University of Tartu
Institute of Philosophy and Semiotics

IS ‘FITNESS’ A PRIMITIVE OR A PROPENSITY?
DIAGNOSING THE ROLE OF EXPLANATORY REDUCTIONISM ON DIFFERING STANDARDS OF SCIENTIFIC DEFINITIONS

Master’s Thesis in Philosophy

Anna Elise Rohtmets

Supervisors: Edit Talpsepp PhD
Riin Kõiv MA

Tartu 2021

TABLE OF CONTENTS

INTRODUCTION

The subject of this thesis is the diagnosis of the disagreement between the two earliest attempts to explain the meaning and explanatory role of evolutionary fitness in the theory of natural selection.

During the 1960s, the theory of evolution was under philosophical scrutiny for its alleged tautological explanations and inability to make testable claims (Manser 1965; Popper 2005; Williams 1970). Gradually, these critiques were debated and solved in the new field of philosophy of biology, which began to emerge in the beginning of the second half of the 20th century (Odenbaugh, Griffiths 2020). However, the criticism that fitness is defined in a manner that renders explanations in evolutionary biology circular, named in this thesis the charge of circularity, sparked a series of debates that last until present day.

The series of debates have centred around the alleged circularity of the theory of natural selection, its core insight captured by the phrase 'survival of the fittest'. The charge of circularity goes as follows: fitness is defined by biologists as reproductive success, which is why it cannot explain reproductive success and by extension the survival of reproductively successful organisms throughout history. 'Survival of the fittest' would translate into 'survival of those who survive' and therefore the central insight of Charles Darwin would turn out to be a mere tautology parading as science.

In the early phases of the fitness debate, two solutions to the charge were presented. The first solution, the propensity interpretation, is to interpret fitness as a propensity for organisms to survive and reproduce (Mills, Beatty 1979; Beatty, Brandon 1984). It breaks the circularity by claiming fitness causes survival – 'survival of the fittest' means that those with the greatest ability to survive, will most likely do so. The second solution, the primitivist interpretation (Williams 1973; Rosenberg 1982, 1983; Williams, Rosenberg 1986), holds that fitness has no definition in the theory of natural selection that could solve the charge and finds the charge itself to be misguided. The articles under discussion were written during 1965 to 1986 and are an example of the kind of philosophy of biology that deals with conceptual problems found within the theory of evolution.

In recent years, philosophers of biology have shown interest in the history of this debate (Otsuka 2016), (Elgin, Sober 2017). My interest is also historical, and I chose to study an early period of this debate which is the disagreement between the propensity and the primitivist interpretation of fitness. I argue that the difference between the propensity interpretation of fitness and the primitivist interpretation of fitness stems from their different understanding of what a successful definition should achieve. According to the proponents of primitivism, a definition of fitness that solves the charge of circularity should provide a reductive account of what fitness is within the theory of natural selection. According to the proponents of the propensity interpretation, a definition of fitness that solves the charge of circularity need not provide a reductive account within the theory. Therefore, the decision which account to favour should be based on whose account of definitions is correct, instead of which interpretation has a more correct understanding of fitness.

This thesis is divided into three chapters. In the first chapter, I will introduce the Inductive-Statistical Model of scientific explanations, explain how biologists understand what fitness and natural selection are, and explicate the charge of circularity. I will also differentiate between the term ‘fitness’ and its referent the property fitness. In the second chapter, I will describe the propensity and the primitivist interpretations. In the third chapter, I will first argue that both interpretations share an understanding of what the referent of ‘fitness’ is, and that their disagreement lies in how to successfully define fitness to dispel the charge of circularity. Finally, I will argue that the disagreement stems from their differing commitment to explanatory reduction.

1. THE TROUBLE WITH FITNESS

The two competing interpretations of fitness I will be discussing in this thesis are examples of philosophy of biology that deal with conceptual problems in the theory of evolution. Following their arguments requires knowledge of evolutionary theory. Therefore, the aim of this chapter is to introduce key evolutionary concepts as well as some theoretical background of philosophy of biology. I will first acquaint the reader with Hempel's model of inductive explanations which appears both in critiques and defences of fitness explanations. Next, I will introduce the theory of natural selection and the concept of fitness as they appear in biology. After I have introduced to the reader how biologists have understood these terms and how they are logically related, I will introduce the criticism that fitness explanations are circular.

1.1. The Inductive-Statistical Model of Scientific Explanations

Philosophers of science have tried to explicate and model what they think are legitimate scientific explanations, thus demarcating descriptions from explanations and scientific explanations from non-scientific ones. For decades, the most influential model for scientific explanations was the Deductive-Nomological Model (DNM), sometimes called the Popper-Hempel Model (Woodward 2019). According to the DNM, an explanation is scientific when it clearly shows how the statement that we wish to explain, the *explanandum*, is deductively derived from other statements, the *explanans* (Hempel 1965: 247). The statements in the *explanans* should be true empirical observation statements and feature at least one law of nature (ibid.: 247-248).

Relevant for the forthcoming discussions about fitness is Hempel's modification of the DNM to include the Inductive-Statistical Model (ISM) where the *explanandum* is inductively derived from the *explanans*, which must feature at least one statistical law and contain empirical observation statements (ibid.: 412). Here is an example:

$$1) P(B) = 93\%$$

$$2) B$$

$$3) Therefore: A\ will\ very\ likely\ occur.$$

1) The efficacy of the antibiotic amoxicillin to treat sinus infections is 93%, i.e., the chance (P) to recover from a sinus infection (A) given that there is treatment with amoxicillin (B) is 93%.
2) Amy had a sinus infection for which she was prescribed amoxicillin to take twice daily for 7 days.
3) Therefore, she will very likely make a full recovery.

Statements 1) and 2) constitute the *explanans*, where 1) is the statistical law and 2) is the observation statement. Statement 3) is the *explanandum*.

Hempel's model has philosophical ramifications that are touched on throughout this thesis. Firstly, it sets in place the relation between causation and explanation: when only A causes B, A explains and predicts B. It is important to keep in mind that Hempel's models of explanations do not mirror scientific methodology. Causation is taken as a relation between observed events that is confirmed experimentally. When a causal chain has been discovered and described, this description enables the formation of explanations and predictions. In Hempel's model, the *explanans* explaining and predicting the *explanandum* is the same thing – the explanatory power of science comes from its ability to make predictions.

When fitness explanations like the kind presented in the last section were called under question, the golden standard of science was still physics (Mayr 1969), and the philosophy of scientific explanations was largely regarded as finished by Carl Hempel. Wesley Salmon has written that the DNM and its modifications have shaped philosophy of science in the latter half of the 20th century and called the original paper by Hempel and Oppenheim in 1948 "truly epoch making" (Salmon 2006: 3). The model peaked in popularity in the first half of the 1960s. Labelled as the received view, later critiques and alternative models were framed either for or against the DNM and its supplements, but always in relation to it . It is not an overstatement according to Salmon that the Hempelian Standard was almost the consensus at the time critiques of evolutionary biology and its concepts, including fitness, were written. (ibid.: 3-4)

1.2. Natural Selection and Fitness in the Theory of Evolution

1.2.1. The Significance and Meaning of Natural Selection and Fitness

Natural selection is the process of evolution where certain traits are selected by the environment and carriers of these traits survive and pass on these traits. It is one of the most important ideas in the theory of evolution because it has been regarded as the most relevant mechanism that brings about evolutionary change.[1] Despite being very closely related, natural selection and evolution are not synonyms. Charles Darwin made two distinct claims. First, he defined evolution as descent with modification from a common ancestor (Darwin 1859: 123-124; 331). This statement also holds true today (Futuyma et al. 2017: 7-8). Second, he believed that "natural selection has been the main but not exclusive means of modification" (Darwin 1859: 6). Evolution does not inevitably presuppose any certain mechanism. Change can happen through many different and subtle processes which is why

[1] For discussion on the extent of natural selection's role in evolution, see Orzack, Steven Hecht and Patrick Forber 2017.

Darwin's idea of natural selection as the chief mechanism of evolution is logically distinct from his idea of descent from a common ancestor. Due to natural selection being Darwin's main insight to evolution, the word 'Darwinism' is used to specifically mean evolution by natural selection.

The theory of natural selection is understood as a set of propositions that are used to describe and explain the distribution of favourable traits of organisms in a population. 'Trait' in biology commonly refers to a discreet observable property of the organism such as wings or a snout[2,] and 'favourable trait' refers to a trait that gives an organism a reproductive edge over those who do not possess it. Sometimes philosophers of biology talk about the principle of natural selection, the law-like phrase that carries the 'core Darwinian insight' of the theory of natural selection, most famously in the form of 'survival of the fittest'. The principle, taken to be a component of the theory, conveys the prediction that those who are fittest will survive. Therefore, in order to understand the theory of natural selection and thereby Darwin's core insight, we need to understand what fitness is.

In contemporary biology, the definition of fitness is often presented in conjunction with the definition of natural selection. The logical relationship between the two is exemplified in the widely used Douglas J. Futuyma and Mark Kirkpatrick's textbook on evolution:

"Natural selection is any constant difference in fitness among different classes of biological entities," where fitness is defined as "/../the number of offspring an individual leaves in the next generation, /../sometimes called reproductive success" (ibid. 2017: 60).

This definition of fitness has been prevalent both in academic articles and textbooks since the Modern Synthesis[3]. It is also the definition of fitness that sparked criticism. It is important to distinguish 'fitness' the term as it is used in explanations and theories from fitness, its referent in nature.

'Fitness' is an indispensable component of the theory of natural selection which describes when evolution is driven by selection. Evolution by natural selection has three necessary and sufficient conditions: heritability of traits, variation in traits, and differences in fitness due to traits (Endler: 1986: 6).[4] Darwin observed that organisms are like their parents in that certain traits are passed on from generation to generation. He also observed that organisms

---

[2] The total set of observable traits is called a phenotype (Futuyma et al.: 18).

[3] Also called Neo-Darwinism, the Modern Synthesis marks the unification of Mendelian genetics with Darwinian evolution and the start of modern evolutionary biology.

[4] There are various formulations of Darwin's conditions for natural selection which can be found in the chapter "Recapitulation" (Darwin 1859). I based my list of three conditions on ethologist John Endler's book on fitness (1986), but perhaps most famously the conditions for natural selection have been formulated by geneticist Richard Lewontin (1970).

have a variety of different traits, as no one looks exactly the same. Lastly, he observed that not all traits are equally beneficial for survival and reproduction. (Darwin 1859) The fact that some types of organisms systematically throughout generations fare better than others is evidence that they possess traits the struggling organisms do not. Natural selection is only present if there is some selection between favourable and unfavourable traits that is indicated by fitness differences.

Fitness[5] cannot be separated from natural selection because differences in reproductive success between organisms in a given environment indicate natural selection at work. If there are no differences in fitness in a population, natural selection is not present, but differences in fitness can occur without natural selection, like where they are the result of rarely occurring natural catastrophes; or like in the case of genetic drift, where reproductive success is random and independent of traits that have evolved to facilitate living in a given environment. So, we know for certain that natural selection is not shaping evolution where fitness differences do not occur, but natural selection can be one of the means of evolution when fitness differences do occur. If fitness differences occur over a long period of many generations when the heritability and trait variation conditions are also met, natural selection is said to be present and causing evolutionary change.

1.2.2. Fitness Explanations

'Fitness' also figures in many explanations in biology which have been charged with being circular. I will call these explanations fitness explanations and they are scientific explanations in evolutionary biology where differences in fitness explain differences in reproductive rates. These explanations are used to further argue for the presence of evolution by natural selection of a given population. Based on Hempel's ISM, 'fitness' is one empirical element of the *explanans,* and reproductive rates are the *explanandum*.[6] In the *explanans*, fitness is conceptualized with "a mathematical representation that allows predictions and explanations to be formulated" (Sober 2001: 1). This is because fitness, though defined as reproductive success, may be mathematically represented in various ways. Sometimes, fitness is mathematically represented as the organism's number of offspring relative to the population's average number of offspring, over a specified number of generations.

[5] Throughout biological and philosophical literature, 'fitness' is sometimes equated with 'adaptation', 'adaptability' or 'adaptedness' - especially in older texts.

[6] 'Fitness explanations' is synonymous with 'natural selectionist explanations', but I will use the former for clarity, because the focus of this thesis is on evolutionary fitness and its interpretations.

Sometimes, further calculations are made to sort individual organisms into groups and then weigh their average fitness over the average fitness of the population.

To better grasp what fitness explanations are and the charge that they are circular, I will present a Hempelian formulation of a textbook fitness explanation, where fitness is mathematically represented as the *average* number of offspring individual organisms of a species contribute to the population each generation (Day, Otto 2001: 4). Let there be two species of bacteria A and B. Both live in environments where there is a fluctuation of wet and dry periods. The fitness of A in a wet environment is 2 and in a dry environment 0.5. The fitness of B is 1.5 and 1, respectively. Starting populations, marked T0, are 100 organisms for both A and B.

What is the predicted population of A and B in two generations T2? The answer is provided by statement 5., which statements 1.-4. are meant to explain.

1. T0(A)=100; T0(B)=100 The population is 100 individuals in the starting generation T0.
2. f_{wet}(A)=2; f_{wet}(B)=1.5 Fitness of A and B in the wet environment.
3. T1(A)=200; T1(B)=150 The population is $f_{wet} \times$ T0 in the generation T1.
4. f_{dry}(A)=0.5; f_{dry}(B)=1 Fitness of A and B in the dry environment.

----

5. T2(A)=100; T2(B)= 150 The population is $f_{dry} \times$ T1 in the generation T2.[7]

In this example, the difference in fitness of A and B in the dry and wet environment is used to explain the difference in reproductive rates of A and B in T0. 1.-4. were the *explanans* and 5. the *explanandum*. Statement 5. reads that the predicted population of A and B is 100 and 150 individuals, respectively. As the cycle repeats[8], B has a starting advantage of 50 organisms. Over time there will be a steady increase in species B.

Remembering Hempel's model, when fitness explains reproductive rates, it is also posited as a cause for reproductive rates. When fitness causes differences in reproductive rates over many generations, it indicates natural selection at work. Therefore, conceptual problems with fitness have long been taken seriously. If something as important in biology as the idea of natural selection turned out to be logically reliant on a faulty concept like fitness, then the

[7] Hempel's schema requires at least one law-like statement in the *explanans*. In this example the Hardy-Weinberg law is already assumed according to which in the absence of any evolutionary influences the frequency of genes in the population will stay the same.

[8] It does not matter if we start with the dry or wet environment; the result over two generations will be the same.

legitimacy of the theory of natural selection would also be under question – and by extension most of evolutionary biology.

1.3. The Charge of Circularity

The charge of circularity is the charge that the concept of fitness has a definition that renders explanations like the one previously presented circular. Presentations of the charge are scattered between different authors and articles both in and out of academia and there is no single clearly formulated version that the philosophers trying to solve the problem cite. For my thesis I have chosen to introduce the problem based on the critiques made by A.R. Manser (1965) and Karl Popper (2005).

What does it mean for explanations to be circular? Circular reasoning is a form of begging the question where propositions in the premise and the conclusion have only the support of each other. René Descartes exemplified this type of reasoning with the following example:

"/../and although it is absolutely true that we must believe that there is a God, because we are so taught in the Holy Scriptures, and, on the other hand, that we must believe the Holy Scriptures because they come from God /../, we nevertheless could not place this argument before infidels, who might accuse us of reasoning in a circle" (Descartes 2003: 55).

Circular reasoning can occur in a logically valid argument, because if the individual premises are true, the conclusion must also be true. Instead of being a formal logical fallacy, circular reasoning causes a pragmatic failure to persuade and explain, because the premises do not provide evidence for the conclusion. In other words, to accept the premise, we already need to accept the conclusion.

Manser's critique is that if fitness is defined as the number of offspring an individual leaves in the next generation (or reproductive success), then fitness cannot explain reproductive success and fitness explanations are therefore not real scientific explanations. The result is that these kinds of explanations fail to demonstrate that natural selection is present and causes evolutionary change. Consider this simplified dialogue meant to demonstrate the circularity of fitness explanations:

A: There are more black-coloured peppered moths in Birmingham than light ones.

B: Why?

A: Because the black moths are fitter.

B: How do we know that?

A: Given that the black-coloured moths clearly have a higher reproductive success, as there are more of them, they are fitter.

A’s explanation is circular, because she expounds the greater reproductive success of black-coloured moths to them having greater fitness, but then uses reproductive success to explain why she attributes them greater fitness. This happens because fitness and reproductive success are synonyms.

The aim of this example is to introduce the underlying logic of A.R. Manser’s critique of one of the most famous papers in evolutionary biology: H. B. D. Kettlewell’s article on the industrial melanism[9] of the peppered moths[10]. (Kettlewell 1958) In his paper, Kettlewell first states that there has been a surge of darker varieties of moths in Britain’s industrialized areas and his aim is to test the hypothesis that the darker, i.e., melanic, forms are fitter due to selective predation (ibid.: 323-324). In a controlled environment, he observes that on the darkened tree trunks of Birmingham, birds were able to capture the lighter variations of the moth more easily. A clear statistical difference emerged in predation rates that confirmed the hypothesis that the darker forms have better camouflage and therefore a selective advantage over the lighter forms (ibid.: 339). In the summary section, he concludes that the effect on natural selection on his tested variety of moth cannot be denied (ibid.: 341).

Manser, relying on Hempel’s model, first argues from a methodological standpoint that Kettlewell’s fitness explanation is operationally[11] circular, because to know fitness levels that are in the *explanans,* which are supposed to explain and predict reproductive rates in the *explanandum*, we must first assume the reproductive rates that are to be predicted in the *explanandum*. His take is that only *after* observing the greater survival in the melanic moths throughout generations do biologists proclaim this variation to be fit: “Survival [and reproduction] and fitness are necessarily connected” (Manser 1965: 26). Fitness, or ‘adaptability’, as Manser phrased it, has no independent epistemic criterion other than looking at actual survival rates of organisms and stating in hindsight that those creatures who survived and reproduced more must have been fit. “We cannot use the explanation to predict what will happen if a new change happens in the environment or what changes could happen in another species in the same environment” (ibid.: 25-26).

Manser’s critique has also a semantic side: Kettlewell’s explanation of natural selection causing the evolution of the peppered moth “is only a description with slightly theory-laden

[9] Industrial melanism is a dark pigmentation occurring in arthropods that develops due to industrial pollution in the environment.

[10] Manser chose to critique this paper because it was and remains one of the most famous cases of proof of natural selection. To critique this paper successfully would have been to critique one of the best and well-regarded examples of natural selection.

[11] ‘Operationally’ means here that there is no independent practice or operation to measure fitness besides reproductive rates.

words" (Manser 1965: 25). Using fitness rates to predict reproductive success becomes a simple induction from earlier reproductive success to future reproductive rates, because fitness and reproductive rates are the same thing in biology. This is describing changes in the population. For Manser, 'explaining' means to "enable events to be deduced from a set of initial conditions together with universal laws, in the way that physics and chemistry explain within their respective fields" (ibid.: 31). For Manser inductive explanations are not simple inductions from past rates to future rates, because the past rates do not *cause* the future rates. For fitness to explain reproductive rates, fitness needs to be a conceptually separate property from reproductive rates that causes reproductive rates.

Originally published in 1974, Karl Popper critiqued the theory of evolution being not a true theory, but a metaphysical research programme (Popper 2005). In addition to Popper arguing that "Darwinism [theory of natural selection] does not really *predict* the evolution of variety," and "it therefore cannot really *explain* it" (ibid: 199), he also critiqued the concept of fitness, by making the example that selection is defined in such a way that if a species has died out, biologists say it was not fit, and if selected, it is necessarily fit (ibid.). Fitness could not explain and predict reproductive rates and by extension the survival of different types of organisms, because fitness is defined and measured by reproductive rates. Because fitness and reproductive rates have an analytic definitional connection, not a causal one, it is impossible to test the claim whether those who are fit, did really survive. Similarly to Manser, Popper requires the theory of natural selection to predict the variety of the forms of life which would succeed in a given environment. Instead, the theory provides a circular explanation why those who survive are fit and vice versa.

Popper and Manser critique Darwin's central idea informally and without explicitly stating what content fitness should possess to escape the charge of circularity. However, both critics argue that fitness should possess independent content apart from actual survival rates, as well as have some independent criterion for testability. When 'fitness' and 'reproductive rate' cease to have the same referent, fitness explanations would no longer be mere restatements from present reproductive rates to future ones.

2. TWO SOLUTIONS FOR THE CIRCULARITY PROBLEM

In the previous chapter, I described the central importance of fitness and the theory of natural selection to biology. Fitness is at the heart of the theory of natural selection but charged by critics to possess a definition that renders fitness explanations circular and therefore not scientific. This prompted biology-friendly philosophers of science to look for ways to escape the charge and defend the validity of Darwinism. A strategy for doing this was to deny that fitness is synonymous with reproductive rates and provide a new interpretation of its meaning that would dispel the charge. Two such accounts of fitness gained prominence: the propensity interpretation and the primitivist interpretation of fitness. In this chapter, I will present the view that interprets fitness as a propensity to reproduce. Second, I will describe the view according to which fitness is a primitive concept of the theory of natural selection.

2.1. The Propensity Interpretation of Fitness

Biologists have defined fitness as the number of offspring an individual leaves in the next generation, or simply as reproductive success. According to the propensity interpretation, biologists define fitness as actual reproductive success, but what biologists really mean is that fitness is potential reproductive success. The revised definition for fitness states that "the fitness of an organism x in an environment E is x's expected number of descendants in E" (Mills, Beatty 1979: 275).[12] The authors then demonstrate that under the new definition the scientific explanations in which fitness rates are meant to explain future reproductive rates are no longer circular, and therefore the circularity problem is solved (ibid.: 279-280). Under the old definition, natural selection occurs "when differential descendant contribution is caused by differential descendant contribution" (ibid.: 283). Under the propensity interpretation, natural selection occurs "when differential descendant contribution is caused by differential disposition to reproduce" (ibid.)[13] They argue that Manser was incorrect in his analysis of Kettlewell's paper, because Kettlewell did not claim that the moths who survived were fit because they survived, but because they had a higher expected number of descendants due to their *disposition* to reproduce. (ibid.: 282-283). Kettlewell's aim was to

[12] They also define type fitness, which is the average fitness of organisms possessing a given type (ibid.:277). It is derived from the definition of organismal fitness.

[13] The formal definition of the natural selection of organisms: "Natural selection is occurring in population P in environment E with regard to organisms x, y, z (members of P) =df x, y, z differ in their descendant contribution dispositions in E, and these differences are manifested in E in P" (ibid.: 283). A derived definition is also provided for types.

prove that this disposition had a causal basis in their melanism, which shielded them from predators in the industrial Birmingham's polluted environment.

To simply say that fitness is the expected number of descendants would not be enough. From where does this expected number come from? What is the referent of 'fitness'? Mills and Beatty (1979) observe that fitness is sometimes referred to by biologists as a trait, a property that organisms possess; and based on this observation they construe fitness as a disposition, an ability of organisms to survive and reproduce. Fitness, like solubility, is a tendency that does not necessarily manifest during an organism's life. They mark that "by saying something has the disposition to behave in a certain way, we are saying that there are physical properties that cause this disposition to manifest in the presence of triggering conditions" (ibid.: 270-271). Like sugar does not dissolve in oil due to its hydrophilic properties, so is a fish not able to fulfil its potential to reproduce out of water, because it is structurally unable to do so.

According to the propensity interpretation, when biologists measure fitness empirically by counting the offspring of individual organisms, they are measuring the organisms' propensity to reproduce. An example of measuring fitness in the field would be the counting and comparing of viable offspring of the female damselflies who laid their eggs to areas that flood in different capacities (Lambret et al.: 2018). After observing that the group of females who laid their eggs in an area that flooded faster left more viable offspring, the scientists concluded that they were fitter than the second group who chose drier areas. The propensity interpretation proponents argue that even though on the surface level it might very well seem that the referent of fitness is the amount of actual descendant contribution to the population, counting descendants is, in fact, a method to indirectly measure the ability to leave descendants, that is fitness. In other words, fitness is measured by its effects.

The notion of dispositions having some sort of causal basis, like a molecular structure in the case of solubility, is in contemporary metaphysics a view that has its proponents and is a relatively non-controversial one (Choi, Fara, 2018). Whether one is a realist or anti-realist[14] about dispositions, or whether one thinks the causal basis is categorical[15] or dispositional, philosophers have presented compelling arguments to regard dispositions and their causes as separate, i.e., dispositions do not cause themselves. Based on the philosophy done on dispositions, the idea of fitness as one is quite intuitive. An organism has properties that are

[14] For a discussion about their distinction see Mumford 1998.

[15] Categorical properties are thought to be properties that are always present, like shape and colour.

a result of their genetic make-up and environment. Some of those properties, when met with triggering conditions, either in the environment or perhaps some properties in the organism itself, cause the disposition to manifest in a certain way that can be observed and measured by scientists.

Prior, Pargetter and Jackson have defined a disposition's causal basis in the following way, where the characteristic stimulus they mention is a synonym for triggering conditions in the context of the present discussion:

"A causal basis for disposition *D* is the property or property-complex that, together with the characteristic stimulus of *D*, is a causally operative sufficient condition for the characteristic manifestation of *D* in the case of "surefire" dispositions, and in the case of probabilistic dispositions is causally sufficient for the relevant chance of the manifestation." (Prior et al.: 251)

Let us turn back to the female damselfly to illustrate what has been discussed so far. During the end of the year in certain climates there are hundreds of female spread-wing damselflies searching for places to lay their eggs. Every one of them is genetically unique, which necessarily brings about differences in their morphological and behavioural properties (also called a phenotype in biology). Those varying properties, biologists agree, are the causes for variation in fitness rates. This would be the causal basis for fitness. When triggering conditions in the environment are met, the disposition to reproduce actualizes. There exist also contingent major impact events in their environment, like forest fires, bulldozers, and rare heavy storms that would suddenly kill off a large part of the population or damage their respective habitat. Such events impact survival rates, but those rates would not say much about the damselflies' ability to reproduce in their habitat, as these freak events are not a stable part of said habitat. Remembering the definition, fitness in environment E is the expected number of offspring in E.

So far fitness has been discussed as an ability, but it is also construed as a propensity. There are unfathomably many interactions and causal chains paired with sheer luck scenarios that lead a damselfly to survive and produce a certain number of viable offspring, but they are causal chains nonetheless – one that is stochastic and probabilistic in its outcomes. In philosophy of statistics, probabilistically actualizing dispositions of individual entities (like fitness) are called propensities[16]. Propensities carry probability values from 0 to 1. A propensity is, for example, a disposition that has 0.7 chance of actualizing. Mills and Beatty

[16] In philosophy of statistics, "propensity interpretations regard probabilities as objective properties of entities in the real world. Probability is thought of as a physical propensity, or disposition, or tendency of a given type of physical situation to yield an outcome of a certain kind" (Hájek 2019). Traditionally, they are thought to be properties of single entities.

define propensities as dispositions of individual objects which makes a propensity a subtype of dispositions. (Mills, Beatty 1979: 272) Their motive to not only call fitness a disposition but specifically a propensity is meant to capture the "unexplicated notion that biologists refer to fitness as a phenotypic trait – i.e., a property of individuals" (ibid.). One consequence of this approach is that the fitness values of whole populations or types are all derivatives of individual fitness, because fitness cannot be a propensity if it is a property of classes or groups of objects[17].

In the context of the accusation that terms in the theory of natural selection only seem to refer to each other, not entities in nature, it would be a bonus for defendants of biology to show with emphasis that fitness refers to something that is 'real'. Therefore, Mills and Beatty have a metaphysical motive to construe fitness as a propensity: "Given propensities apply to individual objects, we also take them to be ontologically real - not merely epistemic properties" (ibid.). How to understand 'real' in this context? Propensities are thought of as real probabilities in the sense that they are a property of a physical system to behave a certain way with a certain frequency and as such can be experimentally validated. Propensities work in favour of an interpretation that takes fitness to be an observation independent property that can be scientifically measured, and which probabilistically causes observable outcomes.

Let us now see how the propensity interpretation is meant to avoid a circular fitness explanation. Here is the dialogue from the previous chapter that illustrated Manser's charge of circularity, but this time with the modification of the propensity interpretation.

A: There are more black-coloured peppered moths in Birmingham than light ones.

B: Why?

A: Because the black-coloured moths have a higher propensity to reproduce, which means they have a higher expected number of offspring (they are fitter).

B: How do we know that?

A: Kettlewell's experiment shows that a structural difference in their wings (melanism) shields them from predators, causing them to have a higher expected rate of offspring compared to the lighter moths.

In this modification, when A answers B why there are more dark coloured moths, she still explains it with their higher fitness, but uses the vocabulary of the propensity interpretation according to which fitness is defined as the organism's expected number of offspring. When asked to explain fitness, she does not fall back on reproductive rates, like in the previous

[17] For different interpretations of propensities for groups or abstract objects see Popper (1959) and Humphreys (1985). At the time when Mills and Beatty's article was published propensities were taken to refer to single entities only.

example, but explains their higher disposition to reproduce through its causal, or categorical base, which in this case is the structural difference in the wings of the darker moths that lets them evade predators.

The aim of defining fitness as the expected number of offspring is to first and foremost provide fitness with explanatory power (Beatty, Brandon 1984: 343). If fitness is understood merely as differential reproduction, then it cannot explain differential reproduction; but if fitness is construed as a disposition that causes differential reproduction, then fitness explains differential reproduction (Mills, Beatty 1979: 270). Propensities capture the sense of fitness as a property of individuals, and they give probabilities a realist and causal flavour.

2.2. 'Fitness' as a Primitive

In parallel with the discussion about the propensity interpretation, a rival primitivist interpretation of fitness was developed and defended by Mary Williams (1973; 1986) and Alex Rosenberg (1982; 1983; 1986). Like the propensity interpretation, the primitivist interpretation denies that fitness is reproductive success, but unlike the propensity interpretation, it does not provide its own definition of fitness to dispel the charge of circularity. Instead, the primitivist interpretation claims the term 'fitness', as it appears in the theory of natural selection[18], is a primitive term without a definition. However, the referent of 'fitness', i.e., fitness, is the degree of the relationship between an organism and its environment – a property with a complex causal base (Williams 1973) (Rosenberg 1982; 1983). The primitivist interpretation aims to solve the charge of circularity by showing how the term 'fitness', despite being a primitive in the theory, still possesses the necessary causal content to explain reproductive rates.

2.2.1. 'Fitness' as a Primitive Term in the Theory of Natural Selection

Based on their shared views on scientific theories, the proponents of the primitivist interpretation argue that 'fitness' is a primitive term. Rosenberg and Williams hold that "evolutionary theory is nothing more nor less than the core theory giving the mechanism of evolution by natural selection" (Williams, Rosenberg 1985: 739). Their view excludes all other biological models and theories that guide the scientific application of the theory of natural selection and its terms (ibid.) – a premise which is key to understanding their view

[18] The theory of natural selection could be taken as a subset of the theory of evolution, but often these two are equated. Alex Rosenberg and Mary Williams both refer specifically to the theory of natural selection when discussing fitness.

of fitness as well as their disagreement on the definition of fitness with the proponents of the propensity interpretation presented in the next chapter. The primitivist interpretation was mainly argued by Rosenberg and this thesis will concentrate on his views the most, but since he builds his argument on his shared premise with Williams, I will first explicate their view on the theory of natural selection.

Rosenberg and Williams view scientific theories as axiomatized systems of sentences[19]. To axiomatize a theory means to show that a total set of claims within a theory can be logically derived from a small beginning set of sentences or axioms. Theories feature terms that can be defined or primitive. Primitive terms appear in axioms that are the logical starting point for the set of sentences that comprise a theory. Primitive terms are not explicitly defined within the theory. Terms which are explicitly defined are *a priori* defined within the logical boundaries of the theory they appear in. For example, a defined term x that figures in the theory T cannot be defined with the theoretical vocabulary of another theory P, because the term x would then cease to be a defined term in the theory T. Any attempt to redefine x would have to be done within the axiomatic boundaries of T to show that it is a defined term in T; otherwise, it would be a primitive term with regards to T.[20]

Rosenberg and Williams hold that the leading explanatory principle of the theory of natural selection is that species evolve when nature selects the fittest organisms to survive and reproduce. William's axiomatization formally explicates this central claim with five axioms (ibid.: 1973). The boundary of the theory of natural selection is roughly marked by what is deducible from these five axioms. In the axiom where Williams explicates the law of natural selection as the "expansion of the fitter population", she introduces 'fitness' as a primitive term (ibid.: 88). From her axiomatization she concludes that 'fitness' is a key concept in the theory of natural selection, but the theory itself does not say much about its referent. This does not mean that fitness cannot be characterized with other theories, just that with respect to the theory of natural selection, it lacks a definition, and that 'fitness' is therefore a primitive of that theory.

One important consequence of Williams' and Rosenberg's view of the structure of the theory of natural selection is that they view the charge of circularity as the charge that the theory of natural selection itself is circular. For a theory to be circular "its leading explanatory principle" has to be circular (ibid.), and as previously mentioned, the leading

[19] Their view is called the syntactic view and it is primarily associated with logical empiricists.
[20] Further discussion in 3.1.

explanatory principle of the theory of natural selection is that species evolve when nature selects the fittest organisms to survive and reproduce. In this reading, the charge of circularity is the following: if fitness is the same as reproductive rates, then the leading explanatory principle of the theory of natural selection is that species evolve when nature selects those with highest reproductive rates to survive and reproduce – which is another way of saying that those who survive, survive.

The propensity interpretation understands the charge of circularity as the charge that fitness explanations are circular. This poses the question of whether the two interpretations of fitness are motivated by a different understanding of the charge. Rosenberg and Williams conclude that the difference is merely semantic, and both interpretations are trying to solve the same charge (Williams, Rosenberg 1986: 413). Since in this case theories are reduced to their leading explanatory principles, and the leading explanatory principle has the same logical structure as fitness explanations that are the focus of the propensity interpretation, the charge that the theory is circular and that fitness explanations are circular, are the same charge. In the next chapter, it is important to keep in mind that although the vocabulary of these two interpretations on this subject varies, they are addressing the same problem.

2.2.2. Fitness and Temperature

While Rosenberg agrees with Williams that 'fitness' is a primitive term in the theory of natural selection, he expands on her work to show that fitness is not reproductive success, but instead causes reproductive success. To Rosenberg, fitness "is a relational property, reflecting the interaction of an organism and its environment" (Rosenberg 1983: 458). Here, the environment means the sum of all the powers that can influence the organism to determine its fitness level. He argues that as there is a myriad of different causes at play for each organism, a complete causal explanation of an organism's fitness level is not achievable (ibid.). Some of these causes could be explained in molecular biology, others in geology, biogeography, etc., but all of them are outside the scope of the theory of natural selection.

Having roughly described what fitness is, he draws a parallel between fitness and temperature in how they are measured. What is sometimes called the multiple realizability argument[21], he argues that an infinitely complex causal chain can lead to the same fitness value in different organisms and therefore fitness must be measured by its effects like

[21] In philosophy of mind, it is the notion that a single mental kind can be realized by many physical kinds. In philosophy of biology, it is the notion that many different lower-level processes are realized as a higher-level process, e.g., different evolutionary paths lead to the same trait in different species, or a trait is coded by different parts of the genome.

temperature (ibid: 459). Temperature causes the substance in the thermometer to expand, and this effect of expansion is what we measure directly. The parallel is that fitness causes organisms to reproduce at different capacities and the actual reproduction rate is what we directly measure. In both cases what is indirectly measured (temperature, fitness) explains its effects that are measured directly (expansion of the liquid, actual reproductive rates). In short, fitness causes survival in nature and therefore the concept of 'fitness' can do the explanatory work in fitness explanations, but the property that is fitness is not directly observable, nor measurable.

Rosenberg's interpretation also serves as critique for the propensity interpretation when he claims that one does not need to evoke metaphysically problematic objective probabilities, i.e., propensities, to explain fitness when the analogy is much simpler. To him, the relation between fitness and actual reproductive rates is the same as between the hypothesis that a die is fair and a finite number of die rolls to prove that hypothesis. To claim that a die is fair is an empirical (observation) statement and therefore that claim is able to explain why each side comes up roughly an equal amount of times, but only an infinite sequence of rolls would prove the die is fair without doubt (ibid: 462). A finite set of rolls can still be used to assess the fairness of the die, just not perfectly. In addition, we could prove the die is fair independently of the rolls by looking at its composition with the aid of the different theories about matter.

Rosenberg stresses that if there were no other way to assess the die's fairness than a finite set of rolls, then there indeed would be a vacuous circular relation between the statement that the die is fair and the finite set of rolls that is supposed to prove it (ibid.). It is the same with fitness: if there were no independent method to measure fitness besides reproductive rates, then 'fitness' would be a truly vacuous term. I believe he means that if there is no possibility to distinguish between something directly observable that we measure and a posited hidden cause, then we should not posit that hidden cause, as it serves no purpose in the theory, nor is its existence verified. According to Rosenberg, there are independent ways to measure fitness with optimality models. Optimality models are a theoretical tool used in evolutionary biology to posit traits that would maximize the fitness of organisms in determined environments (Parker & Maynard Smith, 1990). It is used to measure the cost and benefit of different traits and to correct fitness measurements.

Let us now see how the primitivist interpretation is meant to avoid a circular fitness explanation. Here is the dialogue from the previous chapter that illustrated Manser's charge of circularity, but this time with the modification of the primitivist interpretation.

A: There are more black-coloured peppered moths in Birmingham than light ones.

B: Why?

A: Because they are fitter.

B: How do we know that?

A: Kettlewell measured the higher fitness of black-coloured moths indirectly through its effects which are the black-coloured moths' higher reproductive rates. Throughout the experiment, he took additional steps to correct his measurements of fitness by making sure the higher reproductive rates were not caused by other factors, such as genetic drift. This he did by relying on other theories and models in biology. Once having validated that the dark-coloured moths are fitter because a structural difference in their wings (melanism) shield them from predators, he concluded that natural selection was behind the evolution of these moths[22].

In this modification, when A answers B why there are more dark coloured moths, A refers to their higher fitness which she leaves undefined. When asked to explain fitness, A describes the process of measuring fitness and how with the help of other theories outside the theory of natural selection those measurements are corrected, establishing that fitness causes reproductive rates and can therefore explain reproductive rates.

Both the propensity and primitivist interpretation aim to show that fitness explanations are not circular by denying that fitness is reproductive rates. The propensity interpretation defines fitness as expected reproductive rates and interprets fitness as a propensity to reproduce. On the other hand, the primitivist interpretation does not define fitness anew, but instead interprets the term 'fitness' as it appears in the theory of natural selection, as a primitive term of that theory. The primitivists' aim is to dispel the charge of circularity by explaining how fitness is a complex property that causes reproductive rates and therefore, according to Hempel's schema, explains those rates.

[22] Kettlewell's experiments have been repeated and his conclusion reaffirmed.

3. THE DIAGNOSIS OF THE DISAGREEMENT BETWEEN THE PROPENSITY AND THE PRIMITIVIST INTERPRETATION

In the previous chapter I have described how the two interpretations have aimed to solve the charge of circularity. Their proponents have been in correspondence with each other, where their disagreement became evident. In this chapter, I will introduce their disagreement and argue that according to the proponents of primitivism, a definition of fitness that solves the charge of circularity should provide a reductive account of what fitness is within the theory of natural selection; and according to the proponents of the propensity interpretation, a definition of fitness that solves the charge of circularity need not provide a reductive account within the theory. In the first section, I will present the interpretations' similarities to show that the reason for the disagreement is not a fundamentally different understanding of what the referent of fitness is. Next, I will locate their disagreement. I will compare what each interpretation demands from a successful definition of fitness and I will argue that the proponents of the primitivist interpretation require a reductive definition of fitness due to their commitment to explanatory reductionism.

3.1. The Disagreement Between the Two Interpretations

3.1.1. Points of Agreement Between the Two Interpretations

In the words of proponents of the propensity interpretation: "With the exception of his [Rosenbergs's] conclusion, we find little with which to disagree" (Beatty, Brandon 1984: 342). Proponents of both interpretations believe the concept of fitness is indispensable to evolutionary biology and that its referent and role in scientific explanations has been misunderstood by critics. They clearly agree what the referent of 'fitness' is not: fitness, albeit defined so in various biological literature, is not reproductive success. Instead, fitness is what causes reproductive success and is primarily measured by reproductive success. Both construe fitness as a feature of the natural world that can be measured and is based on the physical properties of organisms. The propensity interpretation sees fitness as a probabilistically actualizing ability with a causal base in the organism's physical properties. To Rosenberg and Williams, fitness denotes the quality of the relationship between an organism and its environment which is grounded in the physical properties of both the organism and its surroundings.

Besides agreeing on what fitness is ontologically, both camps also share a faith in the Hempelian schema of scientific explanations – a faith which they also share with Manser and Popper who charged fitness explanations with being circular. According to the Hempelian schema, the empirical statements with at least one law in the *explanans* predict and explain their results formalized in the *explanandum*. Only then is one dealing with a scientific, i.e., genuine, explanation. If 'fitness' has a legitimate place in the *explanans* of fitness explanations, then it must have causal and empirical content, because that is what the Hempelian schema requires. This is evident from how the proponents of the propensity interpretation use the notion of propensity as an objective probability to give 'fitness' causal and empirical content to meet this requirement (Mills, Beatty 1979); and the proponents of the primitivist interpretation devote a whole article (Rosenberg 1984) and more besides (Rosenberg 1982; Williams, Rosenberg 1986) to explain that fitness has the causal and empirical content required by critics, but that content can only be described with the help of other theories outside the theory of natural selection. In the confinements of the theory of natural selection, 'fitness' is a primitive term.

The two interpretations do not have a substantially different metaphysical view of the referent of 'fitness'. Both interpretations share similarly empiricist views on what kind of content should scientific terms possess to correctly function in scientific explanations and set out in their own ways to demonstrate how 'fitness' possesses that right kind of content.

3.1.2. The Disagreement Over What Makes a Successful Definition of Fitness

Despite agreeing on the fundamentals of what 'fitness' refers to, proponents of the two views disagree over what kind of a definition of fitness solves the charge of circularity. My method for diagnosing their disagreement was gathering textual evidence from articles that were written during the debate, as well as from articles that were written on subjects related to the debate. I will first outline the standards that the proponents of both interpretations require from a definition of fitness and I will point out that there is one standard that they do not share, and I will demonstrate that the core of the disagreement originates from there.

The propensity interpretation has two criteria for the definition of fitness. To break the explanatory circularity, the propensity interpretation is aimed to provide fitness with a definition that would show that fitness is not reproductive rates. Therefore, the first necessary criterion is that 'fitness' needs to be "conceptually independent" from 'reproductive rates' (Beatty, Brandon 1984: 343). By saying fitness is the expected number of offspring they have fulfilled this requirement. However, this is not sufficient, because we could conceptually separate fitness from reproductive rates on paper, but if there is no empirically measurable way to distinguish fitness from reproductive rates, then the conceptual shift would be an empty gesture. Therefore, fitness needs to be "operationally independent" from reproductive rates (ibid.: 344). 'Operationally independent' here means that in scientific practice there needs to be an independent way to measure fitness apart from its effects which are reproductive rates. The propensity theorists claim that this is possible with optimality models (Beatty, Brandon 1984: 343).

Two things are important to note about the requirement of operational independence that will be relevant in my argument. The first important thing to note about the requirement of operational independence is that this requirement does *not* require that there be an operational specification of how fitness values are ascertained as a part of the definition of fitness; it only requires that such independent means to ascertain fitness levels exist. (ibid.: 345). The second important thing to note about this requirement is that the requirement does *not* require there to exist only one operational specification of fitness since there is no one way to measure fitness for different ecological settings and organisms, there is no one operational specification for all cases possible and this is not a problem according to Beatty and Brandon, Instead, it is important that there be *at least* one independent way to measure fitness apart from counting offspring (ibid.).

Let us now look at Rosenberg's three definitional standards (shared by Williams): firstly, fitness needs to be defined operationally, secondly, within the theory of natural selection and, thirdly, without it leading to explanatory circularity to solve the charge. However, he believes such a definition is not logically possible. Therefore, 'fitness' has no definition in the theory of natural selection and is instead a primitive term. The propensity theorists agree completely that such a definition of fitness is not achievable, but do not understand why we would need to adhere to Rosenberg's standards in the first place (Beatty, Brandon 1984). Why is the charge of circularity not solved by taking a concept like 'propensity' from outside the theory of natural selection and using it to define fitness in a manner that is conceptually separate from reproductive rates?

The answer is that to Rosenberg, the propensity interpretation's definition of fitness is not operational, and I am going to demonstrate this by breaking down each criterion and explicating their meaning. One criterion is shared by both interpretations and does not play a part in their disagreement: the definition of fitness needs to be conceptually independent from reproductive rates (Rosenberg 1982: 269). Now we are left with two other criteria that the primitivist interpretation requires and the propensity interpretation does not: first, that the definition must be given within the theory of natural selection and operationality. By breaking down these two criteria I am going to show that the primitivist interpretation's criterion of operationality is the real source of the disagreement.

Rosenberg's requirement that definitions of terms should be given within the theory they appear in is based on his view of the structure of scientific theories. Introduced in the previous chapter, Rosenberg holds that a defined term x that figures in the theory T cannot be defined with the theoretical vocabulary of another theory P, because the term x would then cease to be a defined term in the theory T. Therefore, any attempt to redefine x would have to be done within the axiomatic boundaries of T. Otherwise it would a be primitive term with regards to T. In other words, to Rosenberg, a definition of a term is a definition if and only if it is an explanation of that term within the theory the term figures in.

The logical consequence of this view is that for Rosenberg, if we want to give a definition for fitness, we necessarily need to give that definition within the theory of natural selection where fitness figures in.[23] However, the propensity interpretation accommodates this criterion, because Rosenberg already assumes that the propensity interpretation is also trying to provide a definition of fitness within the theory of natural selection and argues against their definition on other grounds. In this passage, Rosenberg writes why the propensity interpretation does not provide a definition of fitness:

> "Recognizing that the propensity characterization of 'fitness', or any attempt to define the term wholly within the ambit of evolutionary theory is an exercise in "implicit definition /../ Within the theory at best only implicit definitions can be offered for it [fitness], by treating one or another originally contingent statement that the theory makes about fitness as a stipulation about it. This is what Mills and Beatty do of course: this is why their characterization of fitness, when properly understood, is true, but no definition." (Rosenberg 1982: 271)

Rosenberg believes the aim of the propensity interpretation is to provide a definition of fitness within the theory of natural selection, but says it manages to give an implicit definition[24] of fitness which is not a true definition, but a characterization. In this passage, the exact meaning of how Rosenberg understands definitions is left vague, but there is visibly a distinction between a characterization and a definition. Based on his later 1983 article, a definition of a term is an explanation of a term (463). Therefore, it is plausible that to Rosenberg, the difference between the characterization and the definition of a term is that characterisations are descriptions or stipulations about some facet of the referent of the term

[23] Strictly speaking, he is not giving a criterion for a suitable definition of fitness here, he is just presupposing his view of what definitions are.

[24] Based on my reading of Rosenberg, implicit definitions are characterizations of terms within a theory that imply their full explanation found in other theories.

within the vocabulary of the theory the term appears in, but definitions are explanations that should capture, for lack of a better word, the 'constitution' of the referent.[25]

In Rosenberg's view the theory of natural selection is unable to provide a definition, i.e., a sufficient explanation of fitness, even if we introduce propensities to the theory (1982). The idea here is that the theory of natural selection does not say anything about what fitness consists of, but even if we try to introduce a new term to the theory with which to define what fitness consists of, we come to no definition, that is no explanation, for the newly introduced term within the theory. This is what happens with interpreting fitness as a propensity: after introducing the term 'propensity' to the theory, we lack the means to define it further. Rosenberg asks what the causal base is which the propensities are meant to capture(ibid.). What did we really learn about fitness from introducing propensities? It becomes clearer that what he requires from a definition, is some form of explanation of fitness' causal base, because interpreting fitness as a propensity does not on its own add any new empirical content to the theory of natural selection.

Based on his reasoning, it seems we have two options, either we do not define 'propensity' further, which still makes fitness a primitive (just with an extra step that it is a propensity); or we could introduce more terms and their definitions to the theory of natural selection, until we reach some sufficient point where Rosenberg would say we have explained and thus defined fitness. By doing this, however, we have stretched the theory of natural selection so far that it loses its original meaning. Still, the question remains about the sufficient point where Rosenberg would say we have provided a suitable definition for fitness.

To answer that question, we need to look at his requirement of operationality which is evident from this passage critiquing the propensity interpretation:

"Now, it is not a necessary nor a sufficient condition on the definition of a theoretical term that the definition be "operational", or even that it have operational consequences. But this does seem to be required of any definition constructed to show that the theory, in which the defined term is in, is non-tautologous." (Rosenberg 1982: 271)

To Rosenberg, the propensity interpretation provides an "operationally sterile" description of fitness that introduces propensities to the theory but does not tell us what propensities are and how they should be counted (ibid.). 'What propensities are' plausibly refers to the explanation of their causal base. Operational definitions are commonly understood as definitions of scientific terms where the *definiens* explains some facet of measuring the

[25] I mean it in the sense how John Locke's 'real definitions' are definitions where the definiens captures the underlying 'constitution' of the definiendum that the definiendum's properties are based on. (Locke 1689)

definiendum, but judging from Rosenberg's complaint, it seems that his notion of operationality goes a bit further than just explaining some facet of measuring the *definiendum*.

Consider Rosenberg's example of an operational definition of acid, where an acid is defined as a chemical species that donates protons and accepts electrons (ibid.: 270). The idea in this example seems to be that the operational definition cites acid's underlying mechanical behaviour to donate and accept particles. In other words, it cites acid's so-called universal molecular behaviour, whether it be sulphuric or boric acid; it applies in every physical context and in every situation of measurement. Rosenberg's sense of an operational definition is more than just explaining some facet of measuring the *definiendum*; rather, it is a definition that tells us the underlying 'constitutional' property of the definiendum that we are measuring in every measuring instance and this definition should also imply how this underlying constitution could be measured.

The same criterion appears again in the final article of this debate, where Rosenberg and Williams write:

"Beatty and Brandon tell us that there are many operational means of ascertaining [ways of measuring] fitness levels, but no single all-purpose one. True enough, but for all these means to pick out the same property, the definition of fitness cannot simply be a schema into which all these alternative and incompatible operational means can be plugged. The definition must be part of a theoretical explanation of why they [different ways to measure] all work, each in its different way under different circumstances." (Williams, Rosenberg 1986: 416)

In this later article, Rosenberg curiously seems to change his mind by clearly stating that he does not require an operational definition from fitness (ibid.: 415), but then goes on to demand with more clarity the same thing he was implying when he used the term 'operational' in his 1982 article cited previously. It therefore stands to reason that 'operationality', in Rosenberg's 1982 meaning, is not the requirement that the *definiens* should explicitly or implicitly cite some or all means of measuring the *definiendum*, nor is it the requirement that there should be one operational means to measure fitness in all circumstances, like the proponents of the propensity interpretation believed Rosenberg was arguing for (Beatty, Brandon 1986: 345)[26]. A more plausible reading is that the operationality requirement in Rosenberg's 1982 article was the same requirement from his 1986 article where he argues together with Williams that the *definiens* needs to explain why

---

[26] "Why there should be one independent means of ascertaining fitness differences, one all-purpose optimality model, escapes us. That part of Rosenberg's argument is missing" (Beatty, Brandon 1986: 345).

all the different ways of measuring fitness work, i.e., to explain why all the different operations in different contexts work.

The propensity interpretation's criterion of operational independence is different from Rosenberg's criterion of operationality. For the former, operational independence does not require that there be an operational specification of how fitness values are ascertained as a part of the definition of fitness; for the latter, the definition needs to contain an operational specification for fitness. Since both interpretations share the criteria that 'fitness' must be conceptually separate from reproductive rates and the propensity interpretation's definition is compatible with Rosenberg's second criteria that fitness needs to be defined within the theory of natural selection, the disagreement between the proponents' of the two interpretations boils down to how they understand the criteria of operationality differently.

3.1.3. The Role of Explanatory Reduction in the Disagreement

Rosenberg holds that operational definitions of terms are definitions where the *definiens* is a theoretical explanation of the *definiendum's* causal base which demonstrates why all the methods of measuring the *definiendum* measure the same property or underlying constitution of the *definiendum*. As such, I argue, the primitivist interpretation's standard of operationality is a reductive standard. It is reductive in the sense that the criterion of operationality is motivated by a strong version of explanatory reductionism that places greater explanatory power to sentences formed within lower-level theories that feature mechanistic causal explanations rather than higher-level theories that feature statistical explanations. If operational definitions are reductive definitions, then the primitivist interpretation requires a reductive definition of fitness to solve the charge of circularity and the propensity interpretation does not.

Explanatory reduction is a type of epistemic reduction where an explanation from a higher-level scientific domain like evolutionary biology is logically reduced to an explanation from a lower-level scientific domain like physics or chemistry (Brigand, Love 2017). There are many models of reductive explanations, but most commonly a reductive explanation is a causal explanation "where a higher-level feature is explained by the interaction of its constituent parts" (ibid.), e.g., 'fitness', a concept of the higher-level theory of natural selection which could be explained (and therefore defined) by the interaction of its constituent parts, where 'constituent parts' is another way of saying 'the causal base' of fitness. However, the fact that there could logically be a reductive definition of fitness where its causal base is specified does not mean we should necessarily favour it over a non-reductive definition of fitness. Rosenberg's belief that only a reductive definition of fitness could solve the charge of circularity is because he favours a reductive definition of fitness over a non-reductive definition.

To better map Rosenberg's view, I will differentiate the practice of explanatory reduction from the normative view of explanatory reductionism. A weak version of explanatory reductionism is the belief that explanations from a higher-level domain can be reduced to explanations in lower-level domains without losses in meaning. A strong version of explanatory reductionism is the view that explanations from a lower-level domain are more complete and, in that sense, "more explanatory" than their equivalent explanations given in a higher-level domain.

If one were a strong explanatory reductionist about evolutionary biology to organic chemistry, they would hold that explanations in evolutionary biology could in more detail be expressed within organic chemistry. A reductionist of this kind would say that the lower-level domain provides us with more complete descriptions and explanations of phenomena. This could be due to the reductionist favouring causal mechanistic explanations above others – and lower-level theories seem to deliver in that regard[27]. The same would apply to definitions of scientific terms: a primitive term in evolutionary biology would have a more adequate definition in organic chemistry. In this reading, higher-level theories have merit only because they enable scientists to condense information and pursue different questions autonomously, but better and more detailed explanations lie within lower-level theories.

Keeping this picture of a strong explanatory reductionist in mind, let us turn back on the criterion of operationality. The primitivist interpretation of fitness does not provide a definition of fitness to solve the charge of circularity, because according to its proponents an operational definition of fitness is impossible to formulate within the theory. Rosenberg and Williams hold this definitional standard because they are strong explanatory reductionists about what constitutes a sufficient definition of fitness to dispel the charge. Their demand for an operational definition which is "/../a theoretical explanation of why all the different ways of measuring fitness work/../" (Williams, Rosenberg 1986: 416) is reductionist, because this sort of theoretical explanation of fitness' causal base can only be expressed with a lower-level explanation.

[27] It can be argued that physics, being the lowest level domain, provides more causal and deductive explanations than higher level theories where the sheer volume of causal chains affecting the object of study allows only for statistical inferences.

For the proponents of the primitivist interpretation, a definition that captures all the molecular and causal processes that lead to differing fitness levels would adequately satisfy the criteria of “empirical causal content” that is needed of ‘fitness’ to dispel the charge. These processes would be explainable largely by lower level explanations than those found in the theory of natural selection, like by explanations from anatomy and molecular biology that specify the mechanics of an organism’s metabolism and reproductive system.[28] Like the example of an operational definition of acid (Rosenberg 1982) according to which acid is a chemical kind that donates protons and accepts electrons, so should fitness be defined in a fashion that explicates its mechanistic constitution or property that reveals in a lower-level domain how it causes observable fitness rates.

Another example of the primitivist interpretation’s analogy for demanding reductive content from the definition of fitness would be Williams’ and Rosenberg’s comparison of fitness and temperature. On temperature they write: “What we do need to define [temperature] is one (or at most a small number of) underlying mechanisms that constitute or produce temperature” (Williams, Rosenberg 1986: 416). Temperature is defined as mean kinetic energy and they use the definition of temperature to make a point about what sort of a definition would be able to solve the charge of circularity. The reductive notion in their comparison is that the ‘underlying mechanisms that constitute or produce temperature’ can only be given in the lower-level domain of thermodynamics. They add that fitness needs a definition similar to temperature (ibid.).

In comparison, the proponents of the propensity interpretation do not require any reductive content from their definition of fitness. Mills and Beatty write:

> “What do the fittest germ, the fittest geranium, and the fittest chimpanzee have in common? It cannot be any concretely characterized physical property, given that one and the same physical trait can be helpful in one environment and harmful in another. This is not to say that it is impossible that some as yet unsuspected (no doubt abstractly characterized) feature of organisms may be found which correlates with reproductive success. Rather, it is just to say that we need not, and should not, wait for the discovery of such a feature in order to give the definition of "fitness”. (Mills, Beatty 1979: 269)

They believe that propensities, without needing to give a lower-level reductive account of their causal base[29], carry enough causal and empirical content on their own to show that ‘fitness’ has explanatory power to dispel the charge. Propensities by virtue of what they are should do the work to show that ‘fitness’ is a causal term, and the theory of natural selection

[28] Of course, these explanations could be complimented by higher-level theories like geography and geology, but their role would be to provide explanations of environmental influences on the organism. These influences would be formulated in terms of the lower-level explanations that explain the causal processes of the organism influencing its fitness rate.

[29] Be it a single property or otherwise.

is a causal theory that gives causal explanations. A probabilistic higher-level account is enough to explain how fitness causally influences reproductive rates, as long as it makes the definition of 'fitness' conceptually and operationally independent from reproductive rates. In conclusion, the propensity interpretation is not committed to favouring explanatory reductionism and does not share the requirement that a successful definition of fitness needs to be reductive.

The difference between the propensity interpretation of fitness and the primitivist interpretation of fitness stems from their different understanding of what a successful definition should achieve and not from their different understanding of what fitness is in nature. To Rosenberg and Williams, a definition of fitness that solves the charge of circularity should provide a reductive account of 'fitness' within the theory of natural selection; and Mills, Beatty and Brandon hold that a definition of fitness that solves the charge of circularity need not provide a reductive account within the theory. Therefore, the decision which account to favour should be based on whose account of definitions is correct.

3.2. Possible Objection

Williams and Rosenberg present a "third interpretation" under which the propensity interpretation is "a terminological variant" of their primitivist interpretation. Fitness, construed as a propensity and 'fitness' defined as the expected number of offspring will be added to the theory of natural selection and not defined further, thereby making fitness a propensity and 'fitness' a primitive term. In this reading, what is behind the propensity, i.e., its causal base, will be characterized piece-meal by other models and theories outside the theory of evolution. (Williams, Rosenberg 1986: 418)

Even if the primitivist interpretation takes the propensity interpretation under its umbrella, it will be on the terms of primitivist interpretation and the disagreement over definitional standards would persist. According to the propensity interpretation, fitness has a revised definition that solves the charge, but according to the primitivist interpretation, the propensity interpretation fails to give a definition. In their unification, fitness would be a propensity, but it would not have a definition.

CONCLUSION

In this thesis, I argued that the disagreement between the propensity and primitivist interpretation of fitness stems from their differing standards on scientific definitions which solve the problem of circularity. According to the proponents of primitivism, a definition of fitness that solves the charge of circularity should provide a reductive account of what fitness is within the theory of natural selection; and according to the proponents of the propensity interpretation, a definition of fitness that solves the charge of circularity need not provide a reductive account within the theory. Therefore, the difference between the propensity interpretation of fitness and the primitivist interpretation of fitness stems from their different understanding of what a successful definition should achieve and not from their different understanding of what fitness means in biology.

In the first chapter, I described the Inductive-Statistical Model which is the model of fitness explanations in biology and defined fitness as reproductive success and natural selection as the constant difference in fitness in a population in the absence of other evolutionary forces and argued that one cannot be discussed without the other to demonstrate that the charge of circularity touches on both of those concepts.

In the second chapter, I introduced the two interpretations of fitness. According to the propensity interpretation, fitness is a propensity, an ability of organisms to survive and reproduce. Fitness is the expected number of descendants, instead of actual number of descendants, which should make fitness explanations causal explanations. The primitivist interpretation claims fitness is a primitive term of evolutionary theory that does not have a definition, nor does it need to, because the charge of circularity can be explained away by showing how fitness is measured and understood separately from reproductive rates.

In the third chapter I presented my argument that the disagreement between these interpretations does not stem from their differing understanding of fitness, instead the disagreement is about whether the definition of fitness should be operational to solve the charge. I further argued that Rosenberg's and Williams' operationality requirement stems from a strong version of explanatory reductionism – a view according to which lower-level mechanistic causal explanations are more complete or some other way more desirable than higher level explanations. An operational definition would explicate fitness's causal base that is in principle describable only with lower-level theories than the theory of natural selection. The propensity interpretation proponents make no such requirements and find propensities themselves provide enough causal and empirical content to solve the charge.

ABSTRACT

This thesis explores the disagreement between the two earliest attempts by philosophers of biology to explain the meaning and explanatory role of evolutionary fitness in the theory of natural selection. I compare two interpretations of fitness: the propensity interpretation and the primitivist interpretation. Both aimed to solve the charge that due to the way fitness had been construed by biologists, the theory of natural selection offers circular explanations. I argue that their disagreement was not in their different understanding of what fitness is, but in their standards for a successful definition of fitness that would solve the charge. The primitivist interpretation required a reductive definition of fitness, while the propensity interpretation did not.

RESÜMEE

Kas kohasus on kalduvus või primitiiv? Seletusliku reduktsionismi roll kohasuse vastukäivates definitsioonistandardites

Antud magistritöös võrdlen kaht bioloogia filosoofide kõige varasemat püüet selgitada evolutsioonilise kohasuse tähendust ja seletuslikku rolli loodusliku valiku teooras. Esimene püüe oli kohasust tõlgendada kui statistilist kalduvust (*propensity*). Teine püüe oli tõlgendada kohasust kui loodusliku valiku teooria primitiivset ehk definitsioonita mõistet. Mõlema tõlgenduse eesmärk oli lahendada tsirkulaarsusprobleem, mille järgi on kohasus bioloogide poolt defineeritud viisil, et loodusliku valiku teooria pakub tsirkulaarseid seletusi. See töö kaitseb väidet, et tõlgenduste erimeelsus ei seisnud nende erinevas arusaamas sellest, mis kohasus on looduses, vaid milline kohasuse definitsioon võimaldab lahendada tsirkulaarsusprobleemi: primitivistlik tõlgendus taotles reduktiivset kohasuse definitsiooni ja kalduvustõlgendus ei taotlenud reduktiivset definitsiooni.

REFERENCES

Brandon, R. & Beatty, J. (1984). “The Propensity Interpretation of' Fitness--No Interpretation Is No Substitute”. *Philosophy of Science*, 51(2), 342-347

Brigandt, I. & Love, A. "Reductionism in Biology", *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2017/entries/reduction-biology/>.

Choi, S. & Fara, M. "Dispositions", *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2021/entries/dispositions/>.

Creath, R. "Logical Empiricism", *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2021/entries/logical-empiricism/>.

Darwin, C. (1859). *On the Origins of Species*. London: John Murray, Albermarle Street.

Descartes, R. (2003). *Discourse on Method, and Meditations*. New York: Dover Publications.

Endler, J. A. (1986). *Natural selection in the wild*. Princeton University Press.

Futuyma, D.J., Kirkpatrick, M. (2017). *Evolution*. Fourth Ed. Sunderland, Massachusetts: Sinauer.

Hájek, A. "Interpretations of Probability", *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2019/entries/probability-interpret/>.

Hempel, C. (1965). *Aspects of Scientific Explanation, and Other Essays in the Philosophy of Science*. New York: The Free Press.

Humphreys, P. (1985), “Why Propensities Cannot Be Probabilities”, *Philosophical Review*, 94: 557–70

Lambret, Philippe & Rutter, Imogen & Grillas, Patrick & Stoks, Robby. (2018). “Oviposition plant choice maximizes offspring fitness in an aquatic predatory insect”. *Hydrobiologia*. 823

Lewontin, R. C. (1970). “The Units of Selection”, Annual Review of Ecology and Systematics, 1(1): 1–18

Locke, J. (1689). An Essay concerning Human Understanding, edited by P. H. Nidditch, Oxford: Oxford University Press (1975).

Manser, A. R. (1965) "The Concept of Evolution." Philosophy, XL:18-34

Matthen, M., Ariew, A. (2002). “Two ways of thinking about fitness and natural selection,” *Journal of Philosophy*, No. 99 (2): 55–83

Mayr, E. (1969). “Footnotes on the Philosophy of Biology,” Philosophy of Science, 36(2): 197–202

Mills, S. & Beatty, J. (1979). The Propensity Interpretation of Fitness. Philosophy of Science, 46(2), 263-286

Mumford, S. (1998). Dispositions. Oxford: Oxford University Press.

Odenbaugh, J. & Griffiths, P. "Philosophy of Biology", The Stanford Encyclopedia of Philosophy (Summer 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2020/entries/biology-philosophy/>.

Orilia, F. (2000). Meaning and Circular Definitions. Journal of Philosophical Logic Vol. 29, No. 2, pp. 155-169

Orr, H. A. (2009). „Fitness and Its Role in Evolutionary Genetics”. Nature Reviews Genetics 10 (8): 531–39

Orzack, S. H. & Forber, P. "Adaptationism"2017, The Stanford Encyclopedia of

Philosophy (Spring 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2017/entries/adaptationism/>.
Otsuka, J. A. (2016). Critical Review of the Statisticalist Debate. Biol Philos 31, 459–482
Parker, G., Smith, J. (1990). Optimality Theory in Evolutionary Biology. Nature 348, 27–33
Prior, E., Pargetter, R., & Jackson, F., (1982). 'Three Theses About Dispositions', American Philosophical Quarterly, 19: 251–257
Popper, K. (1959). "The Propensity Interpretation of Probability", British Journal of the Philosophy of Science, 10: 25–42
Popper, K. (2005). Unended Quest: An Intellectual Autobiography. Revised edition. Taylor and Francis e-Library.
Putnam, H. (1999). "Problems with the observational/theoretical distinction", in Scientific Inquiry, Robert Klee, ed (New York: Oxford University Press, 1999), 25–29
Rosenberg, A. (1982). "On the Propensity Definition of Fitness", Philosophy of Science, 49 (No.2): 268–273
Rosenberg, A. (1983). "Fitness", Journal of Philosophy, 80 (No.8): 457-473
Salmon, W. C. (2006). "Four decades of scientific explanation". University of Pittsburgh press: Pittsburgh
Smart, J. J. C. (1963). Philosophy and Scientific Realism. London: Routledge and Kegan Paul.
Walsh DM, Lewens T, Ariew, A. (2002). "The trials of life: natural selection and random drift". Philosophy of Science, No. 69(3):452–473
Williams, M. B. (1973). "The Logical Status of Natural Selection and other Evolutionary Controversies: Resolution by Axiomatization." Methodological Unity of Science (Bunge, ed.). Dordrecht: Holland
Williams, M.,B. & Rosenberg, A. (1985). "Fitness' in Fact and Fiction: A Rejoinder to Sober. Journal of Philosophy, No. 82 (12):738-749.
Willims, M.,B. & Rosenberg, A. (1986). Fitness as Primitive and Propensity. Philosophy of Science, Vol. 53, No. 3:412-418
Williams, M., B. (1970). "Deducing the consequences of evolution: a mathematical model", Journal of Theoretical Biology No. 29:343–385.
Sober, E. (2001). The Two Faces of Fitness. Manuscript. https://www.researchgate.net/publication/244305192_The_Two_Faces_of_Fitness
Sober, E. & Elgin, M. (2017). Popper's Shifting Appraisal of Evolutionary Theory. Hopos: The Journal of the International Society for the History of Philosophy of Science, No 7 (1):31-55.
Woodward, James, "Scientific Explanation", The Stanford Encyclopedia of Philosophy (Winter 2019 Edition), Edward N. Zalta (ed.), URL <https://plato.stanford.edu/archives/win2019/entries/scientific-explanation/>.

Non-exclusive licence to reproduce thesis and make thesis public

I, Anna Elise Rohtmets,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

Is 'Fitness' A Primitive or a Propensity? Diagnosing the Role of Explanatory Reductionism On Differing Standards of Scientific Definitions,

supervised by Edit Talpsepp PhD and Riin Kõiv MA.

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Anna Elise Rohtmets
15.05.2021