

Changes in the Concept of “Fitness” in Evolutionary Biology

ISEDA Tetsuji

1. Introduction

Philosophers of science have been speculating on the way scientific theories change, but it is quite recently that they start to found their speculations on concrete historical researches on actual changes in science. Hull's (1988a and 1988b) and Darden's (1991) theories on conceptual changes are two of such recent attempts (Hull uses taxonomy in biology, and Darden uses Mendelism). In this paper I would like to examine their arguments by applying them to another case in biology. The example I use is the concept of "fitness." This concept has been a central concept in evolutionary biology, but at the same time it has been confusing one, and we can find many different usages (Dawkins 1982 distinguished five major usages and added more. 179-194). The history may partly support Hull's account of conceptual change (Hull 1988a and 1988b), but this cannot be the whole story. We will recognize that some theoretical requirements drove the changes, and need to admit something like Darden's view on conceptual

change (Darden 1991). In the course of the tracing of the history, we will also see how philosophical conceptual analyses help the understanding of the history.

In section 2, I summarize Hull's and Darden's views on conceptual change. In section 3 through section 6, I go through case studies from the history of the concept. Section 3 deals with Spencer's original introduction of "fitness" and Darwin's adoption of it. Section 4 deals with Social Darwinism and its use of "fitness" as a normative word. In section 5, I discuss biologists' attempt to measure "fitness" and the problem this attempt caused, namely the tautology problem. Section 6 deals with Hamilton's "inclusive fitness" and its influence. Finally in the last section, I shall summarize the history and factors that have acted on the history. The tables at the end of the paper review and compare different notions of "fitness" dealt with in the paper.

2. Philosophical views on conceptual change -- Hull and Darden

2.1 Hull's evolutionary view of conceptual change

David Hull put forward an explanation of conceptual change in terms of social factors in a scientists' community (Hull 1988a and 1988b). His explanation uses an analogy to the evolution of species. For this purpose, he abstracts essential parts of evolutionary theories. First, he distinguishes *replicators* and *interactors*. A replicator passes on its structure largely intact to

successive generations. An interactor interacts with its environment and this interaction causes differential replication of replicators. *Selection* is a process in which the difference in success of interactors causes differences in replication. As a result of selection, some replicators pass on their structure through time with or without small changes; this temporal succession is called a *lineage* (Hull 1988b, 134-135).

Hull applies these notions to science itself (Hull 1988a, 434). Replicators in science are beliefs, goals, methodologies, and so on. Interactors are scientists. Scientists act for their conceptual inclusive fitness, namely, so as to encourage other scientists to use their work. For example, scientists give credit to other scientists in their own work because this increases the credibility of the work. This will in turn increase the possibility that the work is cited by other scientists (Hull 1988a, 310). Another example is the relationship between a scientist and his/her own graduate students (Hull 1988b, 127-128). On the one hand, the scientists are not required to give credit to the graduate students. On the other hand, graduate students are "likely to be the chief conduits for one's work to later generations" (Hull 1988b, 128). The scientist should balance these two considerations to maximize his/her own conceptual inclusive fitness.

This account of conceptual change lacks an important element in an evolutionary theory, that is, Hull does not explain how new variations come in (Cain and Darden 1988, 165). Scientists come up with new concepts or modification of an old concept, and

without these new variations, science cannot evolve. When and why do they introduce new variations? ¹Hull's view on the mechanism of selection in scientific activities is also unclear. Thus far in the literature critiquing Hull's view, I have found no mention on this point, but a selection takes place in a scientist's mind, when he/she decides which paper to cite. Hull's account is incomplete if he cannot account for their criteria for the selection. Perhaps Hull's answer is that we need to do psychological investigation to know exactly what occurs.

2.2 Darden's view on conceptual change

Hull's account takes into account social factors in conceptual changes, but many philosophers point out that there are also theoretical considerations for conceptual changes. Here we consider Lindley Darden's view on conceptual change (Darden 1991, 168-190). She distinguishes empirical problems and conceptual problems. Empirical problems come from anomalies in evidence. Conceptual problems are not about empirical anomalies, but are "about the adequacy of symbolic representations, the introduction of new theoretical terms, and disputes about the need for and properties ascribed to a new theoretical entities" (170). In the course of examining the conceptual change in "gene" in Mendelian genetics, she distinguishes these problems and their

¹ Actually my point is a little different from Cain and Darden's. Their argument is that variations are essential for selection. I agree with this. But my point is that *new* variations are essential for *evolution*.

solutions. To understand their solutions, she introduces hypothetical "strategies" scientists might have used -- finding the referent of the symbols, changing old components slightly, using analogy, introducing new terminology, postulating an underlying causal factor, and so on (188-190; as for the hypothetical character of the strategies, see also 15-17). She also locates conceptual change by making lists of general properties of several notions (allelomorph, chromosome, factor, gene)(185-186). I will make similar tables for historical changes of fitness at the end of this paper.

Darden's analysis shows that scientists have internal (non social) reasons for conceptual change, and gives good insights into how new concepts are introduced. In this sense, Darden's view complements Hull's view. On the other hand, recognition of conceptual problems and the choice of a strategy may be influenced by social factors as Hull describes. Moreover, scientists might have "social" problems and strategies to solve them along with empirical and conceptual ones. Sometimes such social strategies may give us more plausible hypotheses about the methods scientists used. In such cases, Darden's account may need to be supplemented by Hull's view.

In the next several sections, we will examine the changes in the concept of "fitness." These examples show that both social and internal factors act to produce the change.

3. "Survival of the fittest" and its adoption

3.1 Spencer's "survival of the fittest"

The first usage of the word "fitness" as a theoretical term in evolutionary theory is credited to Herbert Spencer (1864). Herbert Spencer introduced the phrase "survival of the fittest" as an interchangeable phrase to Darwin's "natural selection, or the preservation of favoured races in the struggle for life" (Spencer 1864, 444-445. Spencer's quotation from Darwin is from the latter part of the title of the *Origin of Species*). To see what Spencer meant by this phrase, let us see his discussion of it. Individuals in a species are "necessarily made unlike, in countless ways and degrees" (Spencer 1864, 444). With these variations, when the environment has changed, "some will be less liable than others to have their equilibria overthrown by a particular incident force, previously unexperienced" (444). That is to say, "those will survive whose functions happen to be most nearly in equilibrium with the modified aggregate of external forces" (444). This is survival of the fittest. Except for his peculiar word "equilibrium," this account is almost the same as the account of natural selection by Darwin (Darwin 1859, 80-81). There are several important points here in comparison with later usages of "fitness." First, here is no direct definition of "fitness." It is defined indirectly by defining the "survival of the fittest" as synonymous to "natural selection." Second, this "fitness" is relative to the environment, and individuals "happen to be" the fittest, relative to the new

environment. This point is important when compared with the usage by Social Darwinists. Finally, here is no explicit mention of reproductive success, though this will be a central part of the meaning of "fitness."

3.2 Darwin's original usage of "fitness" in the first edition of the *Origin*

Actually Darwin himself used the word "fitness" and "fitted" several times in the first edition of the *Origin of Species* (Darwin 1859, 88, 91, 472, 480; see also Paul 1994, 112). Once he used the phrase "the continuous preservation of the individuals best fitted" in almost the same sense as Spencer's "survival of the fittest" (Darwin 1859, 91). But this was not a central theoretical term (in his general statement of natural selection, Darwin did not use the word; Darwin 1859, 80-81), and there was an important difference between his usage and Spencer's one. First, take a look at these two examples:

Nor ought we to marvel if all the contrivances in nature be not, as far as we can judge, absolutely perfect; and if some of them be abhorrent to our ideas of fitness. (Darwin 1859, 472)

...we may believe, that the teeth in the mature animal were reduced, during successive generations, by disuse or by the tongue and palate having been fitted by natural selection to browse without their aid... (Darwin 1859,480)

As is obvious from the second quotation, Darwin's "fitness" is caused by natural selection. In this sense, as Paul suggests (Paul 1994, 112), Darwin's "fitted" and "fitness" are

interchangeable with "adapted" and "adaptation." What is the difference between "adaptation" and "fitness"? According to Burian, "adaptation" has two primary meanings in evolutionary context: one is "transgenerational alterations of the features and capacities of organisms" that enable the organisms to solve their problems (Burian 1994, 7); the other is "a trait or capacity" as the product of this process (Burian 1994, 7). So the major difference between two concepts is that "adaptation" implicitly mentions the history of alteration which make the feature adapted, while "fitness" has no such implication. Now, the above quotations from Darwin show that his "fitness" refers to the history of alteration. The first quotation is talking about the reason why we find imperfect adaptations. If Darwin was using the word "fitness" in the same sense as Spencer, this comment does not make sense. For we have no reason to assume the organism which happens to be the fittest one is also the perfect one for the environment, thus an imperfect adaptation is not "abhorrent to our ideas of fitness" in this sense. If we take the "fitness" in the sense of historical alteration, on the other hand, it does make sense to ask why such an alteration does not make a perfect adaptation. The second quotation is about why the calf has teeth which they never use. Darwin's explanation mentions the transgenerational alterations in the way they browse. Thus this "fitted" is also replaceable with "adapted" in Burian's sense. Given these considerations, we can conclude that Darwin's original "fitness" is clearly different from

Spencer's usage in the "survival of the fittest."² This may be a trivial point because anyway "fitness" was not an important theoretical term for Darwin, but the point of this section is to establish the priority of Spencer in introducing a new concept. I think this argument is enough for this purpose.

3.3 Darwin's adoption of "fitness"

It was A. R. Wallace who recommended to Darwin that he adopt the new phrase "survival of fittest" (Wallace 1866, 140-141). According to Wallace, the "survival of the fittest" is "the plain expression of the *fact*," and "natural selection" is "a metaphorical expression of it" and is "to a certain degree *indirect* and *incorrect*" (141, emphases original). Wallace also cited two examples in which critics had misunderstood the word "selection" and had claimed that natural selection requires a chooser (141). Darwin accepted Wallace's proposal (Darwin 1866, 144). He adopted the phrase the "survival of the fittest" in the fifth edition of the *Origin* (Paul 1994, 112). As we can see in the sixth edition of the *Origin*, he changed the title of the chapter 4 from "natural selection" to "natural selection; or the survival of the fittest" (Darwin 1872, 97). He explained the reason saying that the "Survival of the Fittest is more accurate, and is sometimes equally convenient" (Darwin

² Actually Spencer himself used the word "fitness" in a confusing way; "[t]o him [Darwin] we owe the discovery that natural selection is capable of producing fitness between organisms and their circumstances..." (Spencer 1864, p.446, emphasis original). This "fitness" seems to mean something like "harmony." Moreover, as you see in the next section, his usage of "fitness" in his political writings is quite different from these usages.

1872, 77).³ He also added a comment on the misunderstanding which Wallace pointed out (Darwin 1872, 99).

As Wallace's letter and the changes in the *Origin* suggest, the reason for adopting the phrase "survival of the fittest" was two-fold. One is a theoretical reason. By eliminating the metaphorical word "selection," Wallace thought that Darwinism became a more accurate description of the fact of the matter. In this point of view, Darwin's "preservation of the most fitted" or "the preservation of favoured races" was not enough, because these phrases still personify the work of natural selection. But this theoretical consideration is driven by the other reason, a social reason. According to Bowler, the basic idea of evolution was soon accepted by other biologists, while the idea of natural selection as a mechanism for evolution met much resistance (Bowler 1989, 188). So the task of Darwin and Wallace was not only to develop the theory, but also to make the theory accepted by other scientists. From this point of view, if the word "natural selection" can be a stumbling block to understanding the theory, this is a sufficient strategic reason to adopt an alternative expression. And as Wallace pointed out, "selection" did cause a misunderstanding. This misunderstanding might come from the intellectual background of the age. The orthodox theory in the West in the nineteenth century was creationism, that is, the theory that all

species are (separately) created by God *purposively*. With this teleological paradigm, people could easily read the word "selection" with teleological connotation. On the other hand, the "survival of the fittest" does not allow such a teleological reading. To understand Wallace and Darwin's attitude, we need to take account of this background.

Then, what was the result of adopting the phrase? When Darwin accepted Wallace's criticism, he doubted that there would be any effect of replacing natural selection with "survival of the fittest." "The term Natural Selection has now been so largely used abroad and at home that I doubt whether it could be given up, and with all its faults I should be sorry to see the attempt made. Whether it will be rejected must now depend 'on the survival of the fittest'" (Darwin 1866, 144). The history proved that both of the terms are good replicators in Hull's sense. We now still use both. But the "survival of the fittest" and "fitness" seems to be a even better replicator. Mills and Beatty points out that fitness "still plays a major role in explanations of evolutionary phenomena" (Mills and Beatty 1979, 267). My conjecture about the reason for this popularity of "fitness" is that the notion promises us a kind of convenience in thinking. When we see things from a "natural selection" point of view, we should think in macroscopic level, namely the environment and the organisms in it as a whole. We

³ Darwin thought that "the survival of the fittest" is sometimes not equally convenient as "natural selection" because "it [the survival of the fittest] cannot be used as a substantive governing a verb" (Darwin 1866, 144). This consideration may suggest another interesting motive for conceptual change.

need imagination to think in this way. On the other hand, if we see things from a "fitness" point of view, we can start from an organism and its characteristics and then proceed to think about interactions with the environment. This approach may lead us to the biases of reductionistic research Wimsatt has pointed out (Wimsatt 1980, 232-233), but it also make it easy for biologists to imagine the situations and to find the solutions for their problems. This advantage in the economy of thinking will be reflected in the creativity of the biologists who use the "fitness" way of thinking, and will increase their conceptual inclusive fitness (if the advantages outweigh the disadvantages from reductionistic biases). If this conjecture is right, this explains why people keep using the concept of fitness.

4. Fitness and Social Darwinism

4.1 Three versions of Social Darwinism

Besides the theoretical meaning in biology, the phrase "survival of the fittest" acquired social and political meanings (Paul 1994, 113). This application of Darwinism to society is usually called "Social Darwinism," but this is not a single political movement (Jones 1980, Clark 1984). Darwinism was used sometimes to defend laissez-faire capitalism, sometimes liberal reformationism, sometimes eugenics. We shall concentrate on the role of the notion "fitness" played in these applications.

The first example of Social Darwinism is a justification of laissez-faire capitalism. The most famous defender of this position is Herbert Spencer himself. In a paper titled "The sins of legislators," Spencer connects biology and his political claim as follows:

[s]trange to say, now that this truth [evolution by natural selection] is recognized by most cultivated people -- now that the beneficent working of the survival of the fittest has been so impressed on them that, much more than people in past times, they might be expected to hesitate before neutralizing its action -- now more than ever before in the history of the world, are they doing all they can to further survival of the unfittest! (Spencer 1994, 131)

By "the unfittest," he means "the undeserved poor" (Spencer 1994, 134). Here the normative connotation of "the fittest" and "the unfittest" plays an important role in his rhetoric.

Secondly, reformationists used the notion of fitness to justify their theory (this is sometimes called "Reform Darwinism." Clark 1984, 3). They agreed that survival of the fittest is desirable, but they interpreted fitness as superiority in morality. For example, L. T. Hobhouse claimed that "that the morally fittest shall actually survive is the object of good social institutions" (Hobhouse 1893, quoted in Jones 1980, 63).

Finally, the most influential type of Social Darwinism was eugenics. Eugenics was founded by Francis Galton, a cousin of Charles Darwin (Kevles 1985, 3-19). He observed that physically and mentally inferior people were the most fertile, and concluded that natural selection no longer operated in human society (Jones

1980, 99-100). This conclusion shows that Galton associated physical and intellectual superiority with fitness, and thought such superior people should survive in natural selection. Galton's solution was to encourage those who had the desirable qualities to multiply faster than others (Jones 1980, 99). Some eugenicists were aware that this "fitness" was not the same as "fitness" in biology. Kevles summarizes this awareness as follows: "[i]f natural selection yielded the Darwinian fit, only artificial selection --by governmental means, where appropriate -- could multiply the eugenically fit" (Kevles 1985, 91). Later eugenics was used as an ideological basis for Nazi's holocaust in Germany. Zmarzlik describes the ideological components of it as follows:

[A] biologicistic dogma of racial inequality; a moral nihilism invoking the "struggle for existence" and the "survival of the fittest" as a universal law of nature; and -- resulting from both of these -- the conviction that radical extermination of the racially inferior elements and the selection of racially superior elements are justified by the fact that these policies are a vital necessity to a people that wishes to be strong. (Zmarzlik 1972, 435)

Here fitness is associated with racial superiority (whatever it means).

Even though their political positions are conflicting with one another, these three usages of "fitness" have strong similarities. First, there is a curious inversion from biological usage in the logical relationship between selection and fitness. As we saw in section 3.1 of this paper, Spencer originally introduced the concept

of fitness in terms of selection. But Social Darwinists introduce the fitness independent from the environment, and then proceed to say that the fittest ought to be selected regardless of the environment (or, maybe, we should change the environment so that the fittest can survive). Thus Social Darwinists' fitness is logically independent from the environment and selection. This is obvious in Reform Darwinism and eugenics, because their "moral superiority" or "physical and intellectual superiority" refers to human capacity that can be measured independently of the environment. Of course Spencer was much more careful on this point, but it seems to me that he also committed a similar inversion when he called the poor the "unfittest" even though legislators started to make laws to protect them. By this change of political environment, the poor started to be selected for, so if we think of fitness in terms of selection, they are no longer the unfittest. Spencer's "fitness" in Social Darwinism starts to part from his "fitness" in biology here. Another related similarity among these usages is the role the word played in their normative claims. Jones points out: "[t]he idea of 'fitness' tended to be imbued with conventional notions of the desirable and valuable" (Jones 1980, 8). In the original biological usage, of course, "fitness" has no moral evaluative meaning such as this.

It is easy to ignore these usages as abuses of the notion, but we should consider why such abuses were possible. First, as I suggested at the end of previous section, the change from "natural selection" to the "survival of the fittest" seems to include more

than a change of terminology, namely, it also involved a change in the way of thinking. The change enabled us to think from the organism's level. But this change also led Social Darwinists to a kind of bias the sort discussed by Wimsatt (1980). An appropriate bias in this case is what Wimsatt calls "descriptive localization" (Wimsatt 1980, 232). Descriptive localization is to "[d]escribe a relational property as if it were monadic, or a lower order relational property; thus, e.g., fitness as a property of phenotypes (or even of genes) rather than phenotype-environmental relation" (Wimsatt 1980, 232). The bias applies to this case of social Darwinism, as we saw above. This bias is less likely to happen if we keep thinking in the "natural selection" way. Second, the choice of the word "fit" is problematic. The word "fit" had a positive normative meaning before the biological usage, and it was very easy to confuse the biological claim that the fittest tends to survive with the moral claim that the fittest *should* survive. Maybe Spencer had chosen the word deliberately for this purpose.

4.2 Responses from biologists

Biologists struggled to dissociate biology from these forms of Social Darwinism, especially from eugenics. Sometimes they even tried to replace "fitness" with other words, such as "adaptive value" (Paul 1994, 113). For example, Dobzhansky rarely used the word "fitness" in his (1937), and used "adaptive value," "survival value" and so on (Dobzhansky 1937. See pp. 153, 171, 178, 187 and

so on).⁴ J. B. S. Haldane tried to distinguish biological "fitness" and the "fitness" the eugenists talked about (Haldane 1938, 97-99). He claimed that fitness "in the Darwinian sense" is assessed by "average number of offspring left" (Haldane 1938, 78). (We will discuss this measurement in the next section). If we understand fitness in this sense, "we find that in many cases the eugenists are demanding the sterilization of the fit" (Haldane 1938, 99). But he does not intend to object to eugenics for biological reasons. "Man should not follow nature blindly. He should, and does, interfere with natural processes, including natural selection" (Haldane 1938, 99). In short, Haldane's points are two-fold: first, eugenists use the word "fitness" inaccurately; second, biology is neutral about normative judgments. I think that this is a common attitude of biologists.

These responses from biologists show an interesting interaction between social factors and a biological concept. What motivated biologists to dissociate biology from Social Darwinism? Maybe the answer is that Social Darwinism (especially eugenics) was infamous when Dobzhansky and Haldane wrote their books (1930s - 40s), and to be associated with it was disadvantageous for biology. If this answer is correct, then the motivation was a kind of conceptual inclusive fitness, not for a concept, but for biology itself.

⁴ More exactly, he used the adjective form "fit" several times (Dobzhansky 1937, 126, 187). This may come from a practical reason that "adaptive value" and so on have no appropriate adjective form. In his later work, Dobzhansky used the word "Darwinian fitness" as a synonymous phrase to "adaptive value" (Dobzhansky 1955, 119, 122).

5. Reproductive success and the tautology problem

5.1 Population geneticists and "fitness"

Until 1930s, the concept of the fitness remained a vague notion, without exact definition (this is a part of the reason Social Darwinists could interpret the word freely). In the course of the synthesis of Darwinism and Mendelism, the exact measurement of "fitness" was attempted by population geneticists. According to Kimbrough (1980), the first attempt of this kind was perhaps made by R. A. Fisher (Kimbrough 1980, 159; Fisher 1930, 21-47). Fisher's measurement goes as follows. First, he mathematically defines a measurement m as the relative rate of increase (or decrease) of a population (25-26). Next, he introduces the word "fitness" in terms of m : " m measures the fitness to survive by the objective fact of representation in future generations" (34). For example, if two populations have different sets of genes and accordingly have a different relative rate of increase, the population which has the larger rate of increase has also greater fitness. The concept of fitness is applied to a population of individuals, not to an individual organism. The fitness of an individual is the "expectation of offspring" (Fisher 1930, 25) derived from the population's average number of offspring. We should note that Fisher never defined "fitness" itself. As the above quotation shows, he took the concept of fitness as intuitively obvious.

Haldane introduced the phrase "fitness in the Darwinian sense" (Haldane 1932, 90; Haldane 1938, 78). His classical book, *The Cause of Evolution* has a whole chapter named "What is fitness?" (Haldane 1932, 111-143), but he did not try to either define the word, nor give an exact measurement of it.⁵ Instead, his purpose in this chapter is to know "what is actually selected" (Haldane 1932, 111). If these two questions are intended to be identical, then Haldane's definition of fitness should be "what is actually selected." His (1938) gives us a better understanding of the phrase. When he introduces the term, he says, "fitness, assessed in the Darwinian sense on the basis of the average number of offspring left" (Haldane 1938, 78). Thus average number of offspring is an assessment, not a definition. This book has another passage that sounds like a definition of fitness. According to the passage, "fitness" is used "to refer to individuals of such a constitution that they are likely to propagate themselves in larger numbers than their fellows, either as a result of being better adapted to their environment or more fertile, or both" (Haldane 1938, 97). It is not clear if this is supposed to be a definition, for this is found not in a theoretical book, but in political writing. If we can take this as a definition, it is almost the same as the propensity interpretation of Mills and Beatty (1979).

Lastly, let us take a brief look at another leading population geneticist, T. Dobzhansky. As I mentioned in the previous section,

⁵ He introduced Fisher's measurement in another place, but he did not associate it with the "fitness in Darwinian sense" (See Haldane 1932, 172).

he did not use the word "fitness" in his (1937). But he introduced the word "Darwinian fitness" in his later works. For example, in his (1955), he says, "[t]he viability and the reproductive success determine the contribution which the carrier of a genotype make to the gene pool of the next generation of the species or of a population. This contribution is a measure of the *adaptive value*, or *Darwinian fitness*, of the genotype" (Dobzhansky 1955, 119-120, emphasis original). Except for the introduction of the notion of the "gene pool," there is no essential difference between this formulation and Fisher's. Again, the contribution of the viability and the reproductive success to the gene pool is a measure, not a definition, of Darwinian fitness; again, fitness is primarily about genotype.

Thus, these founders of population genetics used the word "fitness" in a very consistent way. This "Darwinian fitness" measured by expectation of reproductive success became an orthodox view of fitness. This status is easily understandable if we consider the importance of population geneticists in the history of Darwinism. Their works revived Darwinism from its eclipse by synthesizing it with Mendelian genetics (Bowler 1989, 307-318). It is natural that this measurement of "fitness" was accepted along with the other parts of this works.

5.2 The tautology problem and two interpretations of fitness

This measurement by reproductive success became a definition of "fitness." But it was not done in a very careful way. For example, Waddington wrote in 1939 "the fitness of the organism as measured by the number of offspring it leaves" (Waddington 1939, 287), so the number of offspring was not a definition of fitness. But in 1957, he changed his mind. "[T]o speak of an animal as 'fittest' does not necessarily imply that it is strongest, or most healthy, or would win a beauty competition. Essentially it denotes nothing more than leaving most offspring" (Waddington 1957, 64-65). Here he almost defines the fitness by the number of offspring. Moreover, we notice that in both quotations he talked about not the expectation of reproduction, but the actual number of offspring. Needless to say, this was not what Fisher and Haldane intended. This definition by actual reproduction caused theoretical problems later, namely a criticism that natural selection is a worthless tautology (Dawkins 1982, 180). If we say "the fittest one is the one who left the most offspring," then we know the fittest only by hindsight and there is no possibility of falsification of the claim about fitness (Dawkins 1982, 184). Mills and Beatty suggest that the problem comes from the definition by actual number of offspring (Mills and Beatty 266-269). If we interpret fitness as propensity of an organism to leave offspring, the tautology disappears. They define fitness primarily for an organism, and extend the definition to types

(Mills and Beatty 272-282). The propensity interpretation has another advantage, namely it is measured by expectation of offspring, thus it is suitable to the usage of population geneticists.

The propensity interpretation removed a part of the problem. But, as Beatty (1994) points out, we should be careful when we reject that the "survival of the fittest" is tautologous because of this propensity interpretation. Suppose we take "the fittest will survive" as a central statement of Darwinism and define "fitness" as "propensity to survive." If we restate the first statement by this definition, we should say, "those who have the largest propensity to survive will survive." This second statement, nevertheless, is clearly false as a universal statement, because it is possible that those who have larger propensity to survive unluckily do not survive. Thus we should understand it as a probabilistic statement, "the probability that those who have the largest propensity to survive will survive is high." This last statement is again tautologous because the propensity to survive and the probability to survive means almost the same thing (for this argument, we need more investigation of the notion of probability itself. See, for example, Salmon 1967 for different philosophical interpretations of probability).

Gould tried to avoid this conclusion by suggesting independent criteria for fitness (Gould 1977, 39-45). In a given environment, certain designs are superior to others a priori, "by an engineer's criterion of good design" (Gould 1977, 42). But this solution just postpones the tautology one step. What does "good design" mean?

Running faster is not necessarily good if it requires sacrifice in other aspects. The organism also needs to eat, to bear children. So the best design should be somewhat balancing these requirements. Where is the maximizing point? The answer to the question amounts to the highest expectation of survival or reproduction. In other words, a good design as a whole organism amounts to the fitness in the sense of the propensity to survive. Thus, Gould's attempt to avoid tautology is not very successful.

It seems to me that the tautology problem is totally misguided. This problem comes from confusion between definition and description. Remember how Spencer introduced fitness. He defined the "survival of the fittest" as synonymous to "natural selection." Thus, the statement "the fittest are those who survive in natural selection" is true by definition. In other words, the "survival of the fittest" is not a description of the fittest, but the definition of the fittest. A definition is naturally tautologous and there is no problem here. Does this mean Darwinism has no empirical content? No. As Kimbrough (1980) and Beatty (1994) point out, as far as natural selection itself can be stated without mentioning "fitness," and has empirical content, the tautologous status of the "survival of the fittest" causes no problem to Darwinism. And, in fact, we can find such a statement of natural selection in Sinnott (et al. 1958): "carriers of different genotypes transmit their genes to the succeeding generations at different rates" (247). This statement can be checked empirically, and we can define the "survival of the fittest" and the "fitness" in terms of

this difference in transmission rate. The tautology problem stems from the misunderstanding of these relationships between central concepts.

6. Inclusive fitness

6.1 Hamilton's "inclusive fitness"

D. Hamilton noticed that the orthodox measure of fitness by reproduction is not enough because it admits "no possibility of the evolution of any characters which are on average to the disadvantage of the individuals possessing them" (Hamilton 1964a, 1), though we find "self-sacrifices" in nature. Hamilton's idea is that if the relatives who receive the benefit have genes "identical by descent," such sacrifices can evolve. To establish this point, Hamilton formulates "neighbour modulated fitness" of an organism in the first part of his paper (2-5). This is an organism's reproductive success as "the sum of the basic unit ⁶, the effect δa of his personal genotype and the total e° of effects on him due to his neighbours which will depend on their genotype" (3). Neighbour modulated fitness is obtained by slightly modifying the orthodox notion of fitness, but "rather unwieldy" because it requires messy calculations (5). So Hamilton proposes another way to see the same situation: "[e]very effect on reproduction

⁶ He means by "basic unit" the degree of fitness "which, if possessed by all the individuals alike, would render the population both stationary and non-evolutionary" (Hamilton 1964a, 2).

which is due to A [an organism] can be thought of as made up of two parts" (5). On the one hand, the genotype influences the reproduction of the organism who has the genotype. On the other hand, the genotype can influence the reproductive success of the relatives of the organism. He introduce a measure $R^{\circ} ij$ of a genotype ij of a single locus (i and j stands for two alleles of the locus). $R^{\circ} ij$ is expressed as $1 + \delta R^{\circ} ij$, where 1 stands for the basic unit (see footnote 6) and $\delta R^{\circ} ij$ stands for the total effect on genes i and j in relatives of the organism which possesses the genotype (including the effect on the organism itself) (5-6). Then he says, " $R^{\circ} ij$ will be called the *inclusive fitness*, $\delta R^{\circ} ij$ the *inclusive fitness effect*" (6, emphases original). Thus this is the definition of "inclusive fitness." He proved that inclusive fitness maximizes in the course of selective change (7). Inclusive fitness is defined for a genotype, but it is applicable to each individual (8). In his idealized model, "we expect to find that no one is prepared to sacrifice his life for any single person but that everyone will sacrifice it when he can thereby save more than two brothers, or four half-brothers, or eight first cousins..." (16). This new concept has remarkably enhanced the explanatory power of Darwinism, especially in the analysis of "altruistic" behaviors of social insects (Hamilton 1964b).

Hamilton's paper suggests interesting points about the introduction of the concept of inclusive fitness. First, he actually

introduced two notions of fitness, namely "neighbour modulated fitness" and "inclusive fitness." The former is almost ignored (even by Hamilton himself) and the latter has prevailed and became a central notion of Darwinism. Where did the difference come from? Dawkins points out these two notions have equivalent results when properly used (Dawkins 1982, 187). So the empirical data cannot discriminate between these two notions. The major difference is usefulness in calculation (this seems to be the major reason Hamilton preferred inclusive fitness). Inclusive fitness is also easier to understand intuitively. For example, if we reconstruct the idealized model in terms of the neighbour modulated fitness, it will go as follows; "if I am drowning by myself, I cannot expect someone to sacrifice his/her life to save me, but if I am drowning with two other brothers, I can expect another brother to come to save..."⁷ This does not seem to be a good way to explain the situation. If Hamilton's paper had been written in this way, the influence of the paper, i. e., the conceptual inclusive fitness of the paper (in Hull's sense --do not confuse this with Hamilton's own usage of inclusive fitness), would have decreased.

Another interesting point is the conceptual problem Hamilton struggled with. He was working on the behaviors of social insects, and he found that some of these behaviors are hard to explain by natural selection. This is an empirical anomaly for Darwinism,

⁷ Another way to put it; if the risk to lose one's own life to save me is 100%, no one will help me; if the risk is less than 50%, then my brother will help me; if the risk is less than 12.5%, my cousin will help me, and so on. This seems to be a more

and he got a solution by calculating the kinship relationship. But why did he introduce a new concept to characterize the solution? We cannot know the exact reason, but we can speculate. Hamilton's result shows the fittest in the orthodox sense is not necessarily selected by natural selection. On the other hand, the notion of fitness is too convenient to give up. So, Hamilton had to introduce a new concept which is as convenient as fitness, and has a direct relation to natural selection. Needless to say, inclusive fitness meets these requirements, and I think this is why Hamilton introduced the word.

6.2 Dawkins's gene's eye view

Hamilton's "inclusive fitness" influenced many biologists. Richard Dawkins is among them. Eventually Dawkins refuses to use this word as a part of his theory, but it is obvious that his "selfish gene" view is a response to the conceptual problem which "inclusive fitness" caused (Dawkins 1989, 1982). According to Dawkins, the unit of the natural selection is not an individual organism, but a gene. He nicely explains biological phenomena from the "gene's eye view" (Dawkins 1989, ix). From this point of view, the orthodox notion of fitness of an individual organism is erroneous (Dawkins 1989, 137). How about inclusive fitness? According to him, inclusive fitness "was technically correct, but complicated and easy to misunderstand" (Dawkins 1982, 194).

accurate explanation of the neighbour modulated fitness, but anyway this explanation is not very attractive

He explains why "inclusive fitness" is misleading (Dawkins 1982, 190). Suppose I have two brothers, A and B. I want to increase my inclusive fitness. I find brother A has more similarities to me than B has. This seems to suggest that A has more genes in common with me. Now, if I support A more than B, does this action increase my inclusive fitness? From individual organism's point of view, the answer seems to be yes. But if we want to keep the neat connection between inclusive fitness and selection, the answer should be no. The genes for facial appearances and the genes for the action to support a brother are independently assorted by Mendel's second law (if they are not on the same chromosome).⁸ This gene for the behavior has the same chance to be in A and in B, regardless of the facial appearance. Thus, the gene to support A instead of B is not selected in natural selection.

I would like to add another paradox which arises when we uncritically talk about an organism's inclusive fitness (to my knowledge, no one had pointed out this paradox before). In a common sense account of inclusive fitness, when I save some people at the cost of my own life, if the total amount of genes they share with me is larger than the amount of genes I have, this behavior increases my inclusive fitness. Now, 99% of human genome are the same in every human being. Therefore, if two

⁸ Dawkins has a nice imaginative example (Dawkins 1989, 89). If a gene for green beardness and a gene to help someone who has a green beard are closely linked, this link can be selected. But usually we cannot assume such a linkage, so the argument here is still practically valid

strangers are drowning, since they have at least 198% of my genes, to save them at the cost of my life should increase my inclusive fitness! Why does this kind of super altruistic genes not evolve? The fallacy is, again, to forget the gene's eye view. For the gene for such a behavior, the amount of shared genes between organisms is irrelevant. When such a super altruistic gene is newly created by mutation, it is among the remaining 1%, thus the probability that a stranger has the same gene is almost 0. Therefore, the super altruistic behavior decreases inclusive fitness of the genotype which includes the gene.

As these considerations suggest, if we restrict inclusive fitness to a genotype ij as Hamilton originally defined, we can avoid these mistakes. But once we start to talk about an organism's inclusive fitness (and Hamilton himself admitted this usage), to commit a fallacy is too easy. And here is the conceptual problem Dawkins tried to solve. When we used the orthodox concept of fitness, we could talk about the fitness of an individual organism and the fitness of a genotype interchangeably. But once we start to think in terms of inclusive fitness, inclusive fitness of a genotype ij is a much clearer notion than that of individual organisms. The latter caused miscalculation even by Hamilton himself (Dawkins 1982, 191-192; see also Hamilton 1964b, 30-31). So the problem Dawkins struggled with was to reconstruct the whole theory in an intuitively understandable way so that we can avoid mistakes. As a solution to this problem, Dawkins adopted the gene's eye view, and abandoned the notion of "fitness" itself.

But here is another interesting fact. He occasionally talks about "survival value" of replicators (Dawkins 1989, 199-200). As we saw before, this is one of the words population geneticists introduced in place of the "fitness." In Dawkins's usage, "survival value" roughly means "advantages at the gene level" (Dawkins 1989, 200). This is almost the notion of "fitness" except that it is a property of genes, not organisms or genotypes. Therefore, he abolishes the word "fitness" at the individual organismal (and genotypic) level, but he still exploits the convenience of the notion at the gene level.

7. Conclusions

This historical overview suggests many interesting relationships between concept changes and theoretical and social factors. The notion of fitness went through various changes in its meaning (see the tables below). What were the causes of these changes?

First, there are many cases in which Hull's account can be applied. Spencer's word choice of "fitness" increased his conceptual inclusive fitness because not only biologists but also social Darwinists used the word. Wallace and Darwin had also a good reason to adopt the word. They thought this word would help in accurate communication of the theory. Later, population geneticists still used the word, but they tried to give it an exact meaning, and sometimes they tried to stop using it. Perhaps this

move was motivated by the consideration that the association with Social Darwinism decreases the conceptual inclusive fitness of biology. Hamilton's introduction of the inclusive fitness gives us another example of inclusive conceptual fitness. He had an enough reason to choose inclusive fitness instead of neighbour modulated fitness, but at the same time this choice seems to increase his conceptual inclusive fitness.

Secondly, these cases and many other cases are also examples of conceptual problems and their solutions. For Wallace and Darwin, the adequacy of the notion of "natural selection" was the problem. They tried to reduce the use of a metaphorical expression by introducing the "survival of the fittest." The problem for population geneticists was to clarify the notion. The propensity interpretation of the fitness was addressed to solve a conceptual problem, i. e., the tautology problem. Hamilton's problem was to solve the anomaly about social behavior without losing the convenient notion of fitness. Dawkins tried to answer the problem that the notion of inclusive fitness caused -- namely the misleading characteristics of the inclusive fitness of an individual organism. His answer was to stop thinking at organismal level, and stop using the notion of fitness itself.

When we analyze the history of conceptual change, both of these two kinds of factors are indispensable to explain the changes. Social factors are not enough to explain the answer the scientists chose. Internal factors are not enough to explain why a particular solution was chosen instead of other possible answers.

The history of the concept of "fitness" seems to exemplify this claim⁹.

⁹ I owe a lot to Prof. Darden and discussions in her seminar. Prof. Darden also gave me many helpful comments on earlier versions of this paper.

Tables --- components of notion of "fitness"

general description

1. definition
2. of what?
3. measurement
4. factors that act on it
5. importance of the notion in the theory
6. other comments

A. Darwin's "fitness" in the first edition of the *Origin*

1. definition none (synonymous to "adaptation")
2. of what? individual organism
3. measurement ?
4. factors natural selection
5. importance not important in his theory
6. comments no direct relation to later usages of the word.

B. Spencer's "fitness"

1. definition "survival of the fittest" = natural selection
2. of what? individual organism
3. measurement ?
4. factors inheritable variations and environment
5. importance central word for his version of Darwinism
6. comments First usage of modern sense of "fitness"

C. "Fitness" in social Darwinism

1. definition ?
2. of what? individual human being
3. measurement economic success or moral superiority or physical and intellectual superiority
4. factors inheritance, relation to other human beings
5. importance normative connotation of "fitness" justified their political claims
6. comments measurable independent from environment

D. "Fitness" or "adaptive value" in population genetics

1. definition ?
2. of what? genotype (secondarily applicable to individual organism)
3. measurement average reproductive success
4. factors environment
5. importance one of theoretical notions
6. comments first scientific measurement of fitness

E. The tautological definition of "fitness"

1. definition actual success in survival and reproduction
2. of what? individual organism
3. measurement success in survival and reproduction
4. factors genotype, environment and luck
5. importance supposed to be a central notion of the theory
6. comments a careless definition

F. The propensity interpretation of "fitness"

1. definition propensity to success in survival and reproduction
2. of what? individual organism (secondarily applicable to type)
3. measurement equivalent to D
4. factors genotype and environment
5. importance central to avoid tautology
6. comments reasonable solution to the tautology problem

G. Engineer's view definition of fitness by Gould

1. definition better design is fitter
2. of what? individual organism (part of an organism?)
3. measurement engineer's criteria
4. factors genotype and environment
5. importance central to avoid tautology
6. comments

H. "Inclusive fitness" by Hamilton

1. definition expectation of an organism's survival and reproductive success
+expectation of its effect to relatives' survival and reproductive success
2. of what? genotype of the organism (also applicable to individual organism
+particular genes)
3. measurement same as 1.
4. factors social behavior associated with the genotype, genetic structure of relatives, etc.
5. importance core notion of his analysis of social behavior
6. comments enhanced explanatory power of Darwinism

I. "Neighbour modulated fitness" by Hamilton

1. definition ?
2. of what? individual organism
3. measurement expectation of the organism's survival and reproduction when we take account of the effect from neighbours.
4. factors genotype of the individual, genotype of neighbours and environment
5. importance almost ignored by author himself
6. comments empirically equivalent to inclusive fitness

J. "Survival value" in Dawkins

1. definition ?
2. of what? replicators (gene, meme)
3. measurement success in leaving copies
4. factors behavior of other genes in the same population
5. importance used occasionally
6. comments

References

- Beatty, J. (1994). Fitness: theoretical contexts. in (Keller and Lloyd 1994), 115-119.
- Bowler, P. J. (1989). *Evolution: the history of an idea*, revised edition. Berkeley: University of California Press.
- Burian, R. M. (1994). Adaptation: historical perspectives. in (Keller and Lloyd 1994), 7-12.
- Cain, J. A. and Darden, L. (1988). Hull and selection. *Biology and Philosophy* 3, 165-171.
- Clark, L. L. (1984). *Social Darwinism in France*. Alabama: University of Alabama Press.
- Darden, L. (1991). *Theory change in science: strategies from Mendelian genetics*, New York: Oxford University Press.
- Darwin, C. R. (1859). *On the origin of species by means of natural selection or the preservation of favoured races in the struggle for life* first edition, reprinted in 1964. Cambridge: Harvard University Press.
- (1866). Letter to A. R. Wallace dated 5 July. in (Marchant 1916), 144-145.
- (1872). *The origin of species by means of natural selection or the preservation of favoured races in the struggle for life* sixth edition, reprinted in two volumes in 1898. New York: D. Appleton and Company.
- Dawkins, R. (1982). *Extended phenotype*, Oxford: W. H. Freeman and Co.

- (1989). *The selfish gene*, new edition. New York: Oxford University Press.
- Dobzhansky, T. (1937). *Genetics and the origin of species*, reprinted in 1982. New York: Columbia University Press.
- (1955). *Evolution, genetics and man*. New York: John Wiley & Sons Inc.
- Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford: Clarendon Press.
- Gould, S. J. (1977). *Ever since Darwin: reflections in natural history*. New York: W. W. Norton & Co.
- Haldane, J. B. S. (1932). *The causes of evolution*. London: Longmans, Green and Co.
- (1938). *Heredity and Politics*. New York: W.W. Norton & Co.
- Haller, M. H. (1963). *Eugenics: hereditarian attitudes in American thought*. New Brunswick: Rutgers University Press.
- Hamilton, W. D. (1964a). The genetical evolution of social behaviour I. *Journal of Theoretical Biology* 7, 1-16.
- (1964b). The genetical evolution of social behaviour II. *Journal of Theoretical Biology* 7, 17-32.
- Hobhouse, L. T. (1893). *The labour movement*. Fisher Unwin.
- Hull, D. L. (1988a). *Science as a process: an evolutionary account of the social and conceptual development of science*, Chicago: University of Chicago Press.
- (1988b). A mechanism and its metaphysics: an evolutionary account of the social and conceptual development of science. *Biology and Philosophy* 3, 123-155.
- Jones, G. (1980). *Social Darwinism and English thought: the interaction between biological and social theory*. New Jersey: Humanities Press.
- Keller, E. F. (1994) Fitness: reproductive ambiguities. in (Keller and Lloyd 1994), 120-121.

- Keller, E. F. and Lloyd, E. (eds.). (1994). *Keywords in evolutionary biology*, Cambridge: Harvard University Press.
- Kevles, D. J. (1985). *In the name of eugenics: genetics and the uses of human heredity*. New York: Alfred A. Knopf.
- Kimbrough, S. O. (1980). The concept of fitness and selection in evolutionary biology. *Journal of Social and Biological Structures* 3, 149-170.
- Marchant, J. (1916). *Alfred Russel Wallace: letters and reminiscences*. New York: Harper and Brothers.
- Mills, S. K. and Beatty, J. H. (1979). The propensity interpretation of fitness. *Philosophy of Science* 46, 263-286.
- Paul, D. (1994). Fitness: historical perspectives. in (Keller and Lloyd 1994), 112-114.
- Salmon, W. (1967). *The foundations of scientific inference*. Pittsburgh: University of Pittsburgh press.
- Sinnott, E. W., Dunn, L. C. and Dobzhansky, T. (1958). *Principles of genetics*, fifth edition. New York: McGraw-Hill Book Company.
- Spencer, H. (1864). *The principles of biology*, vol. 1. London: Williams and Norgate.
- (1994). *Political writings*, John Offer (ed.). Cambridge University Press.
- Waddington, C. H. (1939). *An introduction to modern genetics*. New York: Macmillan Company.
- (1957). *The strategy of the genes*. New York: Macmillan Company.
- Wallace, A. R. (1866). Letter to Charles Darwin dated 2 July. in (Marchant 1916), 140-143.
- Wimsatt, W. C. (1980). Reductionistic research strategies and their biases in the units of selection controversy. in *Scientific Discovery: Case Studies*, T. Nickles (ed.). Dordrecht: D. Reidel Publishing Company. 213-259.

Zmarzlik, H. (1972). Social Darwinism in Germany, seen as a historical problem, in *Republic to Reich: the Making of the Nazi Revolution*. Hajo Holborn (ed.). New York: Random House.