls for the principle of equal probability.

3.4 Probability and its interpretation

The role of probability in 20th century is quite different from other centuries. The theory of error changed and became the most important procedure for getting empirical knowledge. But if someone is asked the true meaning of probability, she can not answer well what probability is. Philosophers have been interested in what the mathematical formalization truly means.¹² This is the problem of the interpretation of mathematical probability. Let’s think some of them.¹³

A proposition or a set means an event and the actual frequency of an event in a population of events is one possible interpretation of probability. On each toss, the coin lands either heads or tails. Let H be the proposition that the coin lands on some arbitrarily selected toss. $P(H)$ can be interpreted as the actual frequency of heads in the 100 tosses. Under this interpretation, all of the above axioms are satisfied. The actual frequency interpretation of probability is an objective interpretation. It interprets probability in terms of how often an event actually happens in a population of events.

Subjective interpretation of probability is another interpretation. We can talk about how much certainty or confidence we should have that a given proposition is true. It describes what our degree of belief ought to be. Degree of belief can be interpreted so that it satisfies the Kolmogorov axioms.

Many philosophers believe that science uses a notion of probability that is not captured by either the idea of actual relative frequency or by the subjective interpretation in terms of degree of belief. A third interpretation of probability says that an event’s probability is its hypothetical relative

¹² Let’s summarize the formal characterization of probability. Taking a set of propositions, closed under the logical or set theoretical operations, we can define a measure P . This function maps the set of proposition on the set of real numbers. And if it satisfies the following conditions for any A, B of the set, P is called a probability measure.

0 $P(A) \leq 1$

If A is true or a whole set, $P(A) = 1$

If A and B are incompatible, $P(A \text{ or } B) = P(A) + P(B)$

¹³ The following classification of the interpretations of probability is due to Sober (1993), pp. 60-69.

frequency. A fair coin need not produce exactly half heads and half tails when it is tossed a finite number of times. A probability value does not entail an actual frequency equal to the value, but it does entail that the frequency in a hypothetical sequence of tosses will converge on the value.

The actual frequency and the degree of belief interpretations of probability say that we can define probability in terms of something else. However, the hypothetical relative frequency interpretation of probability shows that this interpretation offers no such clarification. This interpretation is actually circular. Consider the fact that an infinite series of tosses of a fair coin does not have to converge on a relative frequency of 0.5. If the frequency of heads does not have to converge on the coin's true probability of landing heads, how are these two concepts related? The law of large number provides the answer.

$$P(\text{the coin lands heads} \mid \text{the coin is tossed}) = 0.5$$

$P(\text{the frequency of heads} = 0.5 \pm e \mid \text{the coin is tossed } n \text{ times})$ approaches 1 as n goes to infinity.
(Here, e is any small number.)

The probability concept appears on both sides of . The hypothetical relative frequency interpretation of probability is not really an interpretation at all, if an interpretation must offer a noncircular account of how probability statements should be understood.

The last interpretation of probability is the propensity interpretation of probability. Propensities are probabilistic dispositions. The propensity interpretation stresses an analogy between deterministic disposition and probabilistic disposition. A dispositional property has an associated behavior and a physical basis. We can discover whether an object has a given dispositional property by exploring either of these. The same is true of probabilistic propensities. We can discover if a coin is fair in one of two ways. We can toss it some number of times and gain evidence that is relevant. Or, we can examine the coin's physical structure and find out if it is evenly balanced.

One of the fundamental problems of the propensity interpretation is that "propensity" seems to be little more than a name for the probability concept we are trying to elucidate.

We now face something of a dilemma. The two coherent interpretations of probability are actual relative frequency and subjective degree of belief. If we think that probability concepts in science describe objective facts about nature, which are not interpretable as actual frequencies, we seem to be in trouble. One possible solution to this dilemma is to deny that probabilities are objective. Before embracing the subjective interpretation, another alternative should be placed. Perhaps probability describes objective features of the world but cannot be defined noncircularly. This might be called an objective no-theory theory of probability.

If we used probabilities only because we wish to make predictions, then the subjectivist would have a point. However, there is another reason to use probabilities. Consider the mating pairs in a population in which both parents are heterozygotes. These parental pairs produce different frequencies of heterozygote offspring. Although each obeys the usual Mendelian mechanism, the mating pairs differ from each other in various ways that account for their different frequencies of heterozygote offspring. We could describe these different mating pairs one at a time and list the

unique constellation of causal influences at work in each. However, another strategy is to try to isolate what these parental pairs have in common. We do this when we describe each of them as participating in a Mendelian process in which $P(\text{offspring is } Aa \mid \text{parents are } Aa \text{ and } Aa) = 0.5$. This simple probability statement might be used to describe the parental pairs in the population even if we possessed detailed information about the unique causal factors affecting each of them. Our reason for using probability is not that we are ignorant; we are not ignorant. If we say that heterozygote parents have heterozygote offspring with a probability 0.5, we are making a very general statement that goes beyond what we actually observe in some finite sample of heterozygote parents and their offspring.¹⁴

4. Selection and Drift

Heredity is determined by a molecule called DNA. The DNA molecule can be divided into regions called genes that encode for proteins. The code in the DNA is read to produce a protein in two stages: transcription and translation. The genetic code has been deciphered. New genetic variation originates by mutational changes in the DNA. When two individuals of given genotypes mate together, the proportions of genotypes in the offspring appear in predictable Mendelian ratios. The exact ratios depend on the genotypes in the cross. Different genes are preserved over the generations under Mendelian heredity, which enables natural selection to operate.

Organisms produce many more offspring than can survive, which results in a struggle for existence, or competition to survive. Natural selection will operate among any entities that produce, show inheritance of their characteristics from generation to the next, and vary in fitness according to the characteristics they possess. The members of natural population vary with respect to characteristics at all levels. They differ in their morphology, their microscopic structure, their chromosomes, the amino acid sequences of their proteins, and their DNA sequences. The members of natural population vary in their reproductive success. Some individuals leave no offspring, while others leave many more than average. The new variation that is created by recombination and mutation is accidental, and adaptively random in direction.

In the absence of natural selection, and with random mating in a large population, the genotype frequencies at a locus move in one generation to the Hardy-Weinberg ratio; the genotype frequencies are then stable. It is easy to observe whether the genotypes at a locus are in the Hardy-Weinberg ratio. In nature, they will often not satisfy this ratio, because the fitnesses of the genotypes are not equal, mating is non-random, or the population is small. A theoretical equation for natural selection at a single locus can be written by expressing the frequency of a gene in one generation as a function of its frequency in the previous generation. The relation is determined by the fitnesses of the genotypes. If a mutation is selected against but continues to arise, the mutation settles at a low frequency in the population. This situation is called selection-mutation balance. Selection can maintain a polymorphism when the heterozygote is fitter than the homozygote, when fitnesses of genotypes are negatively frequency-dependent, and when different genotypes are adapted to different niches. Subdivided populations have a higher proportion of homozygotes than an equivalent large, fused

¹⁴ We should notice that there are many criticisms about the frequency interpretation of probability. For instance, see Hájek.

population.

In a small population, random sampling of gametes to produce the next generation can change the gene frequency. These random changes are called genetic drift. Genetic drift has a large effect on gene frequencies if the population size is small than if the population size is large. If a small population colonizes a new area, it is likely to carry all the ancestral population's genes, but the gene frequencies may be unrepresentative. One gene can be substituted for another by random drift. The rate of neutral substitution is equal to the rate at which neutral mutation arises. In a small population, in the absence of mutation, one allele will eventually be fixed at a locus. The population will eventually become homozygous. The Hardy-Weinberg equilibrium does not apply to small populations. The effect of drift is to reduce the amount of variability in the population. The amount of neutral genetic variability in a population will be a balance between its loss by drift and its creation by new mutation. The effective size of a population, which is the population size assumed in the theory of population genetics for small populations, should be distinguished from the size of a population that an ecologist might measure in nature. Effective population sizes are usually smaller than observed population sizes.¹⁵

The above is a brief summary about selection and drift.

5. Relations between natural selection and drift

5.1 Equivalence class and selection

There have been plenty of evidences that the genes or the genotypes are the fundamental units in heredity. We already have accumulated many empirical evidences that hereditary phenomena consist

¹⁵ Molecular evolution and the neutral theory will be summarized as follows. The neutral theory of molecular evolution suggests that molecular evolution is mainly due to neutral drift. The mutations that have been substituted in evolution were selectively neutral with respect to the genes they replaced. Four main observations were originally interpreted in favor of the neutral theory: molecular evolution has a rapid rate, its rate has a clock-like constancy, it is more rapid in functionally less constrained parts of molecules, and natural populations are highly polymorphic. Kimura argued that the high rate of evolution, and the high degree of variability of proteins, would, if caused by natural selection, impose a high genetic load. Neutral drift can drive high rates of evolution and maintain high levels of variability, without imposing a genetic load. The constant rate of molecular evolution gives rise to a "molecular clock." Neutral drift should drive evolution at a stochastically constant rate. Kimura pointed to the contrast between uneven rates of morphological evolution and the constant rate of molecular evolution and argued that natural selection would not drive molecular evolution at a constant rate. The molecular clock for proteins ticks over according to absolute time rather than generational time. For silent changes in DNA, lineages with shorter generation times probably evolve faster. Neutral drift should cause the molecular clock to run according to generational, not absolute, time. Selection can operate without producing impossible genetic loads, and Kimura's original case for the neutral theory is no longer convincing. The neutral theory explains the higher evolutionary rate of functionally less constrained regions of proteins by the greater chance that a mutation there will be neutral. Selectionists explain the higher evolutionary rate of functionally less constrained regions of proteins by the greater chance that a mutation there will be a small, rather than a large, change. The four properties observed in the evolution of proteins (rates of evolution, constancy of rates of evolution, levels of variability, and relation between functional constraint and evolutionary rate) have also been in DNA. Pseudogenes and silent changes in third codon positions may be relatively functionally unconstrained. These parts of the DNA evolve faster than do the first two positions in codons and meaningful third base changes. Neutralists attribute this high rate of evolution to enhanced neutral drift. For amino acids encoded by more than one codon, consistent biases appear in the frequencies of the codons. Changes between the silent codons are, therefore, not completely unconstrained. The neutral theory predicts a positive relation between the degree of variability of a molecule and its rate of evolution. Genes, and parts of genes, that are more polymorphic than others are predicted to evolve at a higher rate. A comparison of the DNA within the *Adh* gene and its surrounding regions contradicts the neutral prediction, but other genes show a better fit to it. (Ridley, Summary of chapter 7)

of these units. This means that if we think of the genes as biologically fundamental units, the existence of the equivalence classes is inevitable not only in the conceptual level but also in the biological level. Equivalence class has many realizations in many different fields, particularly in biology. In population genetics the realization of equivalence classes has the following properties: the realization of equivalent objects based on the same conditions shows that these equivalent objects have the equal probabilities of occurring and the so-called *ceteris paribus* clause necessarily follows from the construction of the equivalence classes by omitting irrelevant conditions.¹⁶

If we want to observe the particular content of a proposition, we have to neglect other irrelevant parts of the proposition. To observe the whole content of a proposition is not the sum of each particular content of the proposition. This is clear when we remember the uncertainty principle in quantum mechanics. This kind of partial model is not the reason why we have to use probabilistic concept in population genetics. Even when we include all the information to make a model, probability exists in the model because of the existence of the genotypic equivalence classes.

What constitutes the equivalence classes in a molecular level? Molecular structures do and so we can call this equivalence class the *structural* equivalence class. On the other hand, macro-level equivalence classes consist of *functional* equivalence in forms, behaviors or patterns.

If the size of an equivalence class in a population is smaller, the effect of drift is bigger. It is an important thing to make the size of an equivalence class to construct a finer model, but by making a size smaller, we have to face with the bigger effect of drift. The fine-grained model has necessarily the bigger effect of drift.

One of the results implied by the hierarchy and the supervenience of biological phenomena is that biological phenomenon is a kind of equivalence class whose members are physical facts. Then we can say that supervenience and equivalence class is almost the same role in genetics. In this situation, selection and drift cannot be separated. We can see the concrete examples of this equivalence class in the genetic models. This is one of the common characters of genetic models, irrespective of deterministic and stochastic models.

5.2 Frequency

Here we will consider the main differences between selection and drift in a probabilistic framework. To consider things simply, let's use examples.

(1) Unique event, or a continuous state

In this case, there is no good way to describe or explain a state except to trace the change of the state in time. This method is what mechanics have adopted. There is neither selection nor drift here. The spatial change of the object is formalized as a function of time and this formalization also is applied to its explanation, description and prediction.

(2) Actual coin tosses

In this case, basically there are two ways to explain the series of tosses. One way is the same as (1). For example, the series of 10 tosses is one event. The other way is to think that each toss is different

¹⁶ Biological statement usually includes the *ceteris paribus* clause. The relation between equivalence class and *ceteris paribus* clause will give a clue to the *ceteris paribus* controversy. See Pietroski, P. and G. Rey and Earman J. and J. Roberts.

event. If the numbers of heads and tails are different, we may think that there is something working on the difference. Selection may be one of such causes. Selection causes the difference of the numbers. Can we imagine that this cause is drift? Unfortunately no one can imagine drift when she observes only 10 times coin tosses. Drift is a kind of deviation and if the expected number is not presupposed, we have to find its cause differently. This means that actual frequency cannot give the existence of drift, although selection can exist only with actual frequency.

(3) Hypothetical coin tosses

In this case, we can say that there is drift, for expected frequency is assumed and the deviation from it is meaningful. Can we reject the hypothetical coin tosses because it is not empirical? To reject it is the same as to reject non-empirical things in sciences. There are many non-empirical things that are indispensable in sciences. Hypothetical frequency is one of them.

	One closed system	Actual frequency	Expected frequency
Natural selection	×		
Drift	×	×	
Migration	×		
Fluctuation	×	×	

Now what can we say about selection and drift from the above table? We can easily find that the tracing description of an object assume only individual's existence, although selection and drift assume not only an object but also a population of objects. In principle, we can trace all of the members of a population, selection and drift has no meaning only by collecting these individual results. They include something more. Frequency or distribution cannot be constructed only from the elements in a population.

Next thing we should notice is the difference between selection and drift. Selection can be characterized with actual frequency, but drift cannot exist only with actual frequency. Drift cannot exist until we assume hypothetical frequency. This hypothetical frequency can be formalized by the mathematical idealization.

Lastly, we have to mention what is hidden in the above table. Selection has different meaning whether we think it with actual frequency or with hypothetical frequency. Let me explain about the difference. Probabilistic concept is deeply included in the concept of selection as follows.

- (1) Chance is presupposed in variation, random mating, large-size population, and Hardy-Weinberg law. Variation happening at random, random mating, and the combinatorial character of Mendelian laws cannot be defined without probabilistic notions. Particularly, fitness needs hypothetical frequency.
- (2) If empirical data do not confirm the models of population genetics, then we have to reject the models. This is as same as other sciences. Unfortunately, models of population genetics are not complete enough to explain all of the actual evolutionary changes. But we can compare models based on actual frequency with models based on hypothetical frequency. Then the latter can

explain more than the former. In this sense, models based on hypothetical frequency are stronger.

(3) Biased sampling presupposes hypothetical frequency. And arguments relied on the law of large numbers need hypothetical frequency. Hypothetical frequency is mathematically much stronger than actual frequency. This strength gives the richer characterization of selection.

Now we can understand that selection has more contents with hypothetical frequency. If someone asks which you should choose between actual frequency and hypothetical frequency, the answer is clear. Hence we may take hypothetical frequency interpretation of probability to characterize both selection and drift. This common framework of selection and drift suggests the following things.

We cannot find any selection in each individual's lives. There is only the survival and the death of individuals. We can find drift neither. With the hypothetical structure of a population, selection and drift emerge. Let's summarize what we have said.

- A. If selection works little, there appears drift. Although the effect of drift depends on the size of a population, we can see the net effect of drift.
- B. If selection works on the particular trait strongly, the effect of drift becomes stronger in the equivalence class of the trait. But by the Simpson's paradox, sometimes we cannot find its effect in the whole population.
- C. Therefore, if there is selection, then drift always exists. And if the effect of drift is stronger, then selection works stronger, although we might not find the effect.

I have already mentioned that the fitness of a trait need not precisely coincide with the actual frequency of individuals possessing the trait. If two genes, A and a , have the same fitnesses, their frequencies may do a random walk. Given long enough, one or the other will go to fixation. If the fitness of a trait were defined as the actual frequency of individuals with the trait that survive, we would have to describe drift as a process in which genes differ in fitness. Evolutionists accept no such implication; they do not interpret the probabilistic concept of fitness in terms of actual frequencies.

How often is natural selection the only factor at work in a population? This question does not have any answer. Populations always are finite in size, which means that a trait's fitness plus its initial frequency do not absolutely determine its frequency after selection.

Selection and drift is connected more tightly if we use the expected frequency interpretation of probability. If we admit the full definition of fitness to formalize the selection concept, then at the same time we have to admit the statistical concepts emerged by the expected frequency interpretation of probability and one of these concepts is drift. In this sense, we can say that there is no drift without selection.

6. Applications

Let's consider two problems as applications of our results. The following problems have been

discussed recently. And still we do not have any clear answer.

(1) Is drift subjective?¹⁷ (2) Can we talk about a particular individual in a genetic model? ¹⁸

Informational characterization of statistical mechanics has the same connotation as Rosenberg's interpretation of drift. According to this characterization, the subject of statistical mechanics is not the physical properties themselves, but their information.¹⁹ Drift is only subjective and so the evolutionary theory is deterministic. This is what Rosenberg claims. We can easily reject this claim according to what we have made clear. Selection presupposes essentially the equivalence classes of the traits of organisms in a population and this equivalence class induces necessarily probability. Within an equivalence class we have to assume the equal probability among each member. And finally this equal probability induces necessarily drift.

We cannot talk a particular object within a genetic model even though selection is cumulative. Here also the existence of equivalence class prevents from talking about a particular thing. We may explain a particular thing by using knowledge supplied by genetic models, but we cannot describe a particular thing. Accumulation of hierarchical classes cannot fix a particular object. Dividing the lower class into more than two classes makes upper classes. Then accumulation of selections means that random drift becomes more effective in the accumulating processes, because the equivalence classes become smaller.

7. Two conclusions

We began our discussion with symmetry and asymmetry. If a system is time-asymmetric, then its origin, history and probability can have the real meaning. Determinism and fatalism can be separated by the asymmetry. And if we assume the hierarchical structure of the world, then supervenience and equivalence class are indispensable to investigate things in special sciences. Moreover, they give the direct basis for introducing probability into special sciences. This is our first conclusion.

Next we applied the above result to the relation of natural selection and genetic drift in population genetics. We may summarize the conclusion in the form of the following argument.

(An argument for selection and drift)

1. We can divide all of the organisms in a population into genotypic subpopulations when we consider evolutionary changes of a particular species in population genetics. For example, the simplest case is that organisms are divided into three classes according to the genotypes *AA*, *Aa* and *aa*.
2. The division of 1 makes equivalence classes modus genotype. We cannot distinguish organisms within an equivalence class in the model. Hence the organisms of genotype *AA* are all equivalent in the model.

¹⁷ Whether the evolutionary theory is deterministic has been discussed mainly around the concept of drift. I do not think that the evolutionary theory is indeterministic like quantum theory. But at the same time I do not think that the evolutionary theory is deterministic like classical mechanics. See Brandon, R. and J. Beatty, Brandon, R. and S. Carson, Graves L., B. L. Horan and A. Rosenberg, Horan, Millstein, Nishiwaki, and Rosenberg.

¹⁸ This problem was proposed first by Neander. See Matthen (1998), (1999), Neander (1995), (1995), Sober (1996).

¹⁹ See Guttman.

3. (Darwin's argument, excess fecundity) If excess fecundity is assumed, the equivalence class of generation $n + 1$ is larger than generation n . If the size of population has to be constant, then some organisms of the same genotype have to be eliminated from the population.
4. But there is no selective difference among the organisms of the same genotype. They are neutral. So random drift or non-biased sampling is required. (Here the principle of equal probability is used.)
5. Therefore, selection and drift occurs at the same time in the population. Selection works among the different genotypes, which we may call the selection between types, and drift works in the same genotype, which we may call the drift within a type.

We have discussed how equivalence class is important. Within an equivalence class, all the genotypic tokens are neutral to each other. And second conclusion is that whenever selection works, there is always drift.

References

- Brandon, R. and J. Beatty (1984): "The Propensity Interpretation of Fitness: No Interpretation is No Substitute", *Philosophy of Science* **51**, 342-347.
- Brandon, R. and S. Carson (1996): "The Indeterministic Character of Evolutionary Theory: No 'No Hidden Variables Proof' but No Room for Determinism Either", *Philosophy of Science* **63**, 315-337.
- Darwin, C. (1859): *On the Origin of Species by Means of Natural Selection*, John Murray.
- Dawkins, R. (1976): *The Selfish Gene*, Oxford University Press.
- Gayon, J. (1998): *Darwinism's Struggle for Survival (Darwin et l'après-Darwin)*, Editions Kimé, Paris, 1992), Cambridge University Press.
- Guttmann, Y. M. (1999): *The Concept of Probability in Statistical Physics*, Cambridge University.
- Earman J. and J. Roberts (1999): "Ceteris Paribus, There Is No Problem of Proviso", *Synthese* **118**, 439-478.
- Findley, A. M., S. P. McGlynn and G. L. Findley (1989): *The Geometry of Genetics*, John Wiley & Sons.
- Graves L., B. L. Horan and A. Rosenberg (1999): "Is Indeterminism the Source of the Statistical Character of Evolutionary Theory?", *Philosophy of Science* **66**, 140-157.
- Hájek, Alan (1997): "'Mises Redux' - Redux: Fifteen Argument Against Finite Frequentism", *Erkenntnis* **45**, 209-227.
- Horan, B. L. (1994): "The Statistical Character of Evolutionary Theory", *Philosophy of Science* **61**, 76-95.
- Kim, J. (1993): *Supervenience and Mind*, Cambridge University Press.
- Kimura, M. (1983): *The Neutral Theory of Molecular Evolution*, Cambridge University Press.
- Mackey, M. C. (1992): *Time's Arrow: The Origins of Thermodynamic Behavior*, Springer-Verlag.
- Matthen, M. (1998): "Biological Universals and the Nature of Fear", *The Journal of Philosophy* **XCIV**, No.3, 105-132.
- Matthen, M. (1999): "Evolution, Wisconsin Style: Selection and the Explanation of Individual Traits", *Brit. J. Phil. Sci.* **50**, 143-150.
- Millstein, R. (1996): "Random Drift and the Omniscient Viewpoint", *Philosophy of Science (Proceedings)*, S10-S18.
- Neander, K. (1995): "Pruning the Tree of Life", *Brit. J. Phil. Sci.* **46**, 59-80.
- Neander, K. (1995): "Explaining Complex Adaptations: A Reply to Sober's 'Reply to Neander'", *Brit. J. Phil. Sci.* **46**, 583-587.
- Nishiwaki, Y. (1998): "Are genetic models the result of classical ignorance?", *Philosophy of Science (Japan)* **31**, 1-16.

- Pietroski, P. and G. Rey (1995): "When Other Things Aren't Equal: Saving *Ceteris Pribus* Laws from Vacuity", *Brit. J. Phil. Sci.* **46**, 81-110.
- Ridley, M. (1996): *Evolution*, 2nd ed., Blackwell.
- Rosen, J. (1995): *Symmetry in Science*, Springer-Verlag.
- Rosenberg, A. (1994): *Instrumental Biology or the Disunity of Science*, Cambridge University Press.
- Roughgarden, J. (1976): *Theory of Population Genetics and Evolutionary Ecology*, Prentice Hall (Reprint, 1996).
- Sober, E. (1984) *The Nature of Selection, Evolutionary Theory in Philosophical Focus*, MIT Press.
- Sober, E. (1993): *Philosophy of Biology*, Oxford University Press.
- Sober, E. (1996): "Natural Selection and Distributive Explanation: A Reply to Neander", *Brit. J. Phil. Sci.* **46**, 384-397.
- Tolman, R. C. (1938): *The Principles of Statistical Mechanics*, Oxford University Press (Dover ed. 1979).
- Williams, G. C. (1966): *Adaptation and Natural Selection*, Princeton University Press.