THE UNIVERSITY OF
SYDNEY

Reconceptualising Evolution by Natural Selection

Pierrick Bourrat

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Faculty of Arts and Social Sciences

The University of Sydney

2015

Abstract

This thesis examines the theoretical and philosophical underpinnings of the concept of natural selection which is pervasively invoked in biology and other 'evolutionary' domains. Although what constitutes the process of natural selection appears to be very intuitive (natural selection results from entities exhibiting differences in fitness in a population), this conceals a number of theoretical ambiguities and difficulties. Some of these have been pointed out numerous times; others have hardly been noticed. One aim of this work is to unpack these difficulties and ambiguities; another is to provide new solutions and clarifications to them using a range of philosophical and conceptual tools. The result is a concept of natural selection stripped down from its biological specificities.

I start by revisiting the entangled debates over whether natural selection is a cause of evolutionary change as opposed to a mere statistical effect of other causes, at what level this putative cause operates and whether it can be distinguished from drift. Borrowing tools from the causal modelling literature, I argue that natural selection is best conceived as a causal process resulting from individual level differences in a population. I then move to the question of whether the process of natural selection requires perfect transmission of types. I show that this question is ambiguous and can find different answers. From there, I distinguish the process of natural selection from some of its possible products, namely, evolution by natural selection and complex adaptation. I argue that reproduction and inheritance are conceptually distinct from natural selection, and using individual-based models, I demonstrate that they can be conceived as evolutionary products of it. This ultimately leads me to generalise the concepts of heritability and fitness used in the formal equations of evolutionary change. Finally, I argue that concepts of fitness and natural selection crucially depend on the grains of description at and temporal scales over which evolutionary explanations are given. These considerations reveal that the metaphysical status of the process of natural selection is problematic and why neglecting them can lead to flawed arguments in the levels of selection debate.

À la mémoire de mon père, Michel,
qui bien qu'il soit parti trop tôt
m'a transmis sa curiosité

Contents

List of figures

N.B. Many figures are COLOUR figures and will be hard to interpret in black and white.

List of tables

Acknowledgements

My first words of gratitude go naturally to my close family: my mother, Christine and my sister, Audrey, who have always been there for me, especially during difficult times. I would also like to thank my grand-mother Marie-Thérèse and my late grand-father Marcel whose generosity and kindness have been without limits. I also thank my aunt and uncle, Annie and Christian and my cousins Fabien and Mathieu.

As far as I remember, I have always been interested in science, but I truly discovered the wonderful world of biology when I met Sébastien Mirkovic in his vet clinic, in Villars, France and started to visit him every day after school when I was about 12 years old. Since then, I have continued to pop in every so often and continue to learn new things about animal medicine and surgery. I can truly say that Sébastien inspired me in many respects. He is an example of a rigorous mind, generosity and honesty. With Sébastien and Valérie Garnier, his assistant, I have spent the most blissful time of my life and for that they will always hold a special place for me.

Although I did not become a vet, I kept my interests in biology, especially under the influence of Jacques Valiergue who was teaching biology in the vet preparatory school, in Clermont-Ferrand, I was attending and who, thanks to his mastering of the subject, was able to make it truly fascinating. During that time, I started to be interested in human behaviour and I soon discovered evolutionary psychology. I found the idea that human behaviours are the product of natural selection, as any other complex traits, fascinating and started reading everything I could find on the subject. After this time I visited and then worked in Oxford (2008 and 2009-2010) on the evolution of religion which was a game changer. There I discovered that

evolutionary theory, particularly when applied to anything different from fictional entities in a model, is a complicated matter, and this is especially true in the domain of human culture. I would like to thank Robin Dunbar, Quentin Atkinson and Harvey Whitehouse for giving me the opportunity of working on the EXREL project. Working on this project has allowed me to see many of the philosophical issues associated with evolutionary theory and this ultimately led me to the project of this thesis. I met in Oxford wonderful and interesting people with whom I had some of the most stimulating philosophical debates. I wish to thank especially my friend Daniel Darg who is another example of rigorous mind and generosity. My level in English today is for a great part due to Daniel. I also wish to thank Ryan Mckay with whom I ran a cool experiment in 2010 mixing experimental economics and cognitive science of religion and who guided me during the process. Sometimes I miss experiments!

During the year I spent in Paris (2008-2009), I have also met decisive people. I thank Dan Sperber for his support and whose work is inspiring. Dan allowed me to work with the people of his lab, the NaSH (Naturalisme en Sciences Humaines) and I would like to thank them all: Nicolas Baumard, Coralie Chevalier, Nicolas Claidière, Hugo Mercier, Olivier Morin, Olivier Mascaro, Anikó Sebestény, Jean Baptiste André and Hugo Viciana. More particularly I thank Hugo Viciana who introduced me to the small world of philosophy of biology and Nicolas Baumard who guided me in my empirical work. I also would like to thank Vito Mandorino with whom I had and continue to have very fruitful discussions on mathematics and Philippe Huneman and Jean Gayon for their support.

I would like to thank Philippe Lachaume and Guila Ganem for their patience and generous time. Guila, besides her support over the last eight years or so, found some of my quirks interesting and this made my internship with her a very enriching experience. During the third

year of my Bachelor (2006), the course on Evolution run by Philippe was a breathing space for me and Philippe noticed my interests in Evolution. Since then I am always happy, when returning to Clermont-Ferrand, to meet and discuss this topic with him.

My coming to Sydney could not have been possible without my supervisor Paul Griffiths. Paul is yet another example of generosity and rigour of mind. Paul is inspiring, a real mentor and an endless source of knowledge. He gave me countless extremely good pieces of advice during these three years and a half. I thank him for his guidance and support, but also for the special care he gives to his students. I can say with confidence that Paul is an outstanding supervisor and a wonderful person to be around.

Many of ideas in this thesis have been built in reaction to the work of Peter Godfrey-Smith. I wish to thank Peter for the time and his critical pieces of advice he gave me, especially during my visit to New York, where we discussed different aspects of evolutionary theory. This thesis would have been quite different without him. Two other works that this thesis has drawn from are the works of Frédéric Bouchard and Samir Okasha. I would like to thank them both for the interactions I had with them. I would like to thank particularly Frédéric for his many professional tips.

In Sydney I also want to thank Maureen O'Malley for her time and pieces of advice, Karola Stotz, Kristie Miller and David Braddon-Mitchell who helped me at different stages of my PhD. I had many interesting and fruitful discussions with Patrick Forber and John Matthewson during their visits to Sydney. I would like to thank particularly Patrick for his advice on the evolution of cooperation. Finally I also would like to thank Paul Rainey, Quentin Atkinson, Russell Gray and Katrin Hammerschmidt for the time they gave me during my visit to Auckland.

There are many other people I wish to thank for diverse reasons both academic and personal I cannot go into details here: Arnaud Pocheville, Adam Hochman, Michael Duncan, Robyn Kath, Qiaoying Lu, Kate Lynch, Andrew Latham, Elena Walsh, Dominic Murphy, Mark Colyvan, John Wilkins, Wesley Fang, Karl Rollings, Linus Huang, Isobel Ronai, Georg Repnikov, Johann Hariman, Susanna Saracco, Steven Orzack, Emanuele Serelli, Mark Olson, Arnon Levy, Adrian Currie, Kim Sterelny, Michael Weisberg, Brett Calcott, Romain Guerreiro, Iván González-Cabrera, Rachael Brown, Silvia De Monte, Mark Tanaka, Steven Hamblin, Inja Stracenski, Angelique Ockert and Susan Cirillo.

I thank more particularly for their comments and help with proofreading: Elena Walsh, Linus Huang, Georg Repnikov, Isobel Ronai, Hugo Viciana, Anikó Sebestény and Olivier Morin for the Introduction and Conclusion; Adam Hochman, Anikó Sebestény and Robin Kath for Chapter 1; Michael Duncan for Chapter 2; Karl Rolling for Chapter 3, Arnaud Pocheville and Kate Lynch for Chapter 4; John Matthewson, Arnaud Pocheville and Olivier Morin for Chapter 5; Kate Lynch, Arnaud Pocheville and Olivier Morin for Chapter 6.

Finally, writing this thesis has been possible thanks to my receipt of the *University of Sydney International Scholarship* (2011-2014), funding from the University of Sydney's *Postgraduate Research Support Scheme* (2012, 2013 and 2014), a *Doctoral Travel Grant* (2012) from the University of Sydney and a top-up scholarship under the Australian Research Council's Discovery Projects funding scheme DP0878650.

Pierrick Bourrat

Sydney
August 2014

0. Introduction

This thesis is about the concept of natural selection. Reference to natural selection is pervasive both in the scientific and non-scientific literature. Yet, it is far from being clear what natural selection in and of itself represents. Natural selection is often presented as the main actor of many evolutionary explanations. But it is often forgotten that it is only one of several processes that can lead to evolution. To that remark, some, including myself, will be tempted to answer that there is something distinctive about natural selection that no other evolutionary processes seem to match and for that reason to regard themselves as adaptationists. The distinctiveness of natural selection is the main topic of this thesis.

Many features make natural selection a very distinctive evolutionary process. Maybe the most obvious one is that explanations invoking natural selection are the only known explanations of adaptations in biology (Dawkins 1986; Ridley 1996), in other words, traits that have the distinctive characteristics of living systems, namely functional complexity and goal directedness. In the absence of an alternative explanation, natural selection occupies thereby a very special place among evolutionary processes.

In 1979, Gould and Lewontin published the most often cited criticism of adaptationism (Gould & Lewontin 1979), that is, the view that "asserts the central importance of adaptation and natural selection to the study of evolution" (Godfrey-Smith 2001; see also Orzack & Forber 2010). The take home message Gould and Lewontin delivered is that adaptation (as the result of natural selection) is only one possible product of evolution. As a consequence, they argued, any adaptationist claim should be rigorously assessed and alternative hypotheses explored. Although

many of the arguments provided by Gould and Lewontin are sound, this criticism of the adaptationist program did not take into consideration the fact that adaptationism can take several distinct forms. Godfrey-Smith (2001) claims that there are three different and independent notions of adaptationism. The three concepts are: empirical adaptationism, which asserts that natural selection is the main force of evolution; explanatory adaptationism, which holds that natural selection is a key process in the apparent design of organism to their environment; and finally, methodological adaptationism, which holds that adaptation is the best way to approach biological questions.

Because adaptationism is a multifaceted concept, even if natural selection did not turn out to be the main architect of the apparent design and the diversity of forms of life surrounding us, that would only undermine one aspect of it, namely empirical adaptationism. Explanatory and methodological adaptationisms would certainly remain untouched, especially because, as I emphasised earlier, there is, so far, no alternative to it. Many biological explanations are functional explanations. Since, for the time being, natural selection is the only process able to explain the origin of functional complexity, explanatory adaptationism and thus the study of natural selection in and of itself, is plainly justified.

For that reason, I should be clear upfront that this thesis is framed within the context of explanatory adaptationism. More specifically, I aim at understanding what natural selection in and of itself is and what it can explain given that its existence is unquestioned. However, I will not commit to empirical adaptationism. This is because, in my view, the question as to whether natural selection is the predominant evolutionary process that explains evolution does not make much sense. In fact, I will argue in Chapter 5 that whether the process natural selection occurs on a trait in a population depends on the choices of scales of description one makes. Since these choices

are to some extent arbitrary or made for practical purposes or material limitations, the answer to the aforementioned question seems to be intractable. I also will not commit to methodological adaptationism. The view that any biological problem *should* necessarily be approached through adaptation and natural selection seems problematic since other approaches have led to important discoveries in biology, most notably the structure of DNA in 1953 (Watson & Crick 1953).

Another distinctive aspect of natural selection, although this is common to other physical processes (e.g., phase transitions) and some other evolutionary processes (e.g., drift), is that it is a population-level concept. In fact, natural selection only occurs once it is applied to a population and as result of difference between members of the population. This makes the process philosophically more demanding than the concept of mutation (another evolutionary process), which ultimately is an individual-level process, since mutation act upon individual entities. Because natural selection is a population-level concept, some authors have argued that natural selection is a population-level cause. We will see however in Chapter 1 that the notion of natural selection as being a population level process resulting from individual level causes can be successfully defended and is superior to the alternative view that regards natural selection as a population level cause.

Another peculiarity of the process of natural selection when compared to other biological or physical processes, is that it has been applied to a wide range of entities. For example, the concept of natural selection has been applied to culture and cultural groups (e.g. Aunger 2000; Blackmore 1999; Boyd & Richerson 1985, 2005; Cavalli-Sforza & Feldman 1981; Dawkins 1976; Lumsden & Wilson 1981; Richerson & Boyd 2005), to clay crystals in order to explain the origin of life (Cairns-Smith 1990), to the universe (Gardner & Conlon 2013; Smolin 1992), in economy (e.g., Hodgson & Knudsen 2010; Nelson & Winter 1982), to the brain and particularly its neural

development (Edelman 1987) and to the immune system (Burnet 1958; Jerne 1955). This demonstrates that natural selection is a rich concept. No other physical or biological process has matched the range of potential applications natural selection has. Let me emphasise that whether these applications have been successful is independent of the question of whether natural selection *can* be in principle applied to population of entities that are not biological one.

Yet, although natural selection is pervasively used in different sciences, one cannot miss the fact that its definition(s) are quite vague. In Chapter 4, I will introduce one of the most widely used formalisms in the theoretical literature on natural selection, namely the Price equation (Price 1970, 1972a). Even though this formalism relies on mathematical notation, there is some controversy regarding which term(s) in the equation should be attributed to natural selection. Grasping the fundamental features of natural selection is rendered even more difficult if one notices that it is defined in many different ways and is often entangled or conflated with the concepts evolution, adaptation and fitness. This is one reason that this Introduction provides several definitions of terms used in evolutionary theory I will strive to use consistently throughout the thesis. I will come back to the definition of natural selection more specifically in Chapter 2 where I evaluate the different propositions of its definition made in the biological and philosophical literature.

Another reason responsible for the lack of unity and clarity in the definition of natural selection is that it is, for the most part, defined in statistical terms of which an objective probabilistic interpretation is usually given. This is for example the case in the Price equation (see Chapter 4) in which natural selection is defined with a covariance, a form of correlation. However, probabilistic explanations combined with the claim that there is nothing above and beyond such explanations is, particularly in the context of evolutionary theory, puzzling. In fact, because of

the pervasiveness of statistics in evolutionary biology, philosophers of biology and theoretical biologists tend to give probabilistic evolutionary explanations without seeking an alternative. It is well known that statistics hinder our understanding of the causal structure of events and it is precisely for that reason that statistics and probabilities are different! Obtaining statistical data does not imply an underlying objective probabilistic process. Tossing a coin and observing that on average it falls on 'heads' half the time does not mean that underlying the coin tossing a probabilistic process is at work! With more precision in the tossing's initial conditions, those probabilities would degenerate into 0 and 1.

Determinism is compatible with statistics and one needs to have good reasons to favour a probabilistic interpretation over a deterministic one. The reason is quite simple: a deterministic explanation of a phenomenon will always allow a higher degree of prediction and it would be irrational to choose a theory or an explanation that only predicts events with a degree of certainty when another theory would predict the event with a higher (perhaps perfect) degree of certainty. This is an *a priori* reason and therefore should not bear on whether it is *a posteriori* possible to obtain a deterministic explanation in some particular circumstances. Consider two algorithms that would allow you to predict with some certainty the lottery draw of tomorrow. Imagine that one is able to predict it with some probability, say 0.5, whilst the other predicts it with certainty. No rational lottery player would choose the first algorithm. The claim that natural selection is a probabilistic or a purely statistical phenomenon is, in my view, like choosing the second lottery algorithm without close scrutiny as to whether there is in principle a better algorithm available that does not require probabilities. As I will show throughout the thesis, I believe this better algorithm exists.

It should thus be clear that having deterministic conceptual tools at hand is by no means a barrier for understanding an indeterministic world. It should also be emphasised that my point is not to deny that evolutionary sciences require *in practice* statistics. Rather, the aim is to question the probabilistic interpretation of evolutionary theory which, we will see throughout the thesis, usually relies on the propensity interpretation of fitness. All the interpretations of the evolutionary concepts I propose in this thesis, with the exception of one (namely probabilistic drift, see Chapter 1), will account for what is classically observed and described statistically. But the difference between my interpretations and the classical interpretations is that mine will rely fundamentally and unambiguously on deterministic processes.

The question as to whether evolutionary theory is fundamentally deterministic or fundamentally indeterministic is much contested in the philosophical literature (e.g., Brandon & Carson 1996; Graves, Horan, & Rosenberg 1999; Horan 1994; Millstein 2000; Rosenberg 1988, 1994). As first noted by Rosenberg (1985, 216-219; 1988), one should distinguish the question of whether the principle of natural selection, in and of itself, is deterministic from the question of whether the 'inputs' of the process are indeterministic. It could well be the case that the process by which natural selection occurs is deterministic, but that the inputs of this process are indeterministic, for instance by the percolation of quantum events into biological processes. This should be clearly distinguished from a concept of natural selection as fundamentally inderministic. What I will demonstrate throughout the thesis and especially in Chapter 1 is that natural selection can be conceived as a deterministic process from an individual entity perspective. And because a deterministic algorithm would always be reasonably privileged over a probabilistic one, I think this is a sufficient reason for privileging my interpretation.

The lack of consensus on what the process of natural selection is and the fact it is a population level process are thus clear motivations to give a philosophical analysis to this concept and to find an alternative to the 'classical' view. In this thesis, I will propose what I believe to be the best alternative view. By 'best' I mean here 'the most consistent one'.

Another startling fact both in the biological and philosophical literature, is that most formulations of the principle of natural selection have been made in a biological context. This of course makes sense historically, for biology was the first discipline in which the concept of natural selection was used. However, as previously mentioned, it has since been used in a number of disciplines that bear no straightforward relationship with evolutionary biology. One may thus be surprised that there is no satisfactory formulation of the process encompassing all cases of natural selection in all disciplines. Without this work being carried out, one cannot even compare whether what is called 'natural selection' in discipline *x* represents the same concept in discipline *y*. To enable comparison, it is necessary to strip down the biological version of the concept of natural selection from its inessential features, namely the biological ones. This work has already been initiated by many scholars, perhaps most famously by Richard Lewontin in his article *The Units of Selection* (1970). Yet, we will see throughout the thesis that even those abstract versions of the concepts of natural selection and evolution by natural selection (ENS) remain to some extent too 'biologised', especially because they involve the concepts of fitness and heritability which themselves have not been reconceptualised adequately in non-biological terms. Thus, one aim of this thesis will be to keep only the essential features of natural selection whilst demonstrating how these features either are perfectly consistent with the biological version of the process or in some cases explain why some particular features usually associated with this process are in fact products of it.

Rather than using one single conceptual tool to reach this aim, I will instead manipulate a range of them. One will naturally be conceptual analysis, especially in the first two chapters and chapters 5 and 6. This will be reinforced by some computer simulations in chapters 3 and 4 and some mathematical formalism in Chapter 4. Chapters 2, 3 and 4 can each be seen as attacking the problem of what natural selection is in and of itself and how to distinguish it from the concept of ENS and adaptation from a different angle. Using the distinctions made in the previous chapters, Chapters 5 will question the ontological status of natural selection and 6 will revisit the 50 years old debate on levels of selection and more particularly, but not only, one of its contemporary form, namely evolutionary transitions in individuality.

My investigation of the process of natural selection will start in Chapter 1, by providing the interpretation of the concept of fitness I find the least problematic. This account is causal and relies heavily both on Bouchard and Rosenberg (2004) and Godfrey-Smith (2009b). Throughout the thesis I will make a commitment to the manipulationist account of causation. I briefly explain why in Chapter 1.

Providing a causal account of fitness will allow me to distinguish natural selection from drift. I then show how this account of fitness, once coupled to causal modelling and more particularly causal graphs, allows separating the evolutionary effects of natural selection from those of drift. It will also provide a demonstration as to why reproductive output in and of itself cannot be associated with natural selection. Further on, I argue that under this account of causation, natural selection (and drift) can be viewed as causes of evolutionary change. But contrary to what has been argued previously by several authors (e.g. Millstein 2006; Reisman & Forber 2005), I see evolution as resulting from individual-level causes not population-level ones and will defend this position in the last section of the chapter.

Chapter 2 will be the central chapter of this thesis. I start, in this chapter, by assessing the Godrey-Smith's (2009b) claim that ENS does not require the perfect transmission of types over generations. Godfrey-Smith call this position *evolutionary nominalism* (Godfrey-Smith 2009b, 35). I argue that this claim is true only if by ENS one means *evolution in which natural selection has a causal role*. Following Godfrey-Smith I call this form of ENS *'minimal* ENS'. However one might want to give a narrower definition of ENS, in which ENS means *evolution in which natural selection is the only causal process*. I call this form of ENS *'pure* ENS'. I show that evolutionary nominalism is incompatible with pure ENS for it introduces new variation in the population that one should associate with mutation rather than natural selection. Later on, I make a brief survey of the different definitions of natural selection found in the biological and philosophical literature. I show that some definitions explicitly include reproduction and/or inheritance as components of the process of natural selection, while some other do not. I demonstrate that once the process of natural selection is clearly distinguished from the more demanding ideas of a process able to generate ENS and complex adaptations, the occurrence of natural selection requires neither reproduction nor inheritance. I use a number of illustrative cases to show that the process of natural selection can occur without reproduction and/or inheritance.

In Chapter 3[1] I demonstrate more rigorously, that is, using simple individual-based simulations, some of the arguments made in Chapter 2. I show, that ENS can occur in populations of entities in which neither heredity nor reproduction are present. Interestingly, I demonstrate this by making these simulations more complex by adding only a few more parameters to the models, namely the ability to mutate at different loci. I show that both

[1] A version of this material has been published as Bourrat (2014).

reproduction and heredity are reliable Darwinian products (more precisely complex adaptations) of populations initially lacking these two properties but in which new variation is introduced via mutations. Later on, I show that if replicators are not required for ENS, they are nevertheless ultimate products of such processes of adaptation. Finally, I briefly assess the value of these models in three relevant domains for Darwinian evolution, namely origin of life studies, evolutionary transitions in individuality (for instance the transition from uni to multicellular organisms) and finally cultural evolution.

The main aim of Chapter 4[2] is the same as of Chapter 3 but this time using different conceptual tools and arguments. By the end of Chapter 2 I will have presented what is known as the classical approach to ENS, a view that distinguishes a set of necessary and sufficient conditions, usually three of them for ENS to occur. These three conditions are variation, differences in fitness and inheritance of such differences. Although this will be the approach mainly used throughout the thesis, there is another approach to ENS in the literature that relies more heavily on mathematics and of which the Price equation mentioned earlier is the most famous instantiation. While the classical approach provides 'recipes' for ENS, this formal approach partitions out selection from other causes of evolutionary changes such as transmission biases, drift and/or different levels of selection. Yet, when comparing the two approaches there seems to be a tension around the concepts of heritability and fitness. After having highlighted these points of tension and more particularly the one surrounding heritability, I assess one claim made by Earnshaw-Whyte (2012) that the classical approach to ENS is flawed and cannot be salvaged. I show that his conclusion is too strong. If one uses a concept of heritability strictly in

[2] A version of this material has been published as Bourrat (in press-a).

line with the formal equations of evolutionary change, the classical approach keeps its validity and generality. To demonstrate that the intuitive concept of heritability as stated within the classical approach to ENS is not a general one, I start from one formulation of the Price equation formulated by Okasha. I show that the concept of heritability in his formulation incorporates both the intuitive notion of heritability as a measure of similarity between parent and offspring characters and a measure of persistence which is often not recognised. I argue that persistence should be a component incorporated in the definition of heritability used in recipes for ENS in the same way inheritance is. I show that this is readily attainable and thereby dissolve any point of tension concerning heritability between the recipe and the analytical approach to ENS. In the last section of the chapter, I demonstrate how the same thing can be done with respect to the tension between the formal and classical recipe approaches to ENS surrounding the concept of fitness.

In Chapter 5, the most controversial of the six chapters, drawing on views developed in Chapter 1 that the concept of fitness relies on differences in intrinsic-invariable properties and in Chapter 2 that a pure case of ENS requires perfect transmission over time, I argue that the concept of fitness (and any measure of it) crucially depends on the concept of type. I show that because there is an indefinite number of types an entity can be categorised in by merely changing the grain (or scale) of description to describe an event of evolutionary change, the concepts of fitness and consequently natural selection both suffers from a similar problem to the reference class problem encountered in probability theory (see Hájek 2007). I call this problem the reference grain problem. After having presented the reference grain problems, I suggest that one way to address the reference grain problem, although not fully satisfactory, would be to conceptualise the process of natural selection as a model of a process that bears some

resemblance at a particular grain of description (that is, when entities have been categorised in a particular way) with evolutionary processes occurring in real populations.

In Chapter 6,[3] I argue against some interpretations of the notion of levels of selection in a context in which this notion has recently been applied, namely evolutionary transitions in individuality. More precisely I identify two major problems with a model of evolutionary transition in individuality developed by Michod and colleagues (Michod 1999, 2005; Michod, Nedelcu, & Roze 2003; Michod, Viossat, Solari, Hurand, & Nedelcu 2006), and extended by Okasha (2006, 2011), commonly referred to as the 'export-of-fitness view'. I first show that it applies inconsistently the notions of viability and fertility across levels of organisation. This leads Michod to claim that once an evolutionary transition in individuality is complete, lower-level entities composing higher-level individuals have nil fitness. I argue that this claim is mistaken, propose a correct way to translate the concepts of viability and fertility from one level to the other and show that once an evolutionary transition in individuality is complete, neither viability nor fertility of the lower level entities are nil. Second, I show that the export-of-fitness view does not sufficiently take the parameter time into account when estimating fitness across levels of selection. As a result fitness at one level is measured over a different period of time, to fitness at another level. The time over which fitness is measured represents a confounding variable that must be linked ultimately to environmental fluctuation. If fitnesses at two levels of organisation are measured over two different periods of time in fluctuating environment, this results in measuring fitness in two different environments. This is what, I argue, misleads Okasha into making the claim that the two levels of organisation represent ontologically distinct levels of

[3] A version of this material has been published as Bourrat (in press-b).

selection. I show that once fitness is measured over the same period of time across levels, the claim about two levels of selection can only be an epistemic one. I finally argue that in some cases, when in an evolutionary context, two levels of organisation are compared they belong to two different biological hierarchies (e.g., 'organism' and 'allele') which renders those comparisons problematic.

The thesis concludes by outlining some exploratory applications of the tools I have developed in the six chapters to the domain of cultural evolution. I show how these tools can be utilised in determining the extent to which cultural evolution is Darwinian.

Before going further, I must first be clear with the terms I will use throughout the thesis, even though I will come back to them when necessary. Natural selection, ENS and evolution are three distinct concepts. It is not uncommon, even today, in discussions about evolution to see the three conflated. Yet in 1930, Fisher, one of the main architect of the modern synthesis, in the preface to *The Genetical Theory of Natural Selection* (1930), clearly distinguishes the three concepts when he writes:

> Natural Selection is not Evolution. Yet, ever since the two words have been in common use, the theory of Natural Selection has been employed as a convenient abbreviation for the theory of Evolution by means of Natural Selection, put forward by Darwin and Wallace. This has had the unfortunate consequence that the theory of Natural Selection itself has scarcely ever, if ever, received separate consideration. (Fisher 1930, vii)

Following Fisher, we can see natural selection as special evolutionary process that *can* lead to a special product: ENS. It is especially important for us to distinguish evolution from natural selection because a large part of the thesis will be dedicated to the relation between the two. Missing this fundamental distinction would undeniably lead to some important confusions. We

will see in Chapter 1 that, in some respects, the same confusion exists with the concept of fitness in which fitness is confused with its effects.

However, the concept of evolution is itself quite ambiguous. Endler (1986) in his introductory chapter to *Natural Selection in the Wild*, while reasserting the point made by Fisher 56 years earlier that natural selection is not evolution, also distinguishes two notions of evolution. He writes:

> Evolution may be defined as any net directional change or any cumulative change in the characteristics of organisms or populations over many generations – in other words, descent with modification […]. It explicitly includes the *origin* as well as the *spread* of alleles, variants, trait values, or character states. […] Population geneticists use a different definition of evolution: a change in allele frequencies among generations. This meaning is quite different from the original; it now includes random as well as directional changes […], but it does not require the origin of new forms. It is roughly equivalent to microevolution. (1986, 5-7)

This way of dividing two notions of evolution echoes Godfrey-Smith (2009b) distinction between distribution and origin evolutionary explanations given in the following terms:

> When we give a distribution explanation we assume the existence of a set of variants in a population, and explain why they have the distribution they do or why their distribution has changed. Some variants may be common, some rare. Some may have been lost from the population, having been present at an earlier time. A distribution explanation explains facts of that kind. An origin explanation, in contrast, is directed on the fact that a population has come to contain individuals of a particular kind *at all*. It does not matter how many there are, or which individuals are the ones bearing the characteristics in question. (2009b, 42)

Based on Godfrey-Smith's and Endler's distinctions,[4] I want to propose that evolution can in principle be given two different but non-mutually-exclusive meanings, a mixture of which was what was meant by Darwin (1859) (see Figure 0.1). First, evolution can result solely through the introduction of new variation in the population through change of the existing substrate of the population. This is what I call *transformative evolution*. In explanations of transformative evolution we merely point to some mechanisms or principles that produce new variants. Origin explanations explain transformative evolution. Events of mutation lead to transformative evolution. Second, evolution can be strictly defined as the change of distribution of existing variants in a population, such as that which is described in a population genetics model in which there is drift and/or natural selection but no mutation. This is what I refer to as *distributive evolution*. Examples of distributive evolution include ENS and evolution due to drift[5] when there are no other evolutionary forces in the population. Distribution explanations explain distributive evolution.

[4] See also Fracchia and Lewontin (1999) who make a similar distinction.
[5] By 'drift' I mean here 'sampling error', although I will propose a new interpretation of drift in Chapter 2.

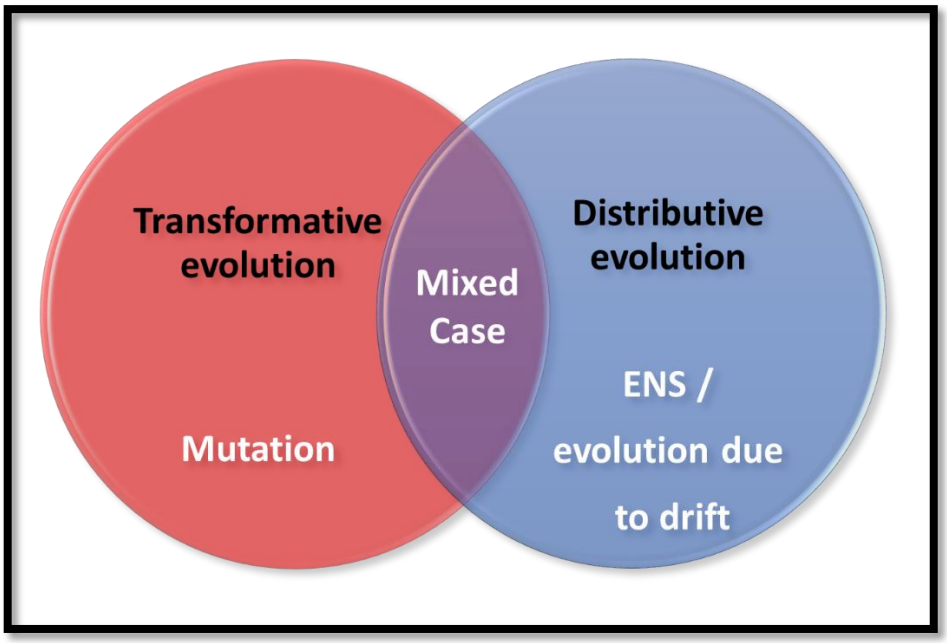


Figure 0.1. Two meanings for 'evolution'

In many cases, evolution refers both to the transformation of an existing substrate and the distribution of variants created. I call this mixed case *transformative-distributive evolution*. An explanation of transformative-distributive evolution will require components of both origin and distribution explanation. An example of transformative-distributive evolution is adaptation. An adaptive explanation will tell us how a complex trait such as an eye has been able to evolve from a population of entities[6] in which none of them has a fully developed eye, to a population in which almost all if not all, have fully developed eyes.[7]

I define the degree of complexity of a trait, following a definition of complexity provided by McShea (1996) and McShea and Brandon (2010) as "the amount of differentiation among its

[6] For now let us suppose they reproduce asexually and in discrete generations.
[7] A very good example of such explanation can be found in Nilsson and Pelger (1994) and a less technical but lengthier version in (Dawkins 1996).

parts or, where variation is discontinuous, the number of part types"[8] (McShea & Brandon 2010, 45). This definition is a definition of complexity in *structure* rather than *function*. An eye is a complex structure but it also has a function that confers an advantage to its bearer: it is an adaptation. A complex adaptation is thus a special case of complex trait in which not only the trait is structurally heterogeneous, but the heterogeneity produces a function.

It is worth noticing that distributive evolution alone can neither produce novelty nor complex traits such as a lensed eye. In fact, under this conception of evolution there is no production of new variation. Without new variation being introduced, natural selection,[9] if there is no other evolutionary process in place, will inexorably deplete the existing variation until only one variant remains. This is one consequence of a simplified version of Fisher's (1930) fundamental theorem of natural selection once correctly interpreted (Edwards 1994).[10] Although I will take an approach different from that of Fisher, the topic of the effect of natural selection in a population in which natural selection is the sole evolutionary process will be discussed in Chapter 2. In many cases, throughout the thesis, distributive evolution is what I will mean by 'evolution'. This is because it is the most straightforward and simpler notion of evolution that is linked to the process of natural selection (and drift).

Transformation evolution can, in contrast to distributive evolution, lead to complex structures. In fact, the production of complex structures is perfectly possible by the sole

[8] McShea and Brandon refer in their definition to the complexity of an organism, but the definition can straightforwardly transposed to traits.
[9] The same can be said about drift.
[10] For correct interpretations of the theorem see also Price(1972b), Ewens (1989) and Okasha (2008).

production of variation.[11] This is the thesis argued by McShea and Brandon (2010). Yet, we should not expect origin evolution by itself to produce traits with functions. Rather, we should expect it to result in an increase of variation or diversity in a population unless some factor(s) explains the particular direction of evolution.

The mixed case, namely transformative-distributive evolution, is without doubt the most interesting form of evolution. In fact by combining both transformative and distributive evolution, evolution can result in adaptation when the production of new variation and natural selection are in a specific equilibrium. This equilibrium must allow at least part of what has been retained by natural selection not to be destructed by mutation (Eigen 1971; Williams 1966). When that equilibrium is reached in a population this population has the property of *memory* (more on memory in Chapter 2). Adaptation is a special case of ENS, in which one of the properties of the population is memory.

In many adaptive explanations the production of variation is considered as random and thus the only known difference maker that explains adaptation is natural selection.[12] Drift once combined with mutation does not produce adaptation. In fact, although it can be causally involved in the production of complex structures, once combined with mutation it does not typically produce functional structures.

[11] Note however that this is not necessarily the case. Transformative evolution can appear as distributive evolution if the same variation is produced over and over. This is one of the reason that led Lewontin to make the distinction between 'reproduction *of*' and 'reproduction *by*' (see Ariew & Lewontin 2004).

[12] Note that this does not mean that natural selection is the only causal factor of adaptation. Rather it means that natural selection is the only one we can point to explain the particular and systematic direction of evolution of a given trait. We will see in the Conclusion of the thesis that the same cannot be said for cultural evolution.

With these definitions in place, I now begin my investigation of the concept of natural selection starting in Chapter 1 by characterising it in causal terms.

1. The Causes of Evolutionary Change, an Individual-Level Perspective

1.1. Introduction

In the Introduction of this thesis, I have mentioned several evolutionary processes and asserted that they can cause evolution. Mutation, I have claimed, causes transformative evolution while natural selection and drift, once they are isolated, cause distributive evolution. This description is in line with population genetics, the mathematical branch of evolutionary theory, that classically distinguishes four causes or forces of evolution:[13] natural selection, mutation, migration and drift (e.g. Crow & Kimura 1970; M. Hamilton 2009; Hartl & Clark 1997).[14] Yet, what constitutes an evolutionary cause or force is not immediately clear. It has been and still is subject to an important debate in philosophy of biology. Most of this literature focuses on natural selection and drift while leaving out mutation and migration. In this chapter I follow suit. I will treat mutation in more detail in Chapter 2. Migration will not be treated in the thesis because I will consider that evolutionary change is a phenomenon that arises in an isolated population without external interventions of other populations. Thus, migration is not relevant at the level of generality I will be working. Furthermore migration is usually not mentioned in the formal descriptions of evolutionary change. For the interested readers, Kerr and Godfrey-Smith (2009) propose to extend the Price equation, a formal description of evolutionary change mentioned in the Introduction of this thesis (see Chapter 4 for more details), to cases of evolutionary change

[13] Distributive evolution and transformative evolution being undistinguished here.
[14] This division into four forces or causes is to some extent arbitrary. Some authors would, for example, add sex or recombination as other forces or causes.

that incorporate arbitrary causal connectivity between ancestors and descendants of a population. This formalism accounts for migrations processes and fill the void left by the classical equations.

It is nearly uncontroversial to both evolutionary biologists and philosophers that natural selection is a phenomenon associated with heritable differences in fitness between the members of a population (Godfrey-Smith 2009b; Sober 1984). Although all the protagonists of the debate on the causal nature of natural selection agree on this point, they regard the concept of cause and its links with fitness and natural selection in different ways. The philosophical landscape on this issue can roughly be divided up into three camps. Under what I will refer as the *statisticalist* view[15] it is simply misleading to consider natural selection and drift as forces or causes. The statisticalists claim that what we call natural selection is a mathematical aggregate of unique events happening to individuals forming populations (e.g. Matthen & Ariew 2002; Walsh, et al. 2002). Each type of organism in a population is assigned a *trait-fitness* value – in its most abstract form an expected growth rate – which is the consequence of different events occurring at the individual organism level such as deaths, births or mating. Yet, those events cannot be equated to natural selection because trait fitness is a non-causal (statistical) property of types at the population level and consequently differences in trait-fitness in a population entail ENS mathematically (statistically) rather than causally (Matthen & Ariew 2009).

Bouchard and Rosenberg (Bouchard & Rosenberg 2004; see also Rosenberg & Bouchard 2005) offer a very different view, in direct opposition to that of the statisticalists. Not only do they argue that natural selection is a causal process, they also claim that fitness should not be understood as a population-level statistical property. They conceive of fitness as a relational

[15] For different arguments from the statisticalist view see: Matthen & Ariew (2002, 2009), Walsh (2000, 2007, 2010), Walsh, Lewens & Ariew (2002).

property between *individual*[16] *entities* forming populations. This property must be distinguished from a growth rate or reproductive output, since Bouchard and Rosenberg view a difference in growth rate between two types of entities not as a difference in fitness, but as a *consequence* of a difference in fitness (see Rosenberg 1985, 158-160 for the view that fitness is measured by its effects). In other words, growth rates are proxies for fitness (a proxy of fitness is also called *realised* fitness).[17] As Bouchard and Rosenberg (2004, 710) put it: "selection [is] a contingent causal process in which individual fitnesses are the causes and subsequent population differences are the effects". I call Bouchard and Rosenberg's view the *individual-level-cause* view (hereafter the 'ILC view').

A number of authors have proposed a third view, namely that natural selection and other evolutionary processes are causes of evolutionary change, not at the individual level, as argued by Bouchard and Rosenberg, but at the population level (e.g. Millstein 2006; Reisman & Forber 2005; Stephens 2004).[18] The main claim of the proponents of the *population-level-cause* view (hereafter the 'PLC view') is that since natural selection and drift can systematically be manipulated at the population level, they represent genuine causes at that level under some legitimate account of causation. An important point to note is that both the ILC and PLC proponents agree that the consequences of natural selection at are the population level. Their disagreement lies in whether the causes of natural selection are population or individual causes.

The main aim of this chapter is to introduce a new conceptual tool to this literature, namely causal modelling, and to use it to elaborate and defend the ILC view. I will also use other

[16] In this chapter, 'individual' should be understood as 'entity below the level of the population'.
[17] See the problem of reifying reproductive output as fitness in Chapter 5.
[18] See also Northcott (2010) and Shapiro & Sober (2007), Huneman (2012).

concepts from the manipulationist account of causation (see Woodward 2003, 2013) to further defend this view (Section 1.4), such as the criterion of invariance (also called stability) in causal explanations developed by Woodward (2000, 2003, 2010). Before starting, I need to make a few remarks. I wrote in the opening sentence of this chapter that natural selection and drift are usually regarded as 'forces' or 'causes' of evolutionary changes by population geneticists. Yet so far, I only have used the notion of cause, and will do it throughout the remaining of the chapter while I will be less careful in further chapters, using 'force' and 'cause' indistinguishably. I make this commitment in this chapter mainly because the concept of cause, although far from being unambiguous, is less vague than the concept of force when 'force' is taken to be something different from Newtonian force. In fact, there are no developed philosophical accounts of what makes a 'force', whereas by sticking to the concept of cause I can draw upon previous work on causation and apply it to evolution.

Although the notion of cause is less vague than the notion of force, it is not unproblematic. This is reflected in the fact that there are many different competing accounts of causation in philosophy. Each of these accounts has (at least) some problems and suffers from counterexamples. In a recent survey of the literature, Hall and Paul (2013) distinguish four families of such accounts, namely regularities accounts, counterfactual accounts, probability accounts and transference accounts. Here is not the place to discuss the difference between each of these accounts. Yet this ambiguity plays a role in the debate over the causes of evolutionary change. Huneman (2012) notes that the statisticalists rely on a particular transference account of causality similar to the one proposed by Salmon (1994, 1998), according to which there is causation when there is transference of some physical quantity between a cause and its effect. In contrast, the PLC camp relies on counterfactual accounts of causation. These different

commitments allows the statisticalists to make the claim that all the causal work occurring during an event of evolution lies at the level of individual entities rather than population, and to point out that these individual events (births, deaths and reproduction) do not themselves constitute natural selection. Walsh (2000) for example considers that natural selection is a pseudo process rather than a genuine process. The PLC camp, on the other hand, claims that following the different counterfactual accounts of causation (e.g., manipulationist), natural selection and drift are causes of evolutionary change.[19]

Bouchard and Rosenberg (2004) do not explicitly state what account of causation they follow. They argue that distinction between natural selection and drift cannot be substantiated from a statistical perspective only,[20] and that natural selection is a causal process relying on *individual* fitness comparisons, not on (non-causal) population level properties such as *trait* fitness. Thus, for them, no notion of population level cause is necessary. Comparisons are at the heart of any counterfactual dependence and assessing whether a property or an event is a difference maker could not be made without them. In fact, to make causal claims under a counterfactual account of causation, assuming that time is asymmetric, is essentially to make a specific comparison between two situations that grants a causal relation. It involves comparing the presence/absence of *c*[21] at time t_1 with the presence/absence of *e* at a later time t_2. In its most general form *c* and *e* can be properties of objects, events or processes. When *ceteris paribus c* and *e* are present *and* had *c* been absent *e* would also have been absent, it is safe, under a counterfactual

[19] Although see Millstein (2013) who uses a Salmonian process account of causality.
[20] Which is a claim clearly endorsed by Walsh, Lewens, & Ariew (2002).
[21] Or under a contrastive account of causation (see for example Schaffer (2005) for the general case or Northcott (2010) in the particular context of evolutionary theory) two different situations c_1 and c_2 rather than presence and absence of *c*.

account, to claim that *c* is a cause of *e*.[22] Thus because Bouchard and Rosenberg's account is both causal and comparative, I will assume that their account relies on a counterfactual account of causation similar to the ones developed by the PLC camp but with a focus on individual properties and events rather than population level ones.

If my interpretation of Bouchard and Rosenberg's account of causation is correct, it seems that the statisticalists initially disagreed both with ILC and PLC camps on what constitutes a cause in evolutionary theory. It is was thus not surprising that they disagreed about whether natural selection and drift are causes of evolutionary change. Although the debate between the statisticalists and the causalists is not yet settled, it seems that in recent days, the statisticalists have acknowledged the virtue of Woodward's manipulationist account of causation in the debate over the causes of evolution (see for example Matthen & Ariew 2009). Furthermore, the concept of cause in science is very often cashed out in terms of counterfactuals (or surrogates of them such as controls, see more on this in sections 1.6 and 6.4). For those reasons, in this thesis I have chosen the counterfactual over the transference approach to causation as the relevant one for deciding whether natural selection and drift are causes of evolution.

With these important remarks in place, I will now argue that the ILC view, once developed with the help of the tools of the causal modelling literature (e.g. Pearl 2000), is superior to both the statisticalist and PLC views. To do so, the remainder of the chapter will be divided into five sections. In Section 1.2, I start by presenting some of the arguments motivating Bouchard and Rosenberg's view that the theoretical underpinning of evolutionary theory, and

[22] Although the *ceteris paribus* clause is regarded as problematic for some in the literature (Reutlinger, Schurz, & Hüttemann 2014), it is important here because it eliminates any potential confounding variable that could be a cause c_2 of *e*. Confounding variables will be a recurrent topic in this chapter and more generally in the thesis.

more particularly the concept of fitness, is deterministic (or sufficiently close to deterministic) and not fundamentally probabilistic as often argued. From there, I present a slightly revised version of Bouchard and Rosenberg's claim that fitness should be defined in causally upstream terms from reproductive output. In Section 1.3, drawing upon one of Godfrey-Smith's (2009b) distinction, I propose a systematic method in individual rather than population terms to distinguish natural selection from other evolutionary causes, and more particularly drift. In Section 1.4, I briefly present causal graphs and depict, from a causal modelling perspective, the problem of teasing apart natural selection from drift with reproductive outputs of organisms as the only available information to the observer that is, without a notion of fitness defined in causally upstream terms from reproductive outputs. In Section 1.5, I show how combining causal graphs and the method proposed in Section 1.3 to partition off natural selection and drift in individual terms can fully explain evolutionary change at the population level. I discuss the methodological and epistemic problems inherently associated with this view, but ultimately conclude that those problems should be kept separate from conceptual considerations. In Section 1.6, I show why the version of the ILC view I propose is superior to the PLC view proposed by Reisman and Forber (2005) and Millstein (2006).

1.2. What fitness is: Bouchard and Rosenberg's causal account of natural selection revisited

To motivate their account of natural selection as a causal process relying on individual fitness differences, Bouchard and Rosenberg argue that the common assumption that fitness is an objectively probabilistic property, that is, a propensity (usually the expected number of

offspring produced after one generation) which has been a very popular concept of fitness in the philosophy of biology (Beatty & Finsen 1989; Brandon 1978; Mills & Beatty 1979) is problematic. In fact they claim that if fitness is an objective probabilistic property, either some quantum events must percolate up in biological processes and have significant effects on fitness differences or there must be brute propensities at the level of organismal fitness difference. Both interpretations, they claim, are problematic if applied to evolutionary theory.[23] Since Bouchard and Rosenberg wrote their article, Strevens (2003) has proposed a new account of probabilities (see also Strevens 2011) and seems to give a third alternative to the two proposed by Bouchard and Rosenberg. If this account proves correct, this could have interesting implications for evolutionary theory. Presenting Strevens' account in details would lead us beyond the scope of this chapter and of this thesis more generally. In Subsection 1.3.4, I provide a very brief introduction to Streven's (2011) account of probability and some of its limitations when providing my account of fitness. For now, I will make two remarks. First, it is not clear whether Strevens's account represent a real alternative to other accounts of probability. Second, it is also far from clear that his account would be successful for the concept of fitness.

Bouchard and Rosenberg note with respect to the view of quantum events percolating up in biological processes, that although this represents a possibility and is at the origins of some mutations for instance, there is no independent evidence that such events have a significant role on fitness differences between biological organisms. The second proposition, that there are brute propensities at the level of organismal differences is even more problematic since to be true, it

[23] For defences of indeterminism in evolutionary theory see Brandon & Carson (1996), Glymour (2001) and Stamos (2001). For defences of determinism in evolutionary theory see Rosenberg (1994), Horan (1994), Graves, Horan and Rosenberg (1999). For arguments in favour a view of compatible with both determinism and indeterminism see Weber (2001).

would requires the emergence in a strong sense of properties at that level, a commitment only a few biologists and philosophers are ready to make. This reasoning leads Bouchard and Rosenberg to consider that the probabilities attached to fitness values are, for a large part, the result of the epistemic state of the observer. Furthermore, one could add that even if some events are brute probabilistic facts about the world and that they have important consequences in terms of differences in fitness, this would only show that some 'inputs' in the process of natural selection are probabilistic, not that the process (or mechanism) in itself is fundamentally a probabilistic one. A similar point is made by Rosenberg (1988, 1994, 2001).

Bouchard and Rosenberg also respond to some statisticalists (e.g. Matthen & Ariew 2002) who have proposed that the probabilistic character of the process of natural selection is analogous to the probabilistic character of the second law of thermodynamics. Bouchard and Rosenberg argue that this represents in fact a disanalogy. Indeed, entropy, contrary to fitness is a property of ensemble of parts that cannot be traced back to some property of the parts and it seems that, based on the theoretical assumptions of thermodynamics, no one could make this reduction. Fitness on the other hand is a property of the parts of an ensemble (a population of organisms) and is not subject to the same irreducibility problems as entropy: although the emergent nature of probabilistic events at the ensemble level in the case of thermodynamics seems unavoidable, it is not the case for natural selection, since it can ultimately be explained by differences in individual fitness. The only commitment of that proposition is that fitness is a property of organisms, a commitment many biologists and philosophers (but not the statisticalists) are ready to make.

To sum up, according to Bouchard and Rosenberg, although there are inherent difficulties associated with any interpretation of probabilities (see Hájek 2012 for a survey of the

limitations of the different interpretations of probabilities), interpreting fitness as a propensity is too strong a commitment and they favour an interpretation as the result of epistemic limitations. Second, although there exists some similarities between thermodynamics and evolutionary theory, it does not entail the claim natural selection is irreducible to some individual causal process as the second law of thermodynamics is.

Now, based on their critique of the different interpretations given to probabilities in evolutionary theory, how exactly do Bouchard and Rosenberg conceive natural selection as being an individual level causal process? They tell us that to know whether natural selection (as opposed to another evolutionary process) is causally responsible for an evolutionary change observed in a population, one must make comparisons between individuals of this population taken two at a time and observe for each comparison which one has the higher fitness. If one individual has a higher fitness than another, then *ceteris paribus*, assuming a positive inheritance between parent and offspring, its type should be favoured and spread in the population (assuming there is heritability of the trait under selection pressure). But fitness is not a measure of reproductive output in Bouchard and Rosenberg's account. Rather, they conceive a reproductive output (or more precisely a difference in reproductive output between two organisms) as a consequence of what they call 'ecological fitness'.[24] They define ecological fitness as follows:

> [*a* and *b* are two individual organisms,] *a* is fitter than *b* in [environment] E = *a*'s traits result in its solving the design problems set by E more fully than *b*'s traits. (2004, 699)

If an individual organism *a* solves the design problems set by the E more fully than another organism *b*, then this will, under the same environmental background, lead *a* to produce

[24] Matthen and Ariew (2002) refer to this concept of fitness as *vernacular* fitness.

more offspring than *b*. Fitness under this account is thus a relational property between two organisms and differences in reproductive output represent merely what evolution is, namely that one type of organisms has a particular reproductive output and another type another reproductive output. Differences in reproductive output do not represent explanations of evolutionary changes but merely state the same facts from a different perspective.

We can see here that the proposition of Bouchard and Rosenberg represents a solution to the famous problem of fitness as being a mere tautology. In looking for the *causes* of evolution, I take it, Bouchard and Rosenberg mean that we need to define fitness in a currency that is 'causally upstream' from reproductive output so that these causes are distinguished from their consequences and thus do not render the account of natural selection tautologous. By causally upstream from reproductive output, I refer to variables that once manipulated will have an effect on the reproductive output of a given organism. Bouchard and Rosenberg propose 'solution to design problems' to be the most general and adequate upstream causes of reproductive output that can be associated to natural selection.

As they acknowledge, 'solution to design problems' is a particularly challenging concept for it is difficult, among other things, to both measure and make precise how many design problems a given organism has to solve. But whether their account in terms of solutions to design problem is successful or not, I do not regard it as fundamental to their position. What I regard as the key to their account, as I have already pointed out, is that fitness must be defined in causal terms upstream to reproductive output, not that one particular currency to measure fitness should be privileged. The motivation behind finding a general currency for measuring fitness is to be able to compare the fitness of organisms from different taxa. This was one of the projects undertaken by Van Valen (1976) and pursued by Bouchard (2004, 2008, 2011). Although I regard

this project as an important one, for my purposes, I define the principle of natural selection as follows:

> **(PNS)** The process of natural selection will occur in a population if and only if there is at least one difference in intrinsic-invariable property, in an environment **E**, between two or more individual entities of the population that can be causally linked to a downstream difference in reproductive output (or more generally relative growth).

What exactly underlies differences in intrinsic-invariable properties and why only those differences can be attributed to natural selection will become clearer in Section 1.2. For the moment, the reader can replace 'intrinsic-invariable properties' with 'biological properties'. Because what constitutes a difference presupposes at least two entities to be compared the claim that, in the ICL framework, the fitness of a given individual is X does not make sense if it is not relativised over at least another individual. To which other individual a given individual is compared will define the currency of fitness for this particular comparison and the trait that is compared (a trait can thus be, under this account, a whole organism). Some comparisons and currencies might be more 'natural' than others (because for instance the organisms compared belong to the same biological population), but ultimately this is a choice made by the observer. If one intends to have a universal currency to measure fitness to be able to compare organisms from different taxa, the suggestions from Van Valen (1976) that the best approach to this end is based on the amount of energy they control seems to be a possible option.[25]

[25] Van Valen (1989, 12) for instance, using what he believes to be the best universal currency for fitness, namely control of energy, writes that "[using energy control as common currency of fitness we can look at natural selection within entire communities in a natural and causal way. We can compare the fitness of a bird with that of a fungus."

Fitness, under my account, is thus a relation between at least (any) two organisms or more generally entities in their ability to reproduce the type they belong to after a given period of time (in the simplest cases reproductive output is a means to that end), assuming here reproduction is perfect. Ability should not be understood here as propensity. It rather means that given the same environmental background, the fittest of two organisms in this environment will have a property, causally upstream to its reproductive output, allowing it to reproduce its type more than the other over a given period of time. We will come to the concept of typehood several times in this thesis.

The payoff of adopting such a definition of fitness might not be immediately obvious to the reader. In sections 1.4, 1.5 and 1.6 , I show using a functional causal modelling approach to causality akin to Pearl (2000), that differences in properties between the members of a population causally upstream from reproductive output, as I have defined it, are fundamental to distinguish ENS and from evolution as the result of drift. But before going further, it is essential to explain why only differences in intrinsic-invariable properties as mentioned in definition of the PNS, are the only ones that can be the basis for natural selection. More reasons will be provided in Chapter 2.

1.3. Fitness as a difference in intrinsic-invariable properties

In this section, following the definition of PNS given in the previous section, I argue that only entities with differences in properties that are intrinsic to the entities of a population and that do not vary in E, that is, that are part of the entities and stable at any time and place in E

See also Bouchard (2008, 2011) for his view that the universal currency of fitness is a propensity to persist through time.

(with E defined both in spatial terms and over time), can undergo natural selection. To make this point, I start from an account of fitness proposed by Godfrey-Smith (2009b), but then show some problems associated with it and propose some revisions to salvage it.

1.3.1. Godfrey Smith on natural selection and drift

In his recent book, Godfrey-Smith (2009, 53-63) develops a principled method for distinguishing natural selection from drift from the perspective of entities forming a population that is perfectly compatible with a fully deterministic set-up. His motivation for doing so is that, in their classical versions, natural selection and drift rely on the notion of expected fitness or expected reproductive output. Yet, the classical interpretation of expected values both in philosophy of biology and theoretical biology is the propensity interpretation of probability which, as we have seen, is highly controversial. Thus some of Godfrey-Smith's and Bouchard and Rosenberg's motivations for revising the classical account of fitness seem to be similar. With his framework, Godfrey-Smith intends to replace the classical interpretation by another one that does not rely on the expected values.

While the classical notion of natural selection relies on differences in expected reproductive outputs between entities of different types, the classical notion of drift is more difficult to grasp. Classically, the notion of drift is often associated with the notions of 'sampling error', 'indiscriminate sampling' or 'random sampling' (e.g. Beatty 1984; Brandon 2005; Crow & Kimura 1970; Gillespie 2004; M. Hamilton 2009; Hartl & Clark 1997; Millstein 2002) although others, such as Rosenberg have proposed a different interpretation, namely that drift is a useful fiction (Rosenberg 1994). In population genetics, differences in reproductive output occurring by 'chance' (random drift) or 'accident' between entities of a population and leading to

evolutionary change are often considered as synonyms for evolution due to drift. Although this statistical description is pragmatically useful in evolutionary theory, it does not capture substantially, that is causally, what the concept of drift amounts to. In fact, although the notions of 'chance' and 'accident' can be quantitatively apprehended via the notion of *expected value*, the interpretation of what expected values on reproductive output mean is highly problematic. The propensity interpretation of fitness (Beatty & Finsen 1989; Brandon 1978; Mills & Beatty 1979; Sober 2002) we encountered earlier is an attempt to make sense of those expectations. It tells us that fitness is a tendency or disposition in the same way that 'resistance to impact' is one. Entities are 'expected' to produce a certain number of offspring in a given environment in the same way that a glass is expected to break under certain conditions.

Yet, propensity interpretations of probability, as we have seen, are controversial for they suffer from the charge of being causally empty in evolutionary theory (e.g. Byerly & Michod 1991) and more generally empty accounts of probability without a clear explanation of what propensities are (Eagle 2004; Hájek 2012). To capture more substantially the notion of drift, we need conceptual tools that do not rely on a particular (and controversial) interpretation of probability, and that are compatible with a fully deterministic world since it is not clear whether quantum events percolating up to biological processes or brute biological propensities have an important role for natural selection (Bouchard & Rosenberg 2004). Godfrey-Smith (2009, 53-63) precisely develops tools that are compatible with those requirements. He holds the view that the distinction between natural selection and drift has something to do, among other things (more on this below) with the notions of intrinsic and extrinsic properties of entities forming populations. Although I believe Godfrey-Smith's account has some problems, I think it is on the right track.

Godfrey-Smith proposes that when in a population there is variation in intrinsic properties between its members, and it leads them to have different reproductive outputs, the resulting evolutionary change should be attributed to natural selection. Conversely, when this difference in reproductive output is due to differences in extrinsic properties, the evolutionary change resulting should be attributed to drift. He defines an intrinsic property as a property that, contrary to an extrinsic one, does not depend on the existence and arrangement of other objects. A good example of intrinsic property is the chemical composition of an organism. Examples of extrinsic properties include being in a particular location or someone's cousin.

The rationale behind this view is that with extrinsic properties as opposed to intrinsic ones, the differences in reproductive outputs they are causally responsible for cannot be systematically attributed to their bearers. In some sense intrinsic properties are constitutive of an entity while extrinsic are not. Another way to understand this distinction is to use counterfactual dependences. When evolutionary change is due to drift, had at least some circumstances the entities are found in been otherwise, the extrinsic properties of their bearers would have been different and led these entities to have different reproductive outputs. When evolutionary change is due to natural selection, had the entities themselves been otherwise (that is, their intrinsic properties), then their reproductive outputs would have been different, independently from the circumstances these entities might have been found in. By associating extrinsicness with drift, we recover the classical notions of accident and chance classically associated with drift, where chance or accident can be understood as 'that does not depend on the entities forming the population'. It follows that two types of entities with different extrinsic properties (that do not ultimately causally depend on intrinsic properties), even in a fully deterministic set-up, should exhibit drift. Under this view, the statistical nature of a type's reproductive output should not be regarded as

the result of indeterminism, but as a measure of the extent to which the member of a population differ in the extrinsic properties insofar as they lead to differences in reproductive outputs.

This view, if correct, has one important consequence: it follows from it that had each extrinsic property been the same for all the members of a type, each member would have had the same reproductive output. This means that this notion of drift does not predict that under a perfectly homogeneous environment for each entity of a population, smaller populations would exhibit higher levels of drift as predicted by a statistical definition of drift (I will come back to this point in Section 1.6). However, it explains why this is observed in most cases. To see this clearly, let us suppose a population composed of two types of entities (that is, with one difference in intrinsic properties) that can be found independently in several different states of the environment. Each state is supposed to lead to different consequences for the reproductive output of an entity. If a population is composed of a small number of entities and there is nothing intrinsic to the entities that determines in which state of the environment it should be, it is very unlikely that for each entity of a given type in a particular state of the environment, there will be another entity of the other type in the same state of the environment. However, the probability of finding matching entities of each intrinsic type in the same environmental state will increase as population size increases. When the population size is infinite, any given entity of one type in a particular environmental state will have a matching entity of the other type in the same environmental state. At that point, any difference in reproductive output between the two types will be attributable to a difference in intrinsic properties only. This notion of drift is thus compatible with its classical statistical interpretation. It should be clear that real populations are usually composed of more than two types and many more than several 'micro-states' of the

environment. Yet, the simple reasoning provided here remains valid for populations with any number of types and environmental states.

Godfrey-Smith's (2009) view on drift is actually more complex than the one presented earlier. In fact the dependence of reproductive output on intrinsic properties is only one of five important features to characterise Darwinian populations, that is, populations able to exhibit ENS[26] (for a more complete assessment of Godfrey-Smith's notion of 'Darwinian population' see Chapter 2). Godfrey-Smith calls the feature measuring the dependence of realised fitness (for our purpose in this chapter this is equivalent to reproductive output) on intrinsic properties 'S'. Along with S the four other features are fidelity of heredity 'H', abundance of variation 'V', competitive interaction with respect to reproduction 'α' and what he calls continuity 'C'. For Godfrey-Smith the parameter C is also involved in drift. But before presenting what this feature entails and why Godfrey-Smith associated it with drift, it is useful to digress a bit and briefly present the three other features.

The features H and V underlie quite straightforward concepts, and represent respectively how reliable heredity is in the population and how much variation there is between the different entities of the population. Without variation there cannot be ENS because there is nothing to 'choose' from. If there is no or a very weak heredity of traits between parent(s) and offspring, then natural selection will have no durable consequences on the population. At least that is what I will assume for now, but I will come back to the notion of inheritance first in Chapter 2 and then in chapters 3 and 4.

[26] According to Godfrey-Smith (2009, 63), this list is incomplete and could include other features.

While the concepts underlying H and V are relatively straightforward, the ones underlying α and C are a bit more demanding. A simple way to understand what 'α' represents using Godfrey-Smith's own words is "the extent to which adding reproductive success to one individual reduces another's" (2009, 52). With this parameter Godfrey-Smith is introducing the Malthusian notion of competition between the members of a population. Finally 'C' represents the level of change in reproductive output induced by small changes in an entity's phenotype (2009, 57). For each of the five parameters in Godfrey-Smith's framework, the higher the population scores on this parameter, the more the population is a paradigmatic case of a Darwinian population, leading to paradigmatic form of ENS as opposed to marginal case of a Darwinian population leading to a marginal form of ENS. Godfrey-Smith represents three of the five parameters (H, C and S) in what he calls a 'Darwinian space' (2009, 63-67) which allow to visualise how far from an ideal paradigm case of a Darwinian population is an actual population. I present a simplified version of it in Figure 1.1. Although the approach to natural selection in terms of a similarity space is very useful, I will examine some of its problems in more details in the next chapter and more particularly in relation to heredity.

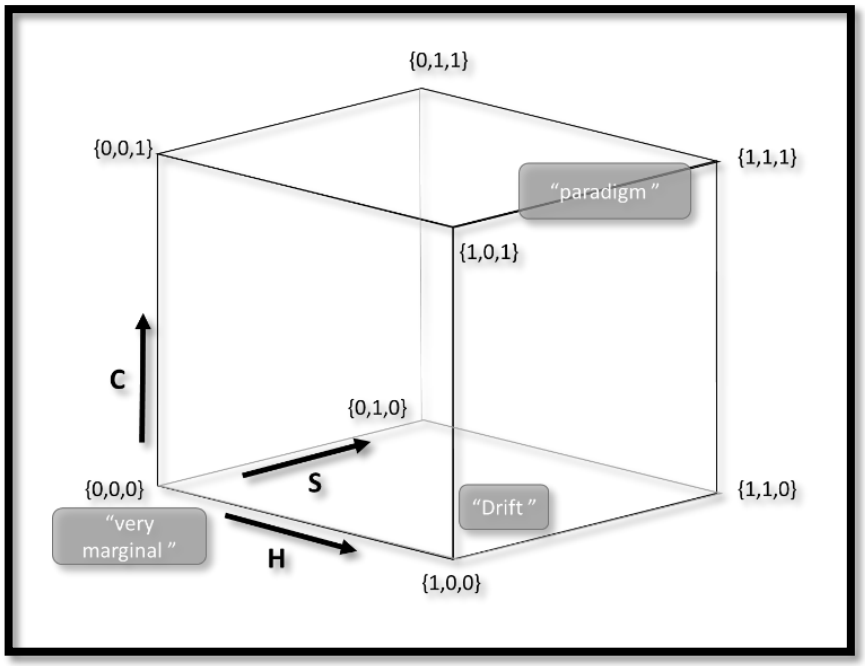


Figure 1.1. Godfrey-Smith's Darwinian space, reproduced and simplified from Godfrey-Smith (2009b). (S: dependence of realised fitness (reproductive output) differences on intrinsic properties; C: continuity; H: fidelity of heredity.)

Going back to drift, Godfrey-Smith holds that a very low C, that is, when small changes in the phenotype lead to large changes in fitness, should be associated with drift especially when S is also low. This can be visualised the Figure 1.1. I disagree with Godfrey-Smith on this point for several reasons. First, he considers that small differences in "everything about an organism [read 'entity']" (2009, 61) (that is, both intrinsic and extrinsic properties) can in some cases lead to large differences in reproductive outputs and should thus be associated with drift. This is problematic. In fact, in cases of large differences in reproductive outputs due to small differences in extrinsic properties, those cases should, in my view, not be considered as of a different nature from cases of differences in extrinsic properties. They represent merely a subset of the cases in which differences in extrinsic properties lead to differences in reproductive outputs. Different levels of C, it seems, modulate the level of drift resulting from difference in extrinsic properties,

but they do not change its nature. Note furthermore that Godfrey-Smith does not provide any principled reasoning as to what thresholds should be used to considered 'small' differences between entities' properties and 'large' differences in reproductive outputs.

Concerning small differences in intrinsic properties leading to differences in reproductive outputs, I find Godfrey-Smith's proposition even more problematic. In fact, these differences should be regarded, as per his own distinction with the parameter 'S', as differences due to natural selection, not drift. If Godfrey-Smith's distinction between intrinsic and extrinsic properties is right, it cannot be the case that small differences in intrinsic properties are sometimes attributed to natural selection and sometimes attributed to drift especially since, in this case too, Godfrey-Smith does not propose a principled way to distinguish which 'small' intrinsic differences should be attributed to natural selection and which ones should be attributed to drift.

Second, it seems that the conceptual work made by Godfrey-Smith's parameter C helps best conceptualising the conditions under-which natural selection can lead to complex adaptations, a point I will briefly develop in the next chapter. If slight differences in properties whether intrinsic or extrinsic lead to dramatic changes in reproductive outputs, that is, if the population exhibits a low C, it is very unlikely that anything like a lensed eye could ever evolve. Godfrey-Smith's (2009) project, as I see it, is a (successful) attempt to delimit the conditions under which natural selection will lead to adaptations as classically understood by evolutionary biologists, that is, through cumulative selection. Yet, the question as to whether natural selection (or drift) is responsible for the pattern observed (complex structures) in populations and the question of whether natural selection (or drift) is an evolutionary process occurring in the population are quite different and should be distinguished. In fact, we will see in the next chapter that the process of natural selection does not necessarily lead to the production of complex

adaptations. I think that populations with low C are instances of such cases. To me, a low C predicts that natural selection cannot lead to the evolutionary process responsible for complex adaptations, not that evolution is drifty, even if empirically ENS without complex adaptations might be extremely hard to distinguish from evolution due to drift.

1.4. Problems with the intrinsic/extrinsic distinction

1.4.1. Drift as intrinsic-variable properties

One general problem with Godfrey-Smith's framework in terms of intrinsic and extrinsic properties, fortunately not a fatal one, is that although I see this distinction as a valuable one to separate natural selection from drift, I believe it is incomplete. In fact, any biological property, say for instance 'height', is obviously diachronically the result of the interaction between the bearer of the property and its environment. Had a given organism been put in a different environment from birth and its height might have been very different. Godfrey-Smith's distinction between intrinsic and extrinsic properties only accounts for 'synchronic' dependences on reproductive output and does not explicitly account for more 'diachronic' dependences on reproductive output. Yet in my view, such dependences matter a lot with respect to natural selection and drift.

To see why this is the case, consider the following intrinsic property of an organism, 'amount of fat'. The amount contained by each individual is generally different between each organism of a population and this might have consequences on reproductive outputs. Using Godfrey-Smith's framework, all the differences in reproductive output due to differences in

amount of fat contained by organisms should be attributed to natural selection. The problem here is that there are cases in which containing a certain amount of fat leading to a different reproductive output, when compared to another organism, should intuitively be attributed to drift. Imagine for example that two organisms have different reproductive outputs due to the fact they contain a different amount of fat. But the difference here is considered as the result of different life histories that cannot causally be traced back to any of their biological properties. For example suppose that the two organisms have the same susceptibility to a disease **V**. Yet, one gets **V** and has to spend more energy to eliminate it. To do so it burns a larger amount of fat than the other organism. As a result the two organisms have different amounts of fat and produce different numbers of offspring.

This case and similar ones – I could have used 'height' for instance – should be considered as cases of drift because they intuitively look like the result of some chance event (one organism got the disease while the other did not). Yet, Godfrey-Smith's distinction is blind to them. Another more dramatic way to make the same point is to take the classical example of drift involving two individuals: one is struck by lightning while the other happily survives and reproduces.[27] In this case one individual is burnt (which will usually have bad consequences for its reproductive output) while the other is not. Both the properties of being burnt or not are intrinsic and thus should be attributed to natural selection if one follows Godfrey-Smith's distinction. This is problematic since this example is considered as the paradigmatic case of drift. This means that Godfrey-Smith's distinction does not allow accounting for all cases of drift. To be associated with natural selection a difference must be a difference in intrinsic properties (or

[27] This example is mistakenly attributed to Scriven (1959), who was rather talking of individuals sitting where a bomb or a tree falls.

causally determined by an intrinsic properties), but this is insufficient. The properties must also be invariable when subjected to an environment over time. Being fat or burnt in the examples above, although they are intrinsic properties, vary over time due to difference in the environment of the organisms bearers of these properties and should thus be associated with drift.

If the reasoning above is correct intrinsic properties should thus be decomposed into two sub-properties that will account for diachronicity: namely, intrinsic-*invariable* properties, such as having a particular gene; and intrinsic-*variable* properties, such as an amount of fat or being burnt due to a particular life history causally independent from any intrinsic-invariable properties of the entity. Both intrinsic-variable and intrinsic-invariable properties should be understood as such while specifying a range of environmental conditions and over a given period of time. Specifying a range of environmental conditions over a period of time is crucial since what is invariable now and here might not be at a later time or under different conditions.[28] It is possible to imagine that a property such as height for instance that does not vary under a range of specific conditions would do so under other conditions such as if organisms were subjected to different gravitational force. It is also known that humans are taller in the morning than in the evening due to the effect of gravitation during the day. Thus height might be considered as invariable if it is measured in organisms that always live at the same altitude and over a short period of time.

Under this modified framework, population differences in intrinsic-invariable properties within an environmental background leading to some differences in reproductive outputs are the ones to be attributed to natural selection. Conversely, differences in reproductive outputs due to

[28] More generally, the notion of 'environment' can be understood both spatially and temporally. From now on by 'environment' I will refer both to its physical and temporal dimensions.

differences between members of the population in intrinsic-variable and extrinsic properties within an environmental background should be attributed to drift.[29]

1.4.2. The possibility of Drift in an Indeterministic Set-up: Deterministic versus Probabilistic drift

The concept of drift I have developed so far presupposed a fully deterministic set-up. Yet it is possible to imagine fundamentally probabilistic intrinsic-invariable characters to have consequences on a probability distribution of reproductive output and ultimately on evolutionary change. Putative, although debated, examples of such phenomena include random foraging strategies with the possibility or quantum events percolating up into biological processes (Glymour 2001). In those cases, differences in reproductive outputs could be due to actual differences in intrinsic-invariable, yet fundamentally probabilistic, character that merely reflects a small sample of the whole probability distribution of this character. If this view is correct, one could imagine that an individual entity produces a given number of offspring simply because indeterministic processes (quantum or otherwise) led this entity to have one out of several possible behaviours. Had the indeterministic process led to another outcome, the same individual could have had a different number of offspring.

To see the difference with the deterministic set-up presented earlier, let us suppose a population composed of two types of organisms able to move freely in a patchy environment, with resources randomly distributed within it. Let us assume that both types are strictly identical in all respects except on their colour which we posit is selectively neutral. In a purely deterministic

[29] Note that drift is defined very broadly here and also entails correlated responses, that is, evolutionary changes which are not causally due to the trait under scrutiny but to another known cause.

set up, if all the organisms were to be put in the exact same initial conditions, their behaviours would be perfectly identical, so that, at any given time, each organism would have the same amount of resources. Assuming that resources are necessary to produce offspring, each organism of the population would produce the exact same number of offspring irrespective of their colour. This is because all their extrinsic (e.g., position), intrinsic-variable (e.g., brain structure depending on environmental resources available) and intrinsic-invariable (e.g., gene coding for a directional bias) properties with consequences on their reproductive output would be identical at all times. According to the definitions developed earlier, in this set-up, there would be no drift and no natural selection. Let us suppose now that the invariable property 'gene for directional bias', because of some quantum events percolating up into the DNA of individuals, can lead them to change their directional bias. In this set-up, each individual can have a direction partially independent from any other individual. These conditions can now lead to substantial evolutionary changes, especially in populations of small size.

Assuming for simplicity a population of two individuals, one of each type, in such a small population, it is in fact very unlikely that each quantum event occurring in each individual would lead them to the same decisions. As a consequence, it is also very unlikely that both individuals would have same amount of resources and consequently the same reproductive output. However, in a population of large size, say infinite for simplicity, composed of the two same types, for each individual of the population going in one direction and gathering a particular amount of resources, there will be another individual of the other type with the exact same behaviour. As a result, overall, the same number of offspring will be produced by each type, which means that no evolutionary change will be observed.

This set-up, and the evolutionary changes it leads to, looks at first glance very similar to the example presented earlier in Subsection 1.3.2.1. In fact, in both cases, some difference in an extrinsic or intrinsic-variable property is involved in a difference in reproductive output between entities, in the present case differences in foraging pathways and location for instance. Yet, the two cases should be clearly distinguished. In Subsection 1.3.2.1, one requirement for a difference in intrinsic-variable properties to be attributable to drift was that the difference *does not* ultimately causally depend on differences in intrinsic-invariable properties. In the present case, the differences in location and foraging pathways depend on the decisions made by the organisms which are entirely decided within their brains and are supposed indeterministic yet invariable and intrinsic. This set-up contrary to the set-up of Subsection 1.3.2.1 is perfectly compatible with the propensity interpretation of fitness because the entities of a given type in a population have a *genuine* probabilistic intrinsic-invariable property leading to the probability distribution.

1.4.3. Two fundamentally distinct notions of drift

We can thus see that there exists another way by which a population, due to differences in indeterministic-intrinsic-invariable properties can lead to substantial evolutionary change that deserves to be associated with drift. I name this form of drift, because it necessarily relies on fundamentally probabilistic events, *probabilistic drift*. It is separate from the form of drift resulting from difference in extrinsic and/or variable-intrinsic properties which I call *deterministic drift*, because it can occur in a purely deterministic set-up. (See Figure 1.2 for a representation of the differences between deterministic and probabilistic drift.)

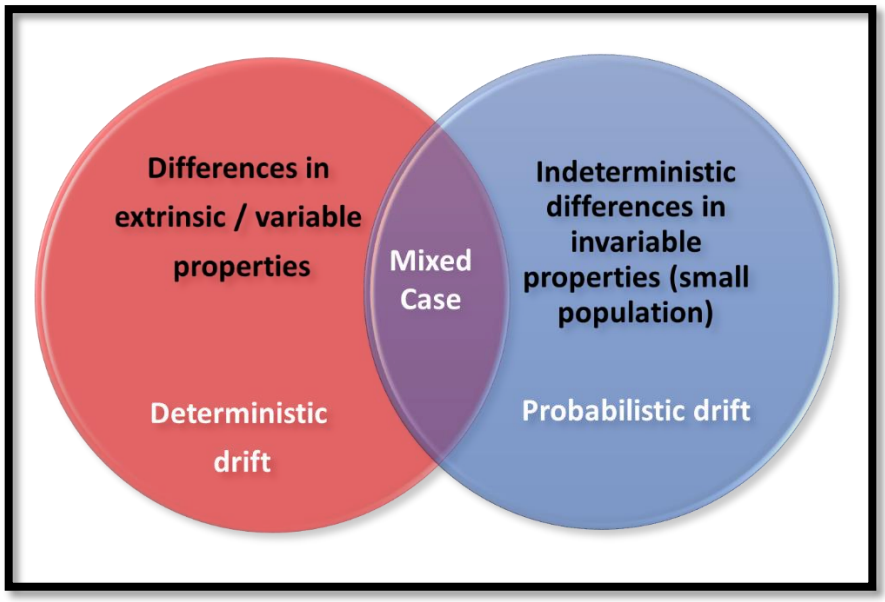


Figure 1.2. Representation of the two possible forms of drift and a mixed case.
('variable' should be read as 'intrinsic-variable' and 'invariable' should be read as 'intrinsic-invariable'.)

Although deterministic and probabilistic drift are conceptually distinct, this does not mean that they are mutually exclusive. In fact, a difference in reproductive output can in principle be due both to differences in intrinsic-variable/extrinsic properties (say differences in initial location) that do not depend on any difference in intrinsic-invariable properties and to some probabilistic differences in intrinsic-invariable or extrinsic properties (say in the present case differences in choices). In those mixed cases (see Figure 1.2 for an illustration), the total differences in reproductive outputs between the entities of the population should be attributed both to deterministic and probabilistic drift. This demonstrates the possibility of deterministic drift in an inderministic set-up[30] (probabilistic drift in a deterministic setup is however impossible).

[30] Another theoretical case of deterministic drift in an indeterministic set-up would involve indeterminism on the states of the environment.

1.4.4. Applying Strevens' account of probability to deterministic drift

Although the work made so far allowed us to distinguish two ontologically distinct notions of drift, it should be stressed that this distinction would be extremely difficult or even impossible to demonstrate empirically. This is related to the problem of the multiple possible causal interpretations associated with any statistical measure. In front of a distribution of reproductive outputs associated with a particular type, knowing whether this distribution should be attributed to ontologically indeterministic events, as in our example of foraging organisms, or to differences in intrinsic-variables or extrinsic properties (e.g., difference in brain structures due to difference in environmental states) is generally untraceable. Because of those empirical limitations, there is thus no doubt that *heuristically* considering fitness as a propensity is useful. But as I have already noted, non-trivial propensity interpretations of probabilities are generally unsatisfactory in deterministic set-ups and one should there have good theoretical reasons to interpret fitness as if it was a propensity in deterministic systems. This is precisely the project of Strevens (2011) in which he gives us nice examples of *deterministic* systems (wheels of fortune and tosses of coins) in which probabilities of events can have a "propensity-like look and feel". Strevens calls these probabilities microconstant probabilities.

For Strevens' account to be applicable, a system should exhibit 'microconstancy'. Strevens' account is very subtle and explaining the subtleties of his account would greatly exceed the scope of this thesis. Yet, it is fair to say that microconstancy is the key to his account. For more on Strevens's account of probabilities see Strevens (2003, 2011). Roughly speaking, microconstancy is the property of a system to produce a given outcome with the same proportions of initial conditions within any small but not too small neighbourhood of initial conditions. This results in a ball stopping with the same frequency on black or red slots however

the ball is launched on a classical roulette wheel and how many turns the wheel is making. If the system exhibits microconstancy, as the roulette wheel does, then it behaves as if it was exhibiting propensities provided that one does not know with precision the initial conditions.

To apply Strevens' account to ecological systems and fitness to be considered as a propensity-like property, that is, that an ecological system exhibits microconstancy over the distribution of initial conditions, one prerequisite is that each entity of the population (or at least each type as a whole) has in principle access to all the states of the environment. This is because if this prerequisite is not met, some outcomes could never be possible (in principle) with at least some entities (or types) and thus not any small neighbourhood of initial conditions would produce a given outcome with the same proportions of initial conditions. It would be equivalent in the case of the roulette wheel to some slots never being reachable for the ball.

A biological population is usually defined as a set of entities within a common selective environment, that is, undergoing the same set of ecological selection pressures (Brandon 1990, 46). This means that the assumption demanded by Strevens' account is a reasonable one. However, cases in which this assumption will be violated should generally be expected. Colyvan (2005) provides several cases in which we know that the condition of microconstancy is violated for biological systems. For instance, if we suppose a population of hares with different types of predators in different health conditions. Unless the average health of the predators are at equilibrium then we cannot consider the system to exhibit microconstancy over time. That would be equivalent to a roulette wheel changing the relative surface of black and red slots over time.

Another problem will arise if the entities of the population are in a patchy environment and different entities for reasons independent from their biology have access to different patches

of the environment. One response to this specific problem could be that if, independently from their intrinsic-invariable properties, the different entities of a set of entities have access to different states of the environment and this leads to differences in reproductive outputs, it might be justified to consider this set of entities not as one but two populations. But this solution will not always be possible. Furthermore, without knowing *a priori* whether the entities (or types) have in principle access only to certain states of the environment, it will be hard to know whether microconstancy is a reasonable assumption for a given set of entities. As a result one will always be at risk of having missed a hidden variable or made a measurement error that would render inaccurate the approximation fitness as a propensity-like property.

Leaving asides the different problems Streven's account leads to, which would require much more work to be done, if one is willing to apply Strevens' account to fitness, deterministic drift should further be separated into *microconstant drift* and *non-microconstant drift*. Microconstant drift would represent a departure from an expectation-like fitness function due to a small sample size of the population and non-microconstant drift would represent the level of error coming out of the assumption of microconstancy. Whether one should distinguish microconstant from non-microconstant drift is debatable at that point.

1.5. Natural selection from a functional causal modelling perspective

With a principled way to distinguish natural selection from drift, I now move to the literature on causal modelling. This literature is very technical and I do not intend here to use all the theoretical apparatus developed by Pearl (2000) and followers to quantitatively study causality. Rather I will only use the parts of his account necessary for my purpose, which mostly is the fact

that causal modelling makes use of causal graphs and will thus render more intuitive my account of natural selection. Causal modelling is in fact a technique that allows to translate statistical data into the language of causality by using the combination of a set of equations (in the simplest form linear equations) often called structural equations and a graphical representation. The combination of the set of equations and the graphical representation is fundamental to the technique, since equations in and of themselves only allow represent symmetrical relations between variables or parameters.[31] Yet causal relations are asymmetric relationships between variable – **X** causes **Y** is not equivalent to **Y** causes **X**. By combining equations to a graph, asymmetric causal relationships causality can be represented quantitatively.

Another particularly interesting property of causal modelling and causal graphs, for our purpose, is that they can be interpreted from a counterfactual perspective. In philosophy of science Woodward (2003) is one proponent of this approach. Apart from Woodward (2003) and Pearl (2000), a good introduction to causal graphs and causal modelling, less technical than Pearl (2000) and more empirically minded than Woodward (2003), can be found in Shipley (2002). In this section, I present very briefly and incompletely the causal modelling approach and propose a simple example of causal graph in the context of evolutionary change to illustrate this approach. This simple example will be made more complex in Section 1.6 to show how the use of causal graphs can help teasing apart natural selection from drift in purely individual causal terms. In

[31] I use indistinguishably 'variable' and 'parameter' to mean a property of organisms or event changing a property of organisms. I make this choice to avoid the confusion between 'intrinsic-variable properties' and a 'variable' in a model.

Section 1.7 I will defend the ILC view against its main rival, namely the PLC view since both approaches can be used in the causal modelling framework.[32]

To start off, one can remark that Judea Pearl, the leading expert of the causal modelling approach, has an interpretation of probabilities throughout his book in terms of epistemic limitations that matches with Bouchard and Rosenberg's view on probabilities in evolutionary theory. Pearl's commitment to this interpretation of probabilities is fundamental to this approach (see Pearl 2000, xiii, 26-27). This suggests, without being decisive, that the reasoning behind functional causal modelling is well suited to test the validity of Bouchard and Rosenberg claims about the statistical nature of natural selection and that if B&R and I are wrong about causal nature of selection and drift, a causal modelling approach to natural selection relying on individual properties should be unsuccessful.

To illustrate the basic reasoning behind causal modelling, let us postulate a very simple case of evolution similar to the most basic models of population genetics. Suppose a population in an environment (**E**) composed of two types of haploid organisms differing only with respect to the phenotype 'colour' genetically determined at a single locus. One type of organisms is 'white' (**W**) and the other 'black' (**B**). For simplicity suppose that these organisms reproduce in discrete generations. Imagine now the case where we observe that although the population was initially composed of 50% of each type, after one generation, it is now composed of 70% of **W** and 30% of **B**, because **Ws** have overall more offspring than **Bs**. Suppose now that we want to know what the causes of this evolutionary change are. The only information we have about the organisms forming the population, besides their reproductive outputs, is that E is not perfectly

[32] Note that the statistical approach being a non-causal one, it cannot be used to separate natural selection from drift in causal terms.

homogeneous that is, has different states with parameters at different values. Call this simple case **EC** for 'evolutionary change'.

Suppose now that the measures we have are exact or precise enough for our purpose. In front of this difference in frequency there are broadly three non-mutually exclusive possible explanations of the evolutionary change in **EC**:

- (*drift*) The two types have the same (ecological) fitness which means that **B**'s traits and **W**'s do not causally lead to any difference in their reproductive output, but **Bs** and **Ws** happened to have on average differences in extrinsic and/or intrinsic-variable properties leading to different reproductive outputs;
- (*correlated response*) There is a difference in the **B** and **W**'s traits (intrinsic-invariable properties) which leads **Ws** to produce more offspring than **Bs**, but the phenotype colour is not causally responsible for this difference, only correlated to it. Besides, there is no other difference in intrinsic-variable properties leading to different reproductive outputs;
- (*natural selection*) the explanation is the same as the previous one, but in this case, the phenotype colour is causally responsible for the difference in reproductive output between the two types.

Without more information on at least one biological mechanism[33] causally responsible for the type **W** to have more offspring that the type **B** and involving the phenotype 'colour', establishing whether the population evolved by natural selection on this trait is impossible unless

[33] I use a definition of mechanism taken from Machamer, Darden, & Craver (2000, 3): "Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions."

one makes blind assumptions on the distributions of events undergone by the two types. This would be equivalent to assume that a correlation between "colour" and "reproductive output" implies causation. Yet, this is famously problematic and the three explanations presented above are *prima facie* all plausible explanations if one has for sole information, differences in reproductive outputs.

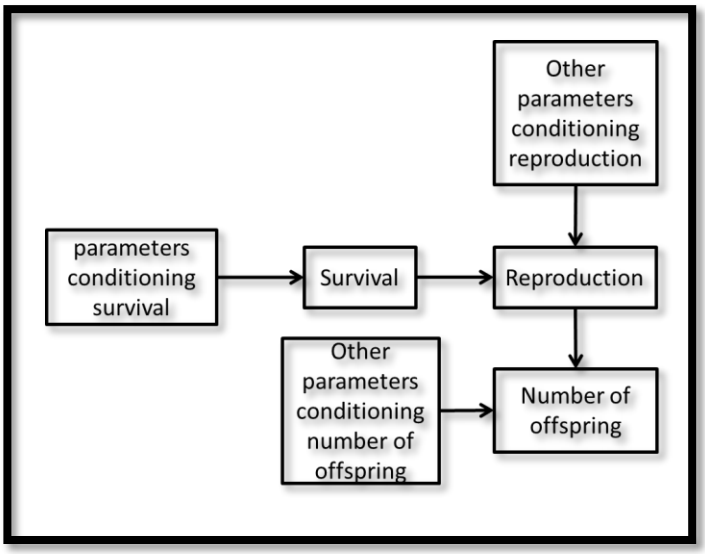


Figure 1.3. DAG representing the variables involved in the reproductive output of an organism in the case EC.

This can be visualised using a directed acyclic graph (DAG) which is the type of graphical representation widely used in the causal modelling literature (see Pearl 2000). Figure 1.3 is a DAG that represents the variables involved in the determination of the reproductive output of a given organism in the case **EC**.[34] Each box represents a variable for a given organism in the population.

[34] Note however that it only represents the reproductive output of one given type of organism in **EC**. Yet evolutionary change can easily be computed by calculating the difference between the numbers of offspring produced by the two types as displayed on Figure 1.8.

Each arrow represents a causal[35] relationship between two variables. The variable preceding directly the bottom of an arrow is called a *parent* of the variable at the top (if it precedes it indirectly it is called an *ancestor*) and the variable directly at the top of this same arrow is called a *child* of the variable at the bottom (or if indirectly, a *descendent*). A parent or ancestor is understood as a cause and a child or descendent as an effect in a relation between two variables.[36] For example, in Figure 1.3, the variable 'survival' which can either take the value 0 (dead) or 1 (alive) for each organism, is a parent (that is, causally influences) the variable 'reproduction' (also binary: "produces no offspring" or "produces some offspring"). Manipulating an ancestor will make a difference in the value of its descendent. The precise relation between ancestors and descendants can be calculated by a set of functions providing the relations between parents and offspring and matching the variables in the DAG. In the case of **EC** the equations (written in a functional form here) are:

- $'Reproduction' = f\begin{pmatrix}'survival',\\'other\ parameters\ conditioning\ reproduction'\end{pmatrix}$
- $'survival' = f('parameters\ conditioning\ survival')$
- $'number\ of\ offspring' =$

  $f\begin{pmatrix}'other\ parameters\ conditioning\ number\ of\ offspring',\\'reproduction'\end{pmatrix}$

Using the distribution for each type of organism on each observed variable, one can assess quantitatively, using this set of functions, the relative effect of manipulating one variable

[35] Bear in mind that by 'causal' no distinction between causal and constitutive relationship is made here since all the relationships described are diachronic.

[36] More abstractly, each variable child is a function (linear or otherwise) of its parents. This function is then interpreted in causal terms: hence the reason why 'causal modelling' is sometimes called 'functional causal modelling'.

on the other variables downstream the graph and thus obtain a measure of the causal relation between them.

The graph on Figure 1.3 shows that the number of offspring a given individual produces is the result of three *exogenous variables*, namely 'parameters conditioning survival', 'other parameters conditioning reproduction'[37] and 'other parameters conditioning number of offspring' which are variables with no parents represented in the graph. This means that those variables are given as inputs into the model. Between the two variables 'parameters conditioning survival' and 'other parameters conditioning reproduction' and the variable 'number of offspring', which represents the number of offspring produced by an organism, there are two *endogenous variables* namely 'survival' and 'reproduction' that is, variables which are functions of exogenous ones or of other endogenous variables.

One important remark to make at this point, is that although functional causal graphs represent the underlying causes of individual variable (events or properties such as the number of offspring produced by one individual organism in our example), causal inference using causal modelling is a statistical technique that requires data on ensembles of entities. This is potentially confusing since we want to establish whether natural selection as a cause should better be understood at the individual or population (ensemble) level. Yet, the causes of a phenomenon should not be confused with the means of epistemologically reaching them. Thus even uncontroversial cases of causal inference of individual level events, population level data are necessary. Another important remark is that a causal graph supposes that the sum of all the causal

[37] 'Other parameters conditioning reproduction' includes all the parameters necessary for an individual to reproduce except the variable 'survival'.

influence of a child comes directly from its parents. Thus there is no missing causal influence in the model.

The type of graph represented in Figure 1.3 is called a DAG because it contains no cycle and has a direction interpreted as a direction in time. Thus there are no feedback loops in a DAG. This feature is important if one assumes that causality is an asymmetric relationship between two variables in time (time's arrow) which is an assumption I will make throughout the rest of the chapter. Once this assumption is made, no genuine feedback loop is possible. This is because a descendant cannot causally influence its ancestors, for its ancestors only exists earlier in time (Shipley 2002).[38] However, a causal descendant can influence another *particular* descendant of its ancestor that bears some structural or constitutive similarities with its ancestors, leading to the impression that a genuine feedback loop exists, when it is in fact a pseudo one. A graphical representation of this kind of causal process such as depicted in Figure 1.4 with 6 variables 'A1', 'B1', 'A2', 'B2', 'A3' and 'B3' makes this idea salient. Once time is taken into account, A at time t_1, is different from A at time t_2 (A at time t_1 is a parent of A at time t_2) and thus no feedback loop from B1 to A1 is possible. Another important property of causal graphs is that a DAG represents a complete causal structure, which is to say that all the sources of dependence are explained by the causal links.

[38] Feedback loops are also possible when one talks about type causes that are not bound to particular space-time locations. Yet, causal modelling is a method developed to deal with token causes and thus bound to particular space-time locations.

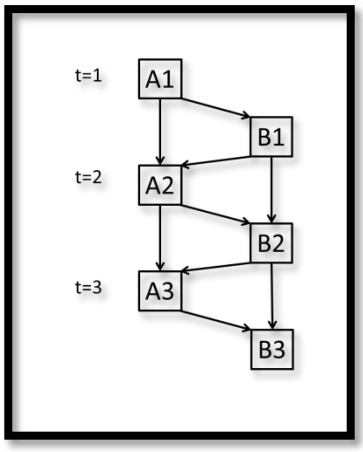


Figure 1.4. DAG of a pseudo feedback loop between A and B when time is taken into account.

1.6. Teasing apart natural selection from drift

After this very short and incomplete introduction to causal modelling, let us go back to our problem, namely to understand why with no biological mechanism involving the phenotype 'colour' leading to differences in reproductive output in **EC** between **W** and **B**, it is not possible to separate natural selection from drift. We can see from the DAG in Figure 1.3, that with no more information than reproductive outputs, there are five possible variables that could have a different distribution between the two different types of organisms and explain a difference in reproductive output between them (all the variables except the variable 'number of offspring').

Imagine now that we get some evidence, or that you have no reason to doubt, that the variable 'other parameters conditioning reproduction' has the same probability distribution in the population between the two types, as well as the variables 'reproduction' once the variable 'survival' is held constant and the variable 'other parameters conditioning number of offspring'. With this supplementary information, the only difference maker remaining in the graph that can explain why the type **B** has less offspring than **W** is a difference in the distribution with respect

to the variable 'parameters conditioning survival'. Yet, establishing that a difference in 'parameters conditioning survival' between the two types is causally involved in a difference in reproductive output between them, is not sufficient to determine whether the evolutionary change observed in the population is the result of natural selection. In fact, the difference in these parameters between the two types could, following the distinction made in the previous section, either be due to some invariable properties of organisms (a trait) and attributable to natural selection or due to variable properties of organisms and attributable to drift. Without more information on whether and to what extent these two parameters contribute to the reproductive output of the two types of organisms, there is no possible way of knowing whether the difference in number of offspring produced in the population is due to natural selection or drift.

In order to distinguish between natural selection and drift, we thus need to have more information on whether and to what extent there is a difference in the distribution with respect to the intrinsic-invariable or extrinsic /intrinsic-variable parameters between the two types. To do so, the first step is to decompose the variable 'parameters conditioning survival' into 'intrinsic-invariable parameters conditioning survival' and 'extrinsic and intrinsic-variable parameters conditioning survival' such as represented in Figure 1.5. Following the distinctions made earlier, differences in intrinsic-invariable parameters conditioning survival between organisms of different types should be associated with natural selection while differences in extrinsic and intrinsic-variable parameters conditioning survival between organisms of different should be associated with drift.

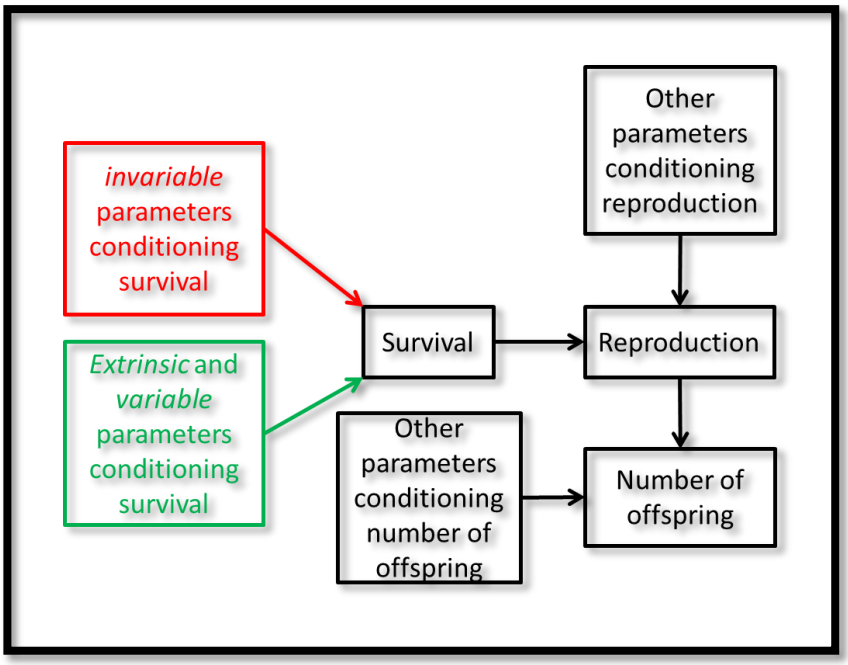


Figure 1.5. DAG representing the variables involved in the reproductive output of an organism in the case EC and allowing the distinction between 'difference in reproductive output due to drift' and 'difference in reproductive output due to natural selection'.
('variable' should be read as 'intrinsic-variable' and 'invariable' should be read as 'intrinsic-invariable'); parameters in green are extrinsic and variables(-intrinsic) parameters; Parameters in red are invariable(-intrinsic) parameters.)

One might think at that point that one way to make the distinction between natural selection and drift is merely to ensure that by controlling that the probability distribution of all the variable parameters is the same between the two types, the only possible causal explanation of the difference in reproductive output between the two types is that there exists a difference in distribution between the two types on some invariable parameter (albeit) unknown to us. Proponents of this solution could argue from that point that having a biological mechanism explaining the downstream difference in reproductive output between the two types is thus non-obligatory. Although *ideally* perfectly valid this solution encounters two problems. First, arguably a substantial part of the research program in evolutionary biology is to identify adaptations, which is impossible with the solution proposed above. In fact, when we say that the type **W** is fitter

than **B** we mean "**W** is fitter than **B** *in virtue* of a difference in colour between the two types." Without a mechanism explaining why a difference in colour leads to a difference in reproductive output, the latter claim can hardly be substantiated and is prone to be confounded with another factor. The problem of hidden confounds is the reason why Fisher (1970) regarded controlled experiments as an inferior method to infer causation than randomised experiments which partially deals with this problem (see Shipley 2002, chap. 1 for more details). The solution allows only making the claim that a difference in *some* invariable property or combination of invariable properties is involved in the difference in reproductive output and thus that there is *some* natural selection in our population with no special trait in mind. The distinction between our hypotheses *correlated response* and *natural selection* presented earlier will be unsuccessful with this solution.

Second, it is impossible to know what variable parameters to attribute to drift without first having in mind potential mechanisms that would lead some individual parameters to remain constant while others vary, since those extrinsic and/or intrinsic-variable parameters might be involved in a mechanism explaining the difference in reproductive output between the two types (remember the parameter 'location' in Section 1.3 which could be under the causal influence of either intrinsic-invariable property or another extrinsic or intrinsic-variable properties). This is one major reason why bringing causal modelling into the philosophical literature on the causes of evolutionary change represents a genuine step forward in this literature.

This naturally brings me to a second solution, the one endorsed by Bouchard and Rosenberg and myself, and consists in independently identifying a causal mechanism (if such mechanism exists) involving the phenotype colour and explaining why in the population, the type **W** has a higher reproductive output than **B** as well as eliminating as much as possible any known confound. Imagine for example that we independently gain the knowledge that predators of the

organisms of our population in **EC** use sight to catch their prey and that because organisms of the two types spend a lot of time on the trunk of birch (white) that makes the **Ws** harder to distinguish and increases their probability of survival. This is a causal mechanism similar to the classical one described in the peppered moth by Kettlewell (1955). This solution, although not perfect because not immune to potential confounds, is nevertheless superior to the previous one for it renders the claims that the type **W** is fitter than **B** *because* it is white more likely to be true. Once such mechanism has been found, it is less likely that a confound variable, through an alternative mechanism would explain the exact same pattern. And if it does one will most likely be able to make further predictions in order to separate the two hypotheses.

In order to make causal claims about the role of selection for the trait colour, it is thus safer to modify the causal graph represented in Figure 1.5 into the one represented in Figure 1.6. This allows to reflect that one cause of survival for organisms is that they are not always visible to predators due in part to their colour which causally depends on the phenotype 'colour',[39] to other invariable properties (such as for example whether there are specific patterns on the body of a given organism), due to some variable parameters (such as the amount of energy each organism has to escape predation) and finally some variable properties (such as location). With the causal graph represented in Figure 1.5 we have all the information necessary to infer from statistical data, by applying causal modelling techniques, whether and to what extant the trait 'colour' is adaptive (the result of natural selection *for* colour), a correlated response of one or some other adaptive trait(s) (the result of natural selection *of* colour)[40] or again the result of difference of distributions in variable properties between the two types of organisms (drift).

[39] I assume that the trait 'colour' is intrinsic-invariable in E.
[40] I borrow the distinction between 'selection for' and 'selection of' from Sober (1984).

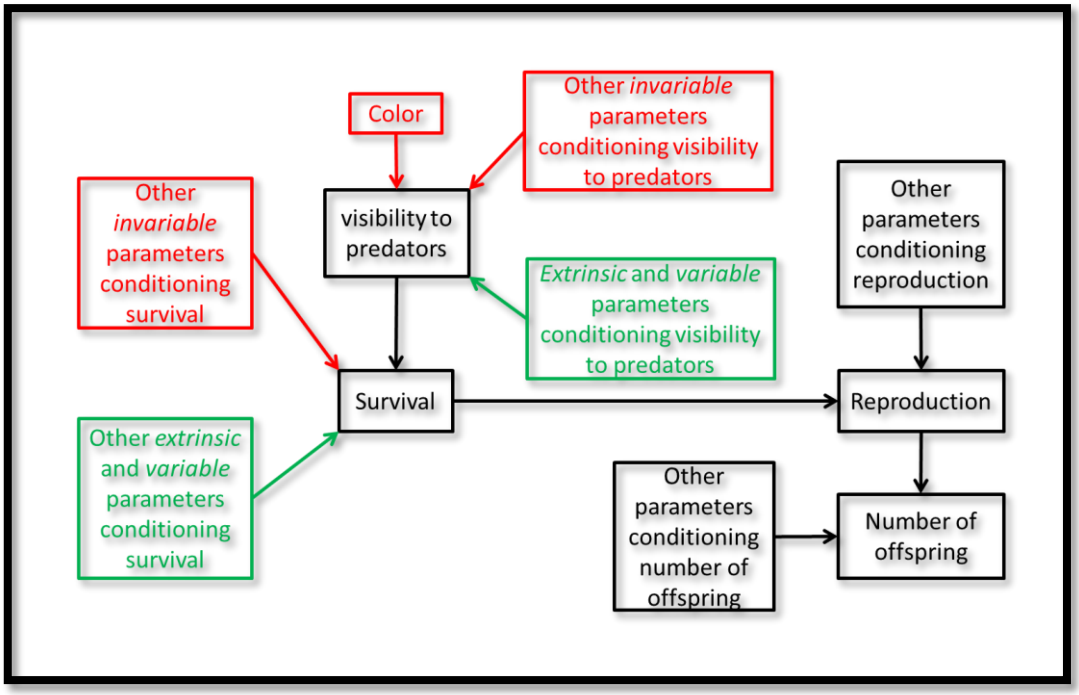


Figure 1.6. DAG representing the variables involved in the case EC allowing to establish whether the trait 'colour' is selected for.
('variable' should be read as 'intrinsic-variable' and 'invariable' should be read as 'intrinsic-invariable'); parameters in green are extrinsic and variables(-intrinsic) variables; Parameters in red are invariable(-intrinsic) variables.)

Although applying causal modelling techniques to the graph in Figure 1.6 allows us distinguishing neatly between natural selection, drift and the correlated response for the phenotype colour, one should bear in mind that it would not allow an observer to make this distinction in practice without having information on the distributions on all the extrinsic, intrinsic-variable and intrinsic-invariable properties of the organisms in the population. Evolutionary biologists, in many cases, only have a record of the reproductive outputs of the organisms they study and some evidence of a mechanism involving an intrinsic-invariable parameter. The only practical solution given those epistemic constraints is to suppose that the mechanism hypothesised is responsible for the evolutionary change observed, while considering

that the distribution for the other variables is the same between the two types. Yet, this solution is plagued with the problem of confounding variables potentially explaining the same data. Thus, because of the inherent epistemological problems it comes with, this solution should be used with great care and updated as one obtains more information on the different variables affecting reproductive outputs of populations under consideration. As poor as this solution is, this is the best available and these practical difficulties should not be confused with the theoretical apparatus underpinning the notion of natural selection and drift I have proposed.

So far this chapter has shown that there is an alternative to the concept of fitness as a propensity. I have argued for a concept of fitness as relational properties between two or more individual entities forming a population. I have shown, using causal graphs that distinguishing between extrinsic and intrinsic-variable properties on the one hand and intrinsic-invariable properties of entities on the other hand could be the basis for distinguishing natural selection from drift. Because those properties are properties of individual entities and not properties of populations, this demonstrates the plausibility of the ILC view. But as noted in the introduction there is another causal approach to natural selection and some might be tempted at that point to use causal modelling from a population level perspective, that is, to use population level variables instead individual level variables in a DAG and claim that it is better than the ILC view. In the next section, I show that although this is possible, the ILC view, if not misrepresented, is nevertheless superior for it leads to more invariant relations between variables.

1.7. Natural selection and drift: population level causes of evolutionary change?

As we have seen in the introduction both the statisticalist and the PLC views give a different interpretation of natural selection from the ILC view. Bouchard and Rosenberg (2004) discuss at length the problems with the statisticalist view and in a later paper (Rosenberg & Bouchard 2005), the problems with Matthen and Ariew's (2002)[41] account more specifically. I will not spend much more time discussing it since I largely agree with their analysis presented in Section 1.2. Bouchard and Rosenberg analysis of the statisticalist view revealed, in a nutshell, that if there is no doubt that natural selection can be characterised statistically, that should not lead us to the conclusion that ecological fitness, involving comparison between entities of a population causally upstream to reproductive output (ecological fitness), is dispensable from evolutionary theory. The causal modelling approach I proposed in the previous section is perfectly compatible with B&R's conclusion.

1.7.1. Reisman and Forber's arguments

The proponents of the PLC view agree that natural selection and drift are causes of evolutionary change but they disagree that causality takes place at the individual entity level. Reisman and Forber (2005), to defend this view, offer three arguments for considering natural selection and drift as population level rather than individual causes. The first is that they succeed in showing that drift can be 'manipulated' at the population level without any reference to individual level properties. Thus under the manipulationist account of causation drift (they claim) can be regarded as a population level cause of evolutionary change. Second, they argue that if population level properties supervene on individual properties they do not necessarily reduce to them. Third, they argue that speaking of natural selection and drift in terms of population level

[41] Incidentally, many of the argument given by Bouchard and Rosenberg can also be applied to the proponents of natural selection and drift as population level causes.

causes is more informative than speaking about them in terms of individual level causes. Below, I argue that Reisman and Forber have succeeded in establishing neither that natural selection and drift are population level causes irreducible to individual level causes nor that the PLC view is superior to the ILC view. At best, they represent two descriptions of one and the same phenomenon. Finally, I present a hypothetical case in which individual level proprieties would be indispensable to separate natural selection from drift.

First of all, in many respects (but not all), I see no fundamental opposition between the PLC view and the ILC view: Bouchard and Rosenberg claim that natural selection is a causal process resulting from differences in individual fitness while Reisman and Forber claim that natural selection is population-level cause of evolutionary change. It is easy to interpret Reisman and Forber's view in Bouchard and Rosenberg's language by equating a difference in individual (ecological) fitness to a population level cause of evolutionary change.[42] The population level causes under the PLC view are the differences in kind (extrinsic and intrinsic-variable on the one hand and intrinsic-invariable on the other hand) of individual properties under Bouchard and Rosenberg's view and both agree that natural selection is a causal process. Figure 1.7 is a representation of my interpretation of Bouchard and Rosenberg's and Figure 1.8 a simplified representation of the PLC view (both applied to the case EC). As can be seen by comparing the two figures, one can easily move back and forth between the two views by remarking first that

[42] Some might argue at that point that 'differences in fitness', because a difference implies at least two individuals, are necessarily, by definition, population-level causes. But that would be missing that Reisman and Forber, who are in the PLC camp, agree that "this tactic of situating all of the causal action in evolution at the individual-level can be repeated for any purported manipulation of population-level causes."(Reisman & Forber 2005, 1120). Furthermore 'manipulating', under the manipulationist account of causation, always involves at least two situations to be compared and thus the concept of 'difference maker' is a relative one. Thus if one is prepared to claim that differences in fitness are population level causes, then they should be also claim, to be consistent, that any difference maker is a population level cause, because it makes a difference relatively to a set of background conditions in the same way an allele, for instance, makes a difference relatively to a population.

under the ILC view, the difference in distribution in intrinsic-invariable parameters and the difference in distribution in extrinsic and intrinsic-variable parameters between the two types correspond respectively to natural selection and drift under the PLC account. Second, one can notice that the reproductive outputs of the two types under the ILC view, once considered together in a population are represented in terms of frequency under the PLC view.

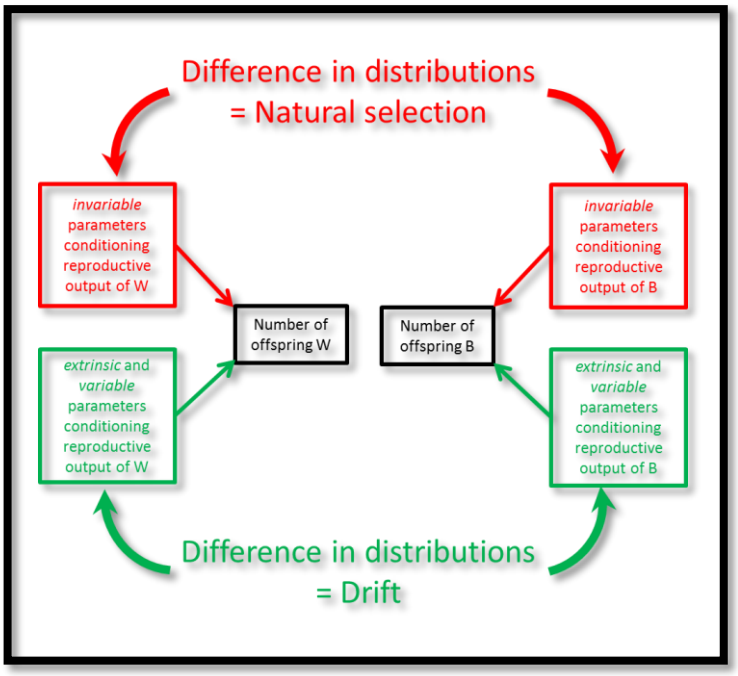


Figure 1.7. Illustration of Bouchard and Rosenberg's view of evolutionary change applied to EC. ('variable' should be read as 'intrinsic-variable' and 'invariable' should be read as 'intrinsic-invariable'); parameters in green are extrinsic and variables(-intrinsic) parameters; Parameters in red are invariable(-intrinsic) parameters.)

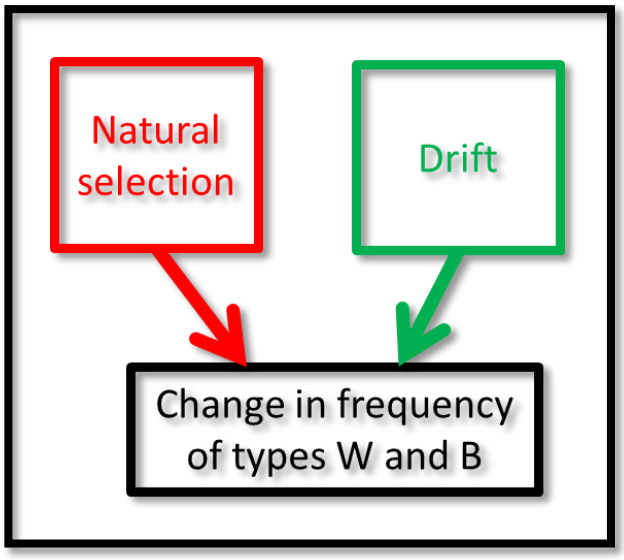


Figure 1.8. Simplified illustration of the PLC view applied to EC.
(initial allele frequencies and population size are not represented.)

To me, the fact that one can move back and forth between the two views is already evidence that the two views are to some extent equivalent and that R&F, at any rate, overstate their positions when they claim that the PLC view is irreducible to and more informative than the ILC view. What I will show in this section is that the opposite is true, that is, that the ILC view is overall more informative than the PLC view and to be fair to R&F, I will use their own example. They present a controlled experiment realised by Dobzhansky and Pavlovsky (1957) which shows that a form of drift (the founder effect) can be manipulated at the population level and causes a deviation from expected evolutionary change at a particular locus. By *drift* I will thus mean here the *founder effect*, which is the target of Reisman and Forber's argument, but it should be clear that the same argument could be run with other phenomena classically associated with drift. Drift in this experiment is manipulated through population size. When the frequency of a particular allele in different replicates of populations of fruit flies reproducing freely and composed initially of 20 (F1) heterozygote individuals sampled from a source population is

tracked over time, the variability over the different replicates in their deviation from the expected frequency of this allele at equilibrium is higher than when the replicates populations are founded with 4000 individuals of the same source population.

Supposing a homogeneous environment,[43] the explanation given by Dobzhansky and Pavlovsky of their result is that when 20 fruit flies are sampled from the source population the sample contains less genetic variability at other loci than when 4000 fruit flies are sampled which is itself dependent on the sample size. Thus, Reisman and Forber argue, drift is a function of the size of the sample. Although this (standard) population level explanation is perfectly legitimate, one should note that an individual level explanation is also perfectly compatible with the finding of Dobzhansky and Pavlovsky. The individual[44] level explanation is as follows. First one can remark that when alleles of both types have the exact same genetic background and are compared (by manipulating whole fruit flies) for the effects they produce on the population, only the effect(s) of the alleles will be observed in the population since they represent the only difference makers. The genetic background being the same, when the manipulation is made, it does not make a difference. This remark is important because when a sample is small (20 fruit flies), the probability that for each allele of one type another allele of the other type with the *same* genetic background is present in the population is smaller than when the sample is large (4000 fruit flies). Thus it is more likely that the frequency of both types diverge from what is expected in the former case due to the influence of the genetic background of the allele studied being more variable and thus being effectively a difference maker.

[43] Note that the environment is homogeneous from the point of view of the *fruit flies* and not from the point of view of the *alleles* of the population

[44] In population genetics, the relevant population is a population of *alleles*. Thus an 'individual' refers here to a token allele.

To make this clear, suppose for instance, for simplification purpose, that each fruit fly is haploid and has only two loci. Call one locus "the focal locus" while the other is the "background locus". Suppose there are two alleles at the focal locus and 10 at the background locus. Finally suppose that each alleles has a different effect on the reproductive output of the fruit fly. In a sample of 20 flies (10 with each allele of the focal locus), it is less likely that for each fly having a background locus with a particular allele there is another fly in the sample with the same allele on the background locus than when the sample is of 4000 flies. Since each allele has a different effect on reproductive output, it is more likely that samples of 20 flies leads to a frequency of the focal allele different from what is expected simply because of the difference in effect on reproductive output coming from the background locus.

Thus by simply considering that the genetic background of an allele is an extrinsic properties of the type, since it is not systematically associated to it, it thus seems that *pace* Reisman and Forber the PLC view is perfectly reducible (in principle) to the ILC one. Although talking about population level cause might be a handy way to express what is happening in the population and this is in fact what Dobzhansky and Pavlovsky do, another more detailed explanation is that drift results from individual differences in some extrinsic properties (here the genetic background of every token allele studied).

At that point Reisman and Forber might argue that because it is handier, yet equivalent, to explain from a PLC perspective, their explanation is superior to mine because it gets to the same outcome at a lesser cost: tediously establishing all individual causes versus readily manipulating the founding population size. But that would be a mistake. In fact, although Reisman and Forber's conceptual apparatus is much less constraining than mine (a brief look at the difference between Figure 1.6 and Figure 1.7 is sufficient to see it), their account is, following

Woodward's criterion of invariance (Woodward 2000, 2010), less explanatory than mine. As a result it misidentifies drift in terms of population size while ultimately population size is confounded with some individual level properties that can fully and more invariantly explain the founder effect. In other words, given a population, individual level properties screen off population level properties in explaining drift and natural selection and thus should be accounted thoroughly by individual level properties.[45]

To see this point, imagine that Dobzhansky and Pavlovsky had used fruit flies (F0) that are initially all clones except on the focal allele, instead of fruit flies with different genetic backgrounds to produce the generation F1. Before the reproduction of the F1 generation, all the token alleles of one type of allele would systematically be associated to one genetic background while the token alleles of the other type to another different genetic background (the variation between the genetic background of two types would only be due to the presence of the other allele at the same locus in the homologous chromosome since all fruit flies would be heterozygotic at that allele). Because of this particular arrangement and in spite of a possible difference in ecological fitness between the two types of alleles, this difference would be compensated by the difference in genetic background between the two alleles. At the organism level, all F1 individuals would have roughly the same viability and fertility, and thus no evolutionary change would be observed at the F2 generation. Yet, after one generation, assuming panmixia, things would change since each allele would not be associated anymore to one particular genetic background and would be found in the two possible genetic backgrounds with equal probability. The effect of each allele on their own reproductive output, in each genetic

[45] For a discussion on the notion of screening off see Brandon (1990) or Okasha (2006).

background, could thus be assessed. From there, supposing a homogeneous ecological environment, because the variability between the sample of 20 and 4000 flies would be the same (each fruit fly in the sample – whether small or large – being identical), following Dobzhansky and Pavlovsky's explanation, the only effect observed would be the effect of natural selection in the two cases, in spite of a difference in initial sample size. There would certainly be more drift in the small samples that the large one initially due to the random mating of individuals, but that sort of drift should be distinguished from the founder effect Reisman and Forber and I are focusing on in their argument. At any rate this form of drift could also be explained purely from an individualistic perspective.

This example, although purely theoretical, flies in the face of the argument that were individual level properties (e.g., changing the genetic background of an allele) manipulated instead of population level ones (e.g., changing the population size), no information would be gained by this manipulation, as claimed by Reisman and Forber (2005, 1121). Not only would the population level manipulation, in this case, yield no difference, but the relation between the founding population size and variability in the frequency of the allele would be lost. On the contrary, varying individual level parameters, such as changing the genetic background of alleles, would yield the deviation from expectation observed by Dobzhansky and Pavlovsky.

According to Woodward (2000, 2010), to be explanatory, a generalisation describing a relationship between two or more variables must be invariant when 'invariant' means that "it would continue to hold – would remain stable or unchanged – as various other conditions change". As Woodward stresses invariance is not an absolute concept. Yet, in evaluating which one of two explanations is the best, invariance will be an important criterion: the more invariant, the better the explanation. With respect to the founder effect, an explanation based on differences

in genetic background is more invariant than an explanation based on initial population size and represents therefore a better explanation.

1.7.2. Millstein's argument

Millstein (2006), like Reisman and Forber, explicitly considers natural selection and drift as population level causes that do not reduce to individual level causes. To fuel this view, she proposes a case of 'non-transitive' population in which there are three types: A, B and C. In the absence of C, A has a higher reproductive output than B. In the absence of A, B has a higher reproductive than C. But C has a lower reproductive output than A, in the absence of B. If one follows the pairwise comparison method proposed by Bouchard and Rosenberg to determine which individuals are the fittest in the population, Millstein claims, it will be necessary to make a *full* ranking of all the individuals in the population based on each pairwise comparison to be able make predictions in non-transitive populations. This would involve, according to Millstein, a lot of comparisons.[46] Besides the practical difficulties of this method, Millstein claims that it collapses into the view that natural selection and drift are population level causes since the whole population must be ranked.

I disagree with Millstein that the account proposed by B&R or more generally the ILC view is problematic for the reason she advances. I think it is based on the wrong assumption that a full ranking of individuals would be necessary to know which organisms are fitter. The method of comparison, as it should be clear by now with the use of causal graphs, does not require a full ranking of organisms' reproductive output. In fact as the causal graphs in Figure 1.5 and 1.6 display, to establish which one of two or more organisms is fitter, it is necessary that the

[46] N(N-1)/2 comparisons are necessary according to her N if is the number of individuals in the population.

organisms are compared with the same values of extrinsic and variable parameters (if the explanation is about a specific trait, then the same distribution of genetic backgrounds should also be assumed), bearing in mind that reproductive outputs are only proxies to fitness. The risk otherwise is to take into account the influence of the of different environmental conditions independent from the biology of the organisms in the fitness comparisons. It is important to stress that it should be avoided by all means. Thus making a full ranking of individuals in terms of reproductive outputs is irrelevant simply because there will be differences in intrinsic-variable and extrinsic properties between them that will explain a substantial amount of the differences in reproductive outputs. The only relevant ranking, in principle, is between individuals with the same intrinsic-variable and extrinsic properties.

With non-transitive populations (leading to a sort of frequency dependent selection) because each type influences the reproductive output of at least one other type, the environment of an organism taken randomly in the population is composed of other organisms and changes when the population evolves, which leads each types to solve the design problems set by the environment[47] differently over time. Yet that does not mean that, *at a given point in time and in a particular environmental context*, one cannot assess in principle, which type, if any, is effectively solving the current design problem in the best way. The facts that the environment of a particular organism is also a part of the population and that at least some information of some other organisms (or of the whole population) is necessary to predict the evolutionary change is not an argument in favour of natural selection as being a population level cause. The process of natural selection still occurs as the result of individual differences that have two outcomes: a change in

[47] I use here B&R's currency of fitness for expository purposes and to follow Millstein's reasoning, but I am not committed to it. All could be framed in terms of upstream causes leading organisms to produce more or less offspring.

frequency between the different types in the population and a change in the environment of each individual of the population. Yet, the outcomes should not be confused with the process by which these outcomes are reached.

1.8. Conclusion of the chapter

In this chapter I have shown using a rudimentary causal modelling approach and a principled distinction in terms of intrinsic-invariable, intrinsic-variable and extrinsic properties, that natural selection and drift could be fully understood as the result of differences in these properties between the members of a population. I have argued more precisely that natural selection results from differences in intrinsic-invariable properties between the members of a population and I proposed that we should distinguish two forms of drift. One that results from difference in extrinsic and intrinsic-variable properties, I call deterministic drift and another one, I called probabilistic drift, that results from differences in indeterministic intrinsic-invariable properties. I have argued that it is reasonable, following a causal modelling approach, to label natural selection and drift, once understood as such, as being 'causes of evolutionary change'. Finally, I have shown the superiority of this view in terms of individual level properties over other views and more precisely over views that consider natural selection and drift as population level causes. With this in place, I propose in the next chapter a set of conceptual tools that will allow us to distinguish between the process of natural selection and its different products, namely ENS and complex adaptations. I show that this distinction, although quite simple, has not been clearly made in the literature and that it poses some conceptual problems.

2. Natural Selection and its Products: Evolution by Natural Selection and Complex Adaptations

2.1. Introduction

In the previous chapter, I have proposed a systematic way to differentiate natural selection from drift. But both processes must, to be in line with the distinction made in the Introduction of the thesis, be replaced in the context of, and distinguished from, evolution and more particularly ENS. This is the main project of this chapter.

Since Darwin (1859) numerous authors have proposed summaries for ENS that all have the following tripartite formulation: for ENS to occur, a population should exhibit variation in character, , which leads to differences in fitness and that is heritable to some extent (e.g., Brandon 1990; ; Dunbar 1982; Endler 1986; Lewontin 1970, 1985; Ridley 1996; see Godfrey-Smith 2007 and 2009 for the subtle differences between each of these summaries). Godfrey-Smith (2009b, 4) calls these summaries 'recipes for change by natural selection'. One remarkable point about those summaries, also referred to as the 'classical' approach to ENS, is that they do not require *perfect* inheritance of character, that is, the exact transmission of character from parent to offspring. For example, if within a population, the taller entities consistently produce offspring that on average are taller than the rest of the population, yet do not have the exact same height as their parent, ENS should nevertheless result. Thus ENS is possible even if "everywhere we look there are degrees of similarity but no variation "faithfully transmitted"", (Godfrey-Smith 2009, 33). This type of reasoning has led Godfrey-Smith to advocate a form of 'evolutionary nominalism' (2009,

35) which holds that "the grouping of entities into types is in no way essential to Darwinian explanation". By 'Darwinian explanation', I take it, Godfrey-Smith means 'ENS'. Evolutionary nominalism is in sharp contrast with an alternative approach to ENS, namely the replicator framework (Dawkins 1976; Dennett 1995; Haig 2012; Hull 1980), which posits perfect inheritance (mutations excluded) at some level (often the genetic level) for ENS to occur. Replicators, according to this view, are transmitted without changes across generations and underlie any evolutionary explanation involving natural selection.

Although both frameworks have been used in evolutionary theory and are sometimes seen as equivalent, their relation remains quite unclear (Godfrey-Smith 2009b, 32). In this chapter, following Godfrey-Smith, I call into question this equivalence. Yet, the conclusion I draw differs to some extent from that of Godfrey-Smith. In fact, I conclude that any evolutionary change resulting from applying the classical approach in which no perfect transmission is supposed is not the result of natural selection alone (and in some cases not at all: see Section 1.3 and Godfrey-Smith 2007, cases 8 and 9), but involves inevitably some other evolutionary causes that are not clearly specified in the classical approach. I relate those evolutionary processes to the classical evolutionary cause of drift, following the definition I gave of it in the previous chapter, and mutation, of which I will provide a definition here. I show that the causal evolutionary process leading to ENS in and of itself, if understood over more than one generation requires perfect inheritance of traits. Later on, I distinguish the process of natural selection from the processes of reproduction and transmission. I show that they have been conflated in the philosophical and biological literature. Natural selection, reproduction and transmission are often part of evolutionary explanations involving complex adaptations which are considered by many to be the most interesting sort of evolutionary explanations. That said, these three evolutionary

processes should nevertheless be conceptually distinguished for natural selection does not necessarily leads to complex adaptation, or even ENS.

To make those different points, I start in Section 2.2, by arguing that Godfrey-Smith's version of the classical approach, culminating with his Darwinian space presented in the previous chapter, is very successful in delimitating the *class of populations* in which *part* of the evolutionary change is due to natural selection. Within this class of population, some will lead to the evolution of complex structures. As successful as this version of the classical approach is, it does not however explain how the *process* of natural selection in and of itself leads to ENS. To get such an explanation, we need a method that will permit us to delimit specifically the process responsible for natural selection. One way to specifically target this process, I will show, is to specify a population in which every evolutionary force except natural selection has been stripped down from the total evolutionary change observed. (I call these populations 'pure Darwinian populations'.) But specifying the properties of a pure Darwinian population requires first having some criteria other than statistical ones for distinguishing the different evolutionary processes that can be at play in the population. The tools developed in Chapter 1 allow us to do precisely this with respect to natural selection and drift.

Yet, nothing has been said so far about the evolutionary cause of mutation.[48] This will be undertaken in Section 2.3, in which I show that ENS without perfect inheritance necessarily involves a process that should be linked to the process of mutation.[49] This is because, as I will demonstrate, imperfect inheritance and mutation – of which a definition in the spirit of those provided in the previous chapter for natural selection and drift will be provided – should be

[48] Migration, for the reasons stated in the previous chapter will not be treated here.
[49] Mutation be will understood here in a broad sense.

regarded as formally equivalent. This will lead me to conclude that a process of pure ENS over more than one generation requires perfect inheritance at some level. To make this latter point, I start by detailing two idealised illustrative examples and then demonstrate it more systematically. I then show that the replicator and the classical approaches need not to be opposed: they just propose two complementary forms of evolutionary explanations. The former strictly focuses on natural selection from a process perspective while the latter contextualises natural selection within real populations in which natural selection is only one among many evolutionary forces.

In Section 2.4, I show that there are different definitions of natural selection in the literature, some of which include reproduction and inheritance and some of which do not. In Section 2.5 and 2.6 I argue that reproduction (Section 2.5) and inheritance (Section 2.6) are conceptually different from the process of natural selection and thus that, strictly speaking, natural selection requires neither reproduction nor inheritance. In fact, as I will demonstrate, they are not associated with natural selection, but with the notions of ENS and complex adaptation, which are both related to natural selection, but ultimately different from it. In particular in Section 2.5, I expose some of the difficulties with the concept of reproduction in relation to the concept of individuality often seen as fundamental for the process of natural selection. I propose some examples illustrating why a population composed of entities that do not reproduce but differentially persist over time should be regarded as cases of ENS. In Section 2.6, I follow the same methodology, but focus on the concept of inheritance. I show that cases in which there is no heritability (or more precisely no biased inheritance) between parent and offspring should not necessarily be regarded as cases in which there is no natural selection.

2.2. Minimal and Pure ENS

In his important recent discussion on natural selection and associated evolutionary processes, already presented in Chapter 1, Godfrey-Smith (2009b) drawing on a long tradition of summaries for ENS, defines a Darwinian population, that is, a population that has the capacity to undergo ENS, in the *minimal sense* as:

> [A] collection of causally connected entity things in which there is variation in character, which leads to differences in reproductive output (differences in how much or how quickly entities reproduce), and which is inherited to some extent. Inheritance is understood as similarity between parent and offspring, due to the causal role of the parents. [...] Any Darwinian population will have these properties plus others. (p. 39)

This set-up for ENS expresses Godfrey-Smith's particular views on a number of philosophical issues in evolutionary theory. For example, differences in fitness are ultimately regarded as differences in reproductive output. Although, I criticised this standard view it in the previous chapter, I will accept it initially (sections 2.2 and 2.3) in this chapter for expository purposes. One could also question the necessity for entities, to be causally connected. I will also accept this for now but will question it in Chapter 6.

Another striking feature of Godfrey-Smith's setup of a minimal Darwinian population, but in line with his evolutionary nominalism, is that perfect inheritance between parents and offspring is seen as unnecessary. This choice seems at first natural since in any real population perfect or quasi-perfect inheritance of most traits seems to be the exception rather than the rule. Following Godfrey-Smith's definition, a population in which there is perfect inheritance between parents and offspring should thus be regarded as a special case of minimal Darwinian population. In fact, with perfect inheritance not only do parents cause their offspring to resemble them as

required by the setup, but the resemblance is perfect. Perfect inheritance seems thus to be a very special case of minimal ENS.

Godfrey-Smith's general strategy with his Darwinian space is to use a 'prototype and similarity' approach to ENS. The minimal Darwinian population is a sort of benchmark against which populations are compared. Although useful in many respects, this approach, which from now I will refer to as the *population perspective* on ENS, has some limits. One of them, relevant for my purpose, is that if ENS should be expected in a minimal Darwinian populations, it might well be the case that other evolutionary forces are partly responsible, in concert with natural selection and in a substantial way, for the production of the total evolutionary change observed (if there is any). This is in fact what happens in a number of cases reviewed by Godfrey-Smith (2007, 2009) in which there is heritable variation in fitness, but no evolutionary change is observed. In those cases, at least one other force 'pushes' in the opposite direction to natural selection. The minimal set-up and more generally the population perspective on ENS, seem unable, on their own, to account for those cases unless one recognises that different evolutionary processes are taking place at the same time. We will examine some of those cases in more detail in Section 2.6.

Another limitation of Godfrey-Smith's approach is visible when he explicitly criticises the replicator view of ENS (2009, 31-36). As I noted earlier, the replicator framework presupposes a (quasi-)perfect transmission of types across generations. Although Godfrey-Smith's criticism is targeted towards the idea that *ENS* does not require perfect transmission, Godfrey-Smith ends up claiming that perfect inheritance is not necessary for *Darwinian explanations* (2009, 36) in general. While I agree with the claim that Darwinian explanations in general do not require the faithful transmission of traits across generations, this does not demonstrate that the part of the explanation due solely to natural selection does not require the faithful transmission of some trait

at some level. Indeed, the fact that there is imperfect inheritance might be entirely causally explained by evolutionary forces distinct from the process of natural selection.

To determine whether ENS requires perfect transmission across generations, I propose that one must use a complementary approach to that of Godfrey-Smith, one that will allow us to evaluate in the first instance the concept of ENS from what I call a *process perspective*. The goal of the process perspective is to delimit what the minimal requirements for a population to exhibit nothing but natural selection (what I will call *pure ENS*) are. Thus, the goal here is to assess whether the evolutionary change observed in a population with imperfect transmission of traits across generations is compatible with pure ENS. One way to carry out this evaluation is to consider a population (albeit a highly idealised one) of entities in which all the evolutionary forces except natural selection have been stripped down: that is, one without mutation, migration and/or drift. In such a case, any remaining evolutionary change in this population will be ENS in its *pure* form. I call this class of populations, which is a subclass of minimal Darwinian populations, *pure Darwinian populations*.

Table 2.1. The difference between the population and process perspective on ENS.

	Population perspective	Process perspective
Corresponding Darwinian population	Minimal Darwinian population	Pure Darwinian population
Evolutionary cause(s) present in the population exhibiting ENS	Natural selection and possibly other evolutionary forces	Nothing but natural selection
Targeted explanation	What the minimal properties of a population to exhibit *some* ENS (minimal ENS) are	What ENS is/What the minimal proprieties of a population for it to exhibit *nothing but* ENS (pure ENS) are

2.3. Two cases of evolution

With these distinctions and definitions in place, we now have at hand all the necessary tools to evaluate whether different cases of evolutionary change with imperfect inheritance are cases of pure ENS. Let me reiterate that the strategy employed here to 'test' this hypothesis is to evaluate whether a population is a *pure* Darwinian populations since, as I argued in the previous section, establishing that a population is a *minimal* Darwinian population does not represent a test of this hypothesis (see the line 'targeted explanation' of Table 2.1). Using a toy model, I start my investigation with two cases of population in which there is imperfect inheritance of variation in reproductive output and show that, for both, natural selection is either absent in the population (non-Darwinian population) or present but mixed with another evolutionary force (non-pure minimal Darwinian population) which I suggest should be associated with the classical force of mutation in population genetics. Further on, I establish more systematically that pure cases of ENS are incompatible with imperfect transmission and thus that evolutionary change due to

imperfect transmission should be attributed to an evolutionary force conceptually distinct from natural selection.

To start off with these two cases, let us imagine a population of entities in a deterministic world (or one sufficiently close to deterministic) in which every entity has a different height. Suppose that each entity reproduces asexually, simultaneously and that there is no overlap of generations (when an entity reproduces it immediately ceases to exist). Suppose also that each entity produces a number of offspring that is a linear function of its height (or any other character): The taller the entity, the greater the number of offspring it produces. Finally, suppose that there is heritability on height in the sense that offspring on average, resemble their parent more average more their parent than they resemble other entities of the parental generation (positive slope of regression line). Call this model '**EC2**'.[50]

Before going further I need to make a small digression on the concept of heritability. The concept of heritability used in evolutionary theory will be a topic I will treat in detail in Chapter 4, but it is crucial to have an idea of what it is in this chapter. There are roughly two notions of heritability used in the literature, namely broad-sense heritability (H^2) and narrow-sense heritability (h^2). In evolutionary theory only the notion of narrow-sense heritability is of interest (for more details see Downes 2009a; Sesardic 2005 and Chapter 4). There are two definitions of narrow-sense heritability. Under one h^2 is the ratio of additive genetic variance[51] on phenotypic variance (Falconer 1981). Under the other, which is more abstract, but formally linked to the first

[50] This is a more detailed version of Godfrey-Smith's (2009, 24) model.

[51] A less biologically centered definition would be the additive variance of intrinsic-invariable characters in a given environment. This is because genes are only one instance of intrinsic-invariable entities that can undergo the process of natural selection. In principle any kind of intrinsic-invariable entity can under this process (see Chapter 4).

one, h^2 represents the slope of regression of average offspring character on parental or mid parental character (Godfrey-Smith 2007, 2009b; Okasha 2006; Rice 2004). This latter definition has been favoured by many authors in evolutionary theory. I follow suit, although I will have something to say about this choice later on. In our model, heritability is positive but offspring have not the same height as their parent. Note that heritability can be of 1, that is perfect, but some or all parents and offspring might still have different heights. This is because heritability is a statistical measures that relies on averages. For example, if on average offspring have the same phenotypic value as their parent(s), then heritability will be of 1, even though none of the offspring has the exact same phenotype as its parent(s).[52]

For simplicity, imagine that in our Model **EC2** the parental generation is composed of only two entities with different heights (say 2 units and 4 units tall respectively). The fact that I use only two individuals to make my point instead of a population composed of a higher number of individuals is for illustrative purpose only and has no consequences for the conclusions drawn later on. Suppose also that the short entity has two offspring, while the tall entity has four offspring and that each offspring is of a height different both from its parent and any of its siblings, yet heritability is 1.[53]

[52] This is only one example in which heritability is 1, but there are many other possible.

[53] In **EC2** there is no difference in viability between the individuals: they all die simultaneously after a unique reproductive event. That makes **EC2** a 'fertility selection' type model, that is, one in which selection comes primarily from a difference in fertility between individuals within the population. Yet all the reasoning involving **EC2** could have been made using a 'viability type' model, that is, one in which selection comes primarily from differences in viability between the individuals.

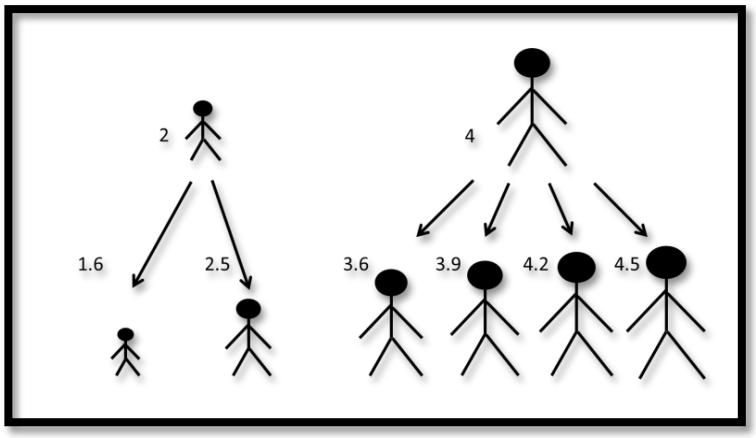


Figure 2.1. Illustration of the model EC2.

Figure 2.1 is an illustration of **EC2**. A simple calculus, using the value of heights presented in Figure 2.1, leads to the conclusion that the population is evolving because the average height at the parental generation is 3 units while it is 3.38 at the offspring generation. Although the population is evolving, can we confidently assume this is a case of ENS as we defined it earlier? And if this case is a case of ENS, should we ascribe all the evolutionary change to natural selection or should we only ascribe part of it? In other words, is it a *pure* case of ENS? As it stands **EC2** is underdetermined and the cases 1 and 2 presented below are possible (but not exhaustive) underlying causal[54] explanations that are compatible with the evolutionary change observed in **EC2**. One of these, we will see, is not a case of ENS while the other is a case of ENS but not a pure one.

2.3.1. Case 1: Evolution due to differences in variable properties

[54] The notion of causality employed here is fully compatible with the manipulationist account of causation (Woodward 2003) presented in the previous chapter.

One possible explanation accounting for the difference in height in **EC2** is that each entity of the population is in a different environmental context (that is, is different with respect to one intrinsic-variable property) causally relevant to height, yet one of which is correlated between parents and offspring. Apart from this difference we suppose no other differences between the entities of the population. Had each entity of the population been in the same environmental context, they would have had the same height. Let us a call this variant of the model '**EC2***'. **EC2*** resembles a case of passive gene-environment correlation distinguished by behavioural geneticists (for more on passive gene-environment correlation see Sesardic 2005).

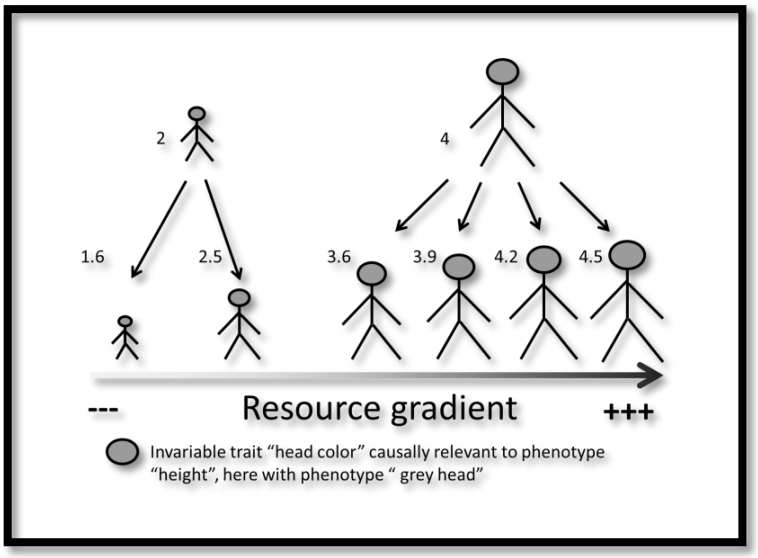


Figure 2.2. Illustration of the model EC2* – all differences in height between entities are due to some differences in intrinsic-variable properties.

Figure 2.2 illustrates this case. Imagine that the trait 'head colour' is causally involved in determining the height of each entity and is intrinsic-invariable within the range of possible environmental states so that two perfect clones growing in any two different states of the

environment have the same phenotype on the trait head colour. Let us also assume that 'head colour' is the only intrinsic-invariable trait involved in the determination of height and that each parent has the phenotype 'grey head' so that there is no variation at the population level in intrinsic-invariable properties between the entities of the population. Suppose also that each entity transmits its intrinsic-invariable properties (the phenotype 'grey head') perfectly to its offspring. Suppose finally that the level of resources received by an entity is causally involved in determining the height of each entity and that there is a gradient of resources in the population (as shown on Figure 2.2): the higher the amount of the resource received, the taller the entity. Each offspring once it is born, moves away from its parent but remains close to it. Because each entity is at a different position along the resource gradient, they all have a different height. Yet because each offspring remains located close to its parents, offspring on average resemble their parent more than they resemble other parental entities.

In this case, we have variation in height, that leads to differences in reproductive output and which is heritable (using the standard regression notion of heritability).[55] Yet this is clearly not a case of ENS, let alone a pure case of ENS. In fact all the differences between entities are differences in extrinsic properties (the amount of resource they obtain depends on their parent's position and their position) and in intrinsic-variable (height). There is also no population variation in intrinsic-invariable properties, thus these differences cannot be attributed to natural selection as per the different distinctions I made in Chapter 1.

This case, although not original (see the very similar although less detailed case 9 in Godfrey-Smith 2007), is useful for our purpose for three reasons. First, it shows us that

[55] See Chapter 4 for the different approaches to heritability.

underlying a continuous trait such as height, there *can* be perfect inheritance of intrinsic-invariable factors (in our case the phenotype 'grey head') within the range of possible states of the environment. Observing a continuous trait is thus insufficient reason to claim that ENS does not require perfect inheritance to occur. This is a fairly obvious point but it is worth mentioning.

Second, it illustrates that unless the different phenotypes of a given trait are intrinsic and invariable within a specified range of environmental conditions, natural selection cannot be invoked as a cause of evolutionary change on this phenotype. If height is an intrinsic-variable phenotype within the different environmental states, over a given period of time, in a given population, and there is no variation in the population on an invariable property causally involved in determining the height of individuals, then height is not a property that is subject to natural selection. This is because whether entities survive or reproduce depends entirely on their extrinsic and intrinsic-variable properties. This point is important for our purpose because the fact that a trait is continuous or discrete is not relevant for ENS if this trait is an intrinsic-variable or an extrinsic one. It only matters if this trait is an intrinsic-invariable one. This means that there is not necessarily any fundamental disagreement with the view that natural selection requires the perfect transmission of types over time and Godfrey-Smith's evolutionary nominalism if the traits under scrutiny, such as height in our case, are not intrinsic-invariable.

Finally, this case demonstrates that even if in a population there is a positive heritability, under the regression approach, and differences in reproductive outputs between the different individuals, this does not necessarily make this population a Darwinian population, since the resemblance between parents and offspring might be due to a correlation between the environment of parents and offspring, rather than caused in the relevant sense by the parents. This suggests that the variance approach to heritability might in some respects be superior to the

regression approach for discriminating the effects of natural selection from drift or correlated responses. This is mainly because the variance approach takes the independent population variable to be an intrinsic-invariable property of entities, which is instantiated by evolutionary genes (see Haig 2012 for details on what this concept entails) while the regression approach takes any population parameter (intrinsic-invariable or intrinsic-variable: that is, any phenotype) as an independent variable. This means that the regression approach to heritability captures more evolutionary processes than natural selection (see Chapter 4 for more details).

2.3.2. Case 2: Evolution due to unreliable channel of transmission

Another possible explanation accounting for the different height of each entity in **EC2** is an unreliable transmission between parents and offspring of an intrinsic-invariable factor involved in the determination of height. Imagine a case in which two entities have a difference in intrinsic-invariable property that is causally involved in their difference in height. In this particular case each intrinsic-variable property causally involved in entities' heights is kept constant across the population so that any difference in reproductive output is due to the differences in intrinsic-invariable property in relation to height. Imagine now that because of the nature of the channel of transmission of height, which is supposed to be intrinsic-invariable and with no variation within the whole population, each offspring produced is different from its parents and different from its siblings, but resemble them more on average than they resemble other entities because of the phenotype of the parent. In other words, in this case, the channel of transmission for height is not perfectly reliable. Let us call this variant of the model '**EC2****'.

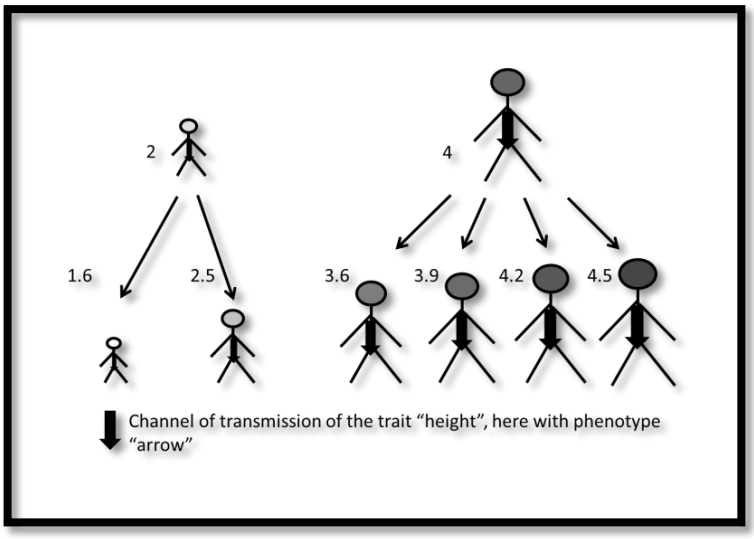


Figure 2.3. Illustration of the model EC2** – all the differences in height between parents and offspring are due to the phenotype of the channel of transmission.

Figure 2.3 illustrates this case in a simplified way. We suppose that in **EC2****, as in the previous case, the trait 'head colour' is the only intrinsic-invariable trait causally relevant in the determination of an entity's height. By 'causally', I mean here, following the manipulationist account of causation encountered in Chapter 1, that this trait is the only intrinsic-invariable difference maker for entities' height in the population. But contrary to the previous case, both entities in the parental generation have a different height because they have a different head colour. Let us postulate that, *ceteris paribus*, height is a linear function of head colour: the darker the head, the taller the entity. Because the two parental entities in Figure 2.3 have different head colours, they have different heights. Suppose also that, contrary to the previous case, there is no resource gradient in the population and that all other variable properties have the same evolutionary consequences on height and reproduction. Although the amount of resource received by an entity might be causally related to its height, this is not an *actual difference maker in this population.* It is part of the environment *sensu* Haig, that is, "all parts of the world that are shared by the alternatives [entities] being compared" (2012, 461). Finally, suppose that each entity

has the same channel of transmission across generations for height (represented by the phenotype 'arrow' in Figure 2.3), which is an intrinsic-invariable property of entities and on which there is no variation in the population that leads them to produce offspring similar but typically not identical to them (for some complex deterministic reasons). Because all the entities have the phenotype arrow they produce offspring with the same pattern of variation for height. Because the entities' heights and the relation between parents and offspring have not changed when compared to case **EC2**, heritability on height, using the regression approach, is in this case still 1. This is because the similarity relations between **EC2** and **EC2**** are strictly the same. Had the phenotype of the channel of transmission been different or had there been some variability in it, heritability would have been different.

As in the previous case there is variation in height that leads to differences in reproductive output and that is heritable. Yet, is this a case of ENS? To answer this question we need to know whether the differences in reproductive output between the entities are due to some difference in intrinsic-invariable property as per our definition given in Chapter 1. This is quite straightforward to establish since all the difference between the reproductive outputs of the two parental entities is accounted for by a difference in height which differs only because of some difference in intrinsic-invariable property (a difference in head colour). This population is thus a minimal Darwinian population and natural selection is causally responsible for at least some of the evolutionary change. With this established, we can now ask whether natural selection is the only causal factor that explains the evolutionary change observed across generations or in other words whether the evolutionary change observed is a pure case of ENS.

Because the channel of transmission for height does not have a high fidelity, new variation is produced in the population every generation by modification of the offspring's intrinsic-

invariable characters when compared to the parent. The production of new variation from existing variation by changing an otherwise intrinsic-invariable character, I propose, can be associated with the notion of 'mutation' which is, as I have already emphasised, one of several possible evolutionary forces among three others: namely, natural selection, drift and migration. Some might disagree that it should be associated with mutation. I will provide a short defence of my claim Section 2.3.3. At any rate, even if one refuses to associate an unfaithful pattern of inheritance with mutation, one should nevertheless agree that it clearly represents an evolutionary process different from natural selection migration and drift, keeping in mind our definitions in terms of difference in intrinsic-invariable properties and differences in extrinsic and intrinsic-variable properties respectively. Thus, if the evolution of height described by our second case depends both on natural selection and mutation, albeit a particular sort due to the existence of a non-random transgenerational unreliable channel of transmission (or for the sceptics another evolutionary process), the evolutionary change observed across generations is not a pure case of ENS. This is a case of ENS with production of new variation at each generation. Although the production of new variation is mandatory for complex adaptations to arise and thus make an evolutionary process 'interesting', it should be conceptually distinguished from natural selection (and it is in fact conceptually distinguished in population genetics).

EC2** is reminiscent of a point made by Godfrey-Smith and Lewontin (1993) in which they argue that in a case of sexual reproduction with at least two alleles at one locus (**A** and **a**) and a difference in reproductive output between the three different genotypes at that locus (**AA**, **Aa** and **aa**), the evolutionary change (or the lack of evolutionary change in some cases) observed is due both to sex and natural selection. Without knowledge about the genetic determinism (contrary to Godfrey-Smith and Lewontin's case where it is known), but only of the different

phenotypes, sexual reproduction would be considered an instance of directed mutations across generations of the same kind as in **EC2**** (Mendel's law of segregation being only one among the infinite possible 'laws of transmission across generations'). In the case described by Godfrey-Smith and Lewontin however, alleles are intrinsic and invariables across generations and thus it is useful to move back and forth between the genetic and genotypic level to better understand the evolutionary change observed. In their case, it can even be argued that, assuming panmixia, all the evolutionary change observed is due to ENS at the gene level without the need to invoke other evolutionary force (e.g., sex) since the regime of reproduction does not represent an actual difference maker.[56] In fact, on average, each type of allele is in the same environment and exhibits the same phenotype in each state of the environment. Yet, because in case 2 we postulated no such invariability of lower level 'entities' across generations, (that is, each intrinsic-invariable property from all levels considered is transformed by the channel of transmission during reproduction) this reductionist move is impossible and one must invoke at least two evolutionary forces to account for the evolutionary change observed (intragenerational selection and transgenerational mutation).

2.3.3. No pure ENS without perfect inheritance across generations

So far I have established that in two cases of evolutionary change in which there is variation, difference in reproductive output and heritability, one (**EC2***) is not a case of ENS, when all the evolution is due to some population variation in extrinsic and intrinsic-variable properties of entities (the population is not a Darwinian population), and the other (**EC2****) is not a pure case of ENS (the population is a minimal Darwinian population but not a pure

[56] Assuming also an infinite population size so that there is no drift.

Darwinian population) since the evolutionary change observed is due partly to natural selection and partly to the production of new variation through an unfaithful channel of transmission for height. Yet, in and of themselves, these two cases do not demonstrate that imperfect inheritance is incompatible with pure ENS and thus that it cannot in principle be exhibited by a subclass of pure Darwinian populations. In fact, at this stage, one could still argue that the two cases I proposed are the result of cherry picking without scope for generalisation. This is a valid objection and to address it, instead of finding new cases of evolutionary change without perfect inheritance across generations and establishing that they do not represent cases of pure ENS, one must provide a demonstration that *any* case in which there is no perfect transmission between parent and offspring necessarily involves an evolutionary force conceptually distinct from natural selection.

To do so, let us start with the requirements for a population to be a pure Darwinian population. For a population to evolve *solely* by natural selection there must be in this population:

1. Natural selection i.e. existing variation in intrinsic-invariable characters within the population's ecological environment that leads to differences in reproductive output;
2. No drift (or correlated response) i.e. no existing variation in intrinsic-variable and/or extrinsic characters within the population's ecological environment that leads to differences in reproductive output. If there is drift together with natural selection, we will have a mixed case of evolution falling in the class of minimal Darwinian population but not the subclass of pure-Darwinian population;
3. No production of new variation or mutations in the population since as with the second requirement, it would mean that this population, assuming there is variation in intrinsic-invariable properties, is a minimal Darwinian population, not a pure one.

Demonstrating that at least one of the three requirements above is incompatible with imperfect inheritance will be sufficient to show that imperfect inheritance is incompatible with a pure case of ENS and thus incompatible with the process of ENS in and of itself. The demonstration is quite straightforward: requirements 1 and 2 are compatible with a form of inheritance that does not presuppose any perfect transmission of intrinsic-invariable properties across generations. In fact there is nothing in these requirements that precludes an offspring from being of a different type from its parent. Existing differences in intrinsic-invariable properties without existing differences in intrinsic-variable or extrinsic properties are logically compatible with imperfect transmission of those differences. However requirement 3 is incompatible. In fact, all conceivable cases involving imperfect transmission of intrinsic-invariable properties will necessary lead to the production of new differences of intrinsic-invariable properties in the population with the production of new offspring entities. If natural selection is associated with *existing* differences in intrinsic-invariable properties, the *production* of such variations during an act of transmission and the evolutionary change it leads to should be conceptually distinct from it. In other words, for natural selection to occur there is no need for new variation to be produced. New variation only allow for natural selection to occur indefinitely, but whether the process of natural selection can occur indefinitely is a question conceptually distinct from the question of what the process of natural selection in and of itself is.

I have suggested with the case **EC2**** that the evolutionary change resulting from the production of new variation during imperfect inheritance should be associated with mutation as understood in population genetics. Although mutations are classically thought of as occurring randomly and due to some factors external to the population, I follow Godfrey-Smith (2007) and Mameli (2004) in their views that there is no fundamental reason to think this. As Godfrey-Smith

puts it, "Darwinian evolution can occur on variation that is directional, even adaptively "directed." In these cases natural selection may have less explanatory importance than it has when variation is random, but it can still exist" (2007, 493). Elsewhere, Godfrey-Smith also writes that "[m]utation processes subtly change intrinsic character" (Godfrey-Smith 2009b, 55). If we replace 'intrinsic' by 'intrinsic-invariable' in my framework, we can see that imperfect transmission satisfies the definition of mutation given by Godfrey-Smith. Imperfect transmission subtly changes the intrinsic character that the offspring would have had if transmission was perfect. I thus generalise here the proposition made with **EC2**** and claim that any case in which there is imperfect inheritance will necessarily lead to the *production* of differences in intrinsic-invariable properties in the population that should be associated with the evolutionary cause of mutation rather natural selection. Thus, because requirement 3 for pure ENS cannot be reconciled with the unfaithful transmission of characters across generations, we have here the demonstration that pure ENS requires the faithful transmission of traits across generations.

It should be noted that the notion of mutation I defend here is not a 'molecular' concept of mutation, in the same way that the notion of gene I use throughout the thesis is not a molecular concept of gene. Rather, I defend the evolutionary concept of mutation of which point mutations in the DNA are one instance.

One could argue at this point that because I *stipulate* that imperfect transmission should be understood as (transgenerational) mutation, the result found here is trivial or just definitional *fiat*. But that would be unfair. In fact, it would miss the point that, by providing a reasonable attempt to make sense of evolutionary processes of natural selection, drift and mutation in causal terms using the variable or extrinsic /invariable distinction, I *did not* stipulate that imperfect transmission over generations should be associated with mutation. It is only because imperfect

transmission can be understood as producing new intrinsic-invariable properties in the population that it led me to link it to concept of mutation. Without the distinction made between natural selection, drift and mutation in terms of variable, extrinsic and invariable properties, that is, by sticking to purely statistical methods, it would have been impossible to understand what (different) evolutionary processes were at play in **EC2*** and **EC2**** and their relation to the process of natural selection.

2.3.4. The replicator and classical approaches to ENS reconsidered

I have shown so far that for ENS to occur with no other evolutionary process being involved, the perfect transmission of traits between generations is necessary. To get to this conclusion, I have distinguished two perspectives on ENS, one that focuses on the existence of natural selection in a population and another that focuses strictly on the nature of the process leading to ENS and the part of evolutionary change it is causally responsible for. I have also provided a new approach to understand the distinction between natural selection and drift (Chapter 1) and clarified the links between imperfect inheritance and the force of mutation classically used in population genetics (subsections 2.3.2 and 2.3.3).

As I showed in Subsection 2.3.3, in a minimal Darwinian population, because there is imperfect transmission from parents to offspring, the total evolutionary change must be separated into at least two parts: natural selection and mutation. In some simple models of viability selection with discrete generations, selection and trans-generational mutation are two processes that act alternately over time. First natural selection chooses the fitter variants; then the channel of transmission produces new variation in the population. ENS represents, in this case, the change in frequency of variants *within* the generation while between two generations new

variation is produced by the unreliable channel of transmission. The variation newly produced is then selected again, but this process of ENS is a new one for the intrinsic-invariable properties of the entities to select in the population are now different.

In cases of minimal Darwinian populations in which there is fertility selection (such as the one used in Section 3.2) or in cases where generations are not discrete, selection and mutation via unfaithful transmission occur simultaneously. In those cases the only possible way natural selection and mutation can be explanatorily dissociated, in my view, is by understanding them as the relative parts of the evolutionary change explained by each evolutionary process in a model as close as possible to the target system in which both natural selection and mutation are causally distinct evolutionary process (see Chapter 5 for more on natural selection and models). Thus, it seems that the only possible form of pure ENS over more than one generation is a case in which there is no mutation between generations so that parents pass on their traits perfectly at the next generation which vindicates the replicator framework.

It is now time to relate this conclusion to the initial oppositions between the classical accounts and the replicator approach to ENS discussed in the introduction. As I pointed out, Godfrey-Smith (2009, 31-36) in arguing for evolutionary nominalism opposes the classical approach to the replicator approach. Yet, once the distinction between the population and the process perspective on ENS is made and natural selection is conceptually separated from mutation, one can see that the replicator and the classical approach need not be opposed to each other. In fact, although replication at some level seems necessary for pure multigenerational ENS (that is, what the process of natural selection essentially leads to), it is not so for minimal ENS (that is, what the minimal requirements are necessary for a population to exhibit *some* ENS). I see Godfrey-Smith's overall project as providing us with the tools necessary to predict when and how

natural selection, once working in synergy with other evolutionary forces, will produce complex (that is, multi-step) adaptations. Complex adaptations require the production of new variation over time. Without it, adaptations become mere changes of frequency of pre-existing variation (distributive evolution) and natural selection without mutation ultimately leads to the exhaustion of that variation (see Chapter 3). Yet, from a process perspective, natural selection only leads to a change in type frequencies which might seem unappealing to some. But using Godfrey-Smith's approach, if not complemented by a process perspective on natural selection, will not always allow prediction of what part of the evolutionary change observed should be attributed to natural selection as opposed to other evolutionary causes and will hinder our understanding of it.

The classical approach should thus, in my view, be the one privileged to understand the origin of complex adaptations. Invoking replicators in those explanations is not necessary because complex adaptations *can* result without any entities having been replicated. And it is thus not surprising that Godfrey-Smith endorses a form of evolutionary nominalism to explain away complex adaptations in a general way. On the other hand, if one is willing to decide whether natural selection is responsible for the evolutionary change observed and if it is, to what extent, the replicator framework seems much more appealing. Both perspectives are thus important to fully understand in what sense natural selection is explanatorily relevant in evolution. Better yet, each perspective can illuminate the other.

2.4. Inconsistent Definitions of Natural selection

I now move to a second sort of problems in my attempt to define the process of natural selection, namely whether reproduction and inheritance are required for it to happen. The general

mechanism I presented in the Introduction of this thesis by which natural selection leads to complex (that is, multi-step) adaptations is hardly mysterious. Here is a very clear way to state it provided by Godfrey-Smith (2009b, 43):

> […] natural selection can reshape a population in a way that makes a given variant more likely to be produced via the immediate sources of variation than it otherwise would be. Selection does this by making intermediate stages on the road to some new characteristic common rather than rare, thus increasing the number of ways in which a given mutational event (or similar) will suffice to produce the characteristic in question. Some kinds of novelty can be produced easily by an evolutionary process without this role for selection, but other kinds – complex and adapted structures – cannot.

This clearly illustrates the necessary role of mutation and natural selection in the origin of adaptation (the mixed case of transformative and distributive evolution). Yet, the definition of the process of natural selection *in and of itself* is not that clear.

Evolutionary biologists have proposed a range of different definitions of this process. Endler for example (1986, 33) defines natural selection as "a *process*, *resulting* from heritable biological differences among individuals which can lead to genetic changes"(emphasis in the original). Haig claims that "[n]atural selection requires heritability of the *causes* of variation in fitness" (2012, 466) and it "extracts the average additive *effects* of genes as the environment 'chooses' among phenotypes"(2012, 475). For Rice (2004, 168) "there is [natural] selection when covariance between fitness and phenotype is due to some causal effect of phenotype on fitness". Futuyma (2005, 251) defines natural selection "as any consistent difference in fitness among phenotypically different classes of biological entities". Philosophers of biology have also proposed definitions of natural selection. Beatty (1984, 190) for example writes that "[n]atural selection is, as its name suggests, a discriminate form of sampling – a sampling process that

discriminates, in particular, on the basis of *fitness* differences." This definition of natural selection has also been defended by Millstein (2002, 2006). Brandon (2005, 2014), relying on Brandon (1990), defines natural selection as "differential reproduction that is due to differential adaptedness (or fitness) to a common selective environment". Finally the definition of the principle of natural selection I proposed in Chapter 1 under the acronym of 'PNS' is "the process of natural selection will occur in a population if and only if there is at least one difference in intrinsic-invariable property, in an environment **E**, between two or more individual entities of the population that can be causally linked to a downstream difference in reproductive output (or more generally relative growth)."

Two interesting facts can be noted when these various definitions are compared. First, in some definitions, inheritance, and more particularly one of its measures, heritability, is considered part of the process of natural selection (Endler's and Haig's definitions), while in other definitions it is not (Rice's, Brandon's, Futuyma's, Millstein's, Beatty's definitions, as well as my version of the PNS). There seems thus to be some variation in the meanings attached to the notion of natural selection as some authors understand it as a process that can occur only intergenerationally while this feature does not appear fundamental for others. Second, most of these definitions include the notion of fitness differences. This means that the concept of natural selection critically depends on what one means by 'fitness differences'. This is intriguing because there is also a great variance of opinions on what fitness is (for reviews of different concepts of fitness existing in the literature see for example Abrams 2012; Michod 1999, 223-225; see also Van Valen 1976 and Chapter 1), with some authors doubting that a general consensus can be found (e.g., Godfrey-Smith 2009b; Krimbas 2004; Stearns 1976).

We had an overview of this variety of opinions in Chapter 1, but it is worth restating that some regard fitness as a concept ultimately bound to the notion of reproductive output. Differences in fitness, for them, necessarily lead in some way or another to differences in reproductive outputs between entities varying on some trait. This is the majority position defended by, among others, Godfrey-Smith (2009b, 30) and Brandon (2014). This was the definition used so far in this chapter, but I will now challenge it. Others regard fitness as a predictive statistical concept that is associated abstractly with a growth rate (e.g. Matthen & Ariew 2002). For them the notion of natural selection could be attached to any collection of types that grow differentially. We saw the kind of problem this view leads to in the previous chapter, but also earlier in this chapter: a purely statistical notion of natural selection will accept cases that intuitively belong to drift. Finally some others such as Bouchard (2008, 2011; see also Van Valen 1976) regard differences in fitness as being neither necessarily bound to differences in reproductive outputs nor as an essentially statistical concept. For Bouchard, differences in fitness cause ENS, but these differences can be the result of differences in survival or differences in developmental growth rates alone. Critically, Bouchard relies on the notion of ecological fitness put forward by Bouchard and Rosenberg (2004), that I have made more precise in the previous chapter, in which fitness is understood as a relational property between two organisms and measures their relative ability to solve design problems set by the environment. Put another way and using the notion of fitness I proposed in Chapter 1, fitness measures a difference in intrinsic-invariable properties that leads to differences in relative growth rate between two or more types. Under this definition, a difference in reproductive output is only one possible effect of fitness differences.

With such a variety of opinions (and the ambiguities they lead to) surrounding both the definitions of natural selection and fitness, it is thus not clear whether inheritance (or more generally transmission over time) and reproduction should be included as parts of the process of natural selection and if not what their relations to it are. At the same time, in spite of the divergences in opinions, it is often assumed that natural selection is a general principle, process or algorithm[57] that is 'substrate neutral' (Dennett 1995, 51) and can be applied to different levels of organisation (Lewontin 1970) as well as any temporal scale. One example of a domain different from biology in which the principle of natural selection has been proposed to apply is cultural evolution (Blackmore 1999; Dawkins 1976; Godfrey-Smith 2009b, 2012; Laland & Brown 2002; Mesoudi, Whiten, & Laland 2004; Richerson & Boyd 2005). If this view is right (as I will assume it is) and natural selection is to be considered as a general principle, I argue in the next two sections that neither reproduction nor inheritance, as classically understood in evolutionary theory, should be included as components of the process of natural selection. Rather, they should be seen as contributing to other processes of evolutionary change. These processes once combined with natural selection can in some cases (but not necessarily) produce evolutionary change and complex adaptations. Although natural selection is necessary for complex adaptations, complex adaptations should be regarded only as *possible* outcomes of the process of natural selection.

2.5. Why reproduction is not essential to natural selection: the problem of individuality

[57] I consider here 'principle', 'process' and 'algorithm' to be roughly synonymous.

Imagine a population of entities, in which two types exist. One type '**R**' *reproduces* with perfect inheritance in discrete generations, but the offspring produced are made of parts of their parent. As a result, they are twice as light as their parent and only able to reproduce once they have reached the mass their parent had before reproducing by gathering resources in their environment. The other type '**G**' does not reproduce but *grows* and gets larger over time. More precisely it doubles its mass every time the other type reproduces (see Figure 2.4). We suppose now that there is no survival advantage in growing or reproducing. Is natural selection occurring in this population? For sure, the type **R** will soon be represented by a much higher number of individuals in the population. Starting from a population of 10 individuals of each types, after 10 generations only, there are 10240 individuals of type **R**, while only 10 of type **G**, assuming that the viability of each organism is very high. There is here a difference in reproductive output between the two types, but the organisms of each type are massively different: one is 1024 times heavier than the other. But is this difference a fundamental one with respect to natural selection? In other words is there a process of natural selection occurring in this case?

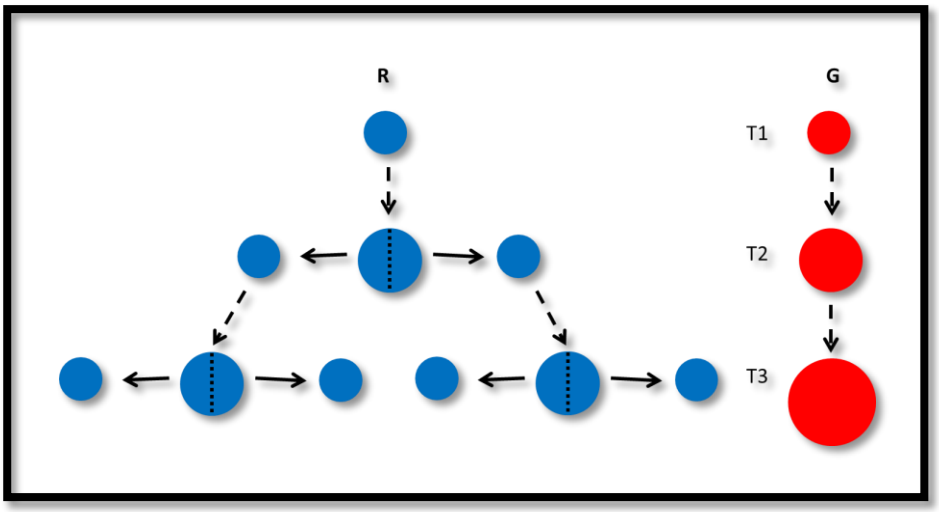


Figure 2.4. Illustration of the model with the types R and G over three units of time.

This theoretical case is inspired by an example proposed by Van Valen (1976, 182) which vividly demonstrates the problem associated with considering reproduction as a feature ultimately bound to the concepts of fitness and natural selection:

> [C]onsider two seedlings of a grass. Each spreads by tillering to occupy a field. In one field the tillers are retained; in the other they are lost. Genetical theory would call the second population and its genotype vastly more fit because there are many more individuals. Yet there is no more than a trivial biological difference. Degree of connection has no sharp demarcation and we can't get around the quasi-problem by calling each shoot an individual; what then should we do with a bush? The natural solution is ecological.

As Van Valen asserts one cannot get around the problem of measuring fitness posed here by claiming that each shoot is an individual for there are no clear criteria to define individuality. And in fact, some organisms challenge our intuitions about individuality. A good example of this is the now classical case of 'slime moulds' formed by the amoeba *Dictyostelium discoideum* (Bonner 1959; Buss 1987). Slime moulds are temporary structures, formed by unicellular amoeba. They

behave like an individual and reproduce as such by forming 'fruiting bodies'. Yet it is hard to make them fit our intuitive concept of individuality as they should, for example, be considered as individuals made up of a large number of different parents. This kind of problem is even more striking in the case of evolutionary transitions in individuality (which I will discuss in more detail in Chapter 6) in which the criteria used to decide when there is a genuine transition from individuality at the level L to individuality at the level L+1 are to some extent arbitrary. Okasha (2006, 237), for example, is close to admitting thus since he claims that there is a 'grey area' between the stage at which the higher level entity is clearly not an individual and the stage at which it is clearly one. Other issues surrounding the concept of individuality in relation to evolution are treated in Godfrey-Smith (2009b, 70-81)

Some authors, confronted by these conceptual difficulties in defining the concept of biological individual, have nevertheless tried to come up with systematic criteria integrated into a framework. It has been argued for example that developmental bottlenecks and the presence of a germ-soma separation (or something equivalent) could be important criteria (along with some others) to determine what an individual is, or rather to what extent an entity is an individual. Godfrey-Smith (2009b, 91-95) uses these two criteria, along with another one he calls 'integration' to build a three dimensional Darwinian space, in a similar fashion to the one presented in the previous chapter, that helps in visualising the degree of individuality of any given collective of entities. His view, as in with Darwinian populations, is that the more highly a collective scores on developmental bottleneck, germ-soma segregation (or equivalent) and integration, the more paradigmatically this collective is a biological individual.

Although there is a sense in which individuality is higher when developmental bottlenecks and germ-soma segregation are present (they facilitate complex adaptations), we know since the

work of Buss (1983, 1987) that developmental modes with a germ-soma separation represent only a minority of ways in the whole spectrum of possible developmental modes to produce multicellular organisms. We also know that some species of fungi and plants such as the quaking aspen *Populus tremuloides* (Bouchard 2008, 2011) have very long life spans in some environments. With those organisms, there is literally no reproduction followed by developmental bottleneck. Yet they represent strong prima facie cases of ENS.

Contrary to Godfrey-Smith, Queller & Strassmann (2009) propose that only two parameters are necessary to define individuality,[58] namely a high degree of cooperation and a low degree of conflict between the parts of the higher level entity. A high degree of cooperation at the lower level leads to adaptations at the higher level and a low degree of conflict means that these adaptations are not disrupted. Although appealing in its simplicity, this framework is not without problems. For example, it leads to the very unintuitive view that, a Tasmanian devil (*Sarcophilus harrisii*) is less an individual than a whale or mouse, because although their cells are highly cooperative, some conflict can emerge easily within them through the transmission of fatal facial cancer (McCallum & Jones 2006).

In a similar way to Godfrey-Smith, Queller & Strassmann use a space (two dimensions instead of three for Godfrey-Smith) to characterise individuality based on degree of cooperation and degree of conflict. Contrary to Godfrey-Smith, they regard developmental bottlenecks and germ-soma separation along with other parameters,[59] as non-essential features for their definition but rather assume that they are, for some, means to reach a low degree of conflict and a high

[58] They actually prefer to use the notion of *organismality* over *individuality* but consider these to be roughly synonymous.
[59] Other parameters include: physical contiguity, indivisibility, clonality or high relatedness, short-term and long-term genetic cotransmission and membership in the same species.

degree of cooperation at the lower level. Another important difference between Godfrey-Smith's and Queller & Strassmann views' is that contrary to the former, the latter does not make explicit reproduction of the higher level entity as being necessary for scoring high on individuality, for the notions of highly cooperation and low conflict do not entail reproduction. This suggests that some biological entities that do not reproduce should be considered as individuals.

We can thus see that in both accounts (for another account called the 'export of fitness view' see for instanceFolse & Roughgarden 2010; Michod 2005 and Chapter 6 where this account is presented in more details),[60] not only is there no unanimous definition of individuality, since Godfrey-Smith and Queller and Strassmann disagree to some extent as to what features are the most important ones to define individuality, but also that the frameworks they develop do not lead to an *absolute* definition of what an individual is for the distinctions they make are relative ones. In fact, those frameworks can be used to evaluate whether a particular collective is *more* or *less* an individual than another collective. The lack of a clear cut definition seems to be a problem if fitness, and consequently natural selection, is defined in terms of new individuals produced. In fact, if it is not clear what an individual in a species with an 'atypical' life cycle is, it will be consequently hard to assess their reproductive output and thus whether natural selection occurs at all in populations made of those organisms and if the evolutionary change observed should be attributed to natural selection or not. I will come back to individuality in Chapter 4 where I will show that depending on what concept of individuality one has, in spite of the same process occurring one ends up with very different predictions of evolutionary change.

[60] The concept of biological individuality is treated in detail in Bouchard and Huneman (2013).

It is *possible* to marginalise the case proposed by Van Valen's, as well as any other case that does not fit the view of fitness as reproductive output and natural selection as resulting from differences in reproductive outputs. There are, however, some important problems with this solution. First, explanations in terms of natural selection will be limited for those 'marginal' cases. In itself, this represents a problem since many of them have complex structures that supposedly are the result of natural selection or a similar process. Second, one worry with this solution (if it is really a solution) is that more and more 'marginal' cases are discovered until they form the majority of cases. Bouchard's (2008, 2011) favourite example of what Godfrey-Smith would call a 'marginal case' of an individual that does not reproduce is the quaking aspen, but there are many similar cases in plants and fungi. One example among hundreds is the case of the herbaceous shrub *Haloragodendron lucasii* (Sydes & Peakall 1998) which is found only in North Sydney and seems to consist entirely of a handful of clones, with no known sexual reproduction.[61] Third, it is not clear in what sense there *could* be a marginal process of natural selection. Godfrey-Smith (2009) speaks about marginal cases of *ENS*, which is not objectionable if this means "evolution in which natural selection has played a non-negligible role in producing evolutionary change". But the concept of complex adaptation as I have outlined it earlier should be distinguished from the concept of natural selection. Knowing that ENS is occurring in a population and whether it plays a minor role or major role in the production of complex structures is a different question from knowing what natural selection is in and of itself. Both questions are legitimate.

It should be noted that survivability (or more generally persistence) and developmental growth are in some cases a function of the reproduction of lower level entities constituting a

[61] I thank Mark Olson for providing me this example.

higher level entity. In those cases one might argue that ultimately they reduce to a case in which reproduction is necessary. For instance imagine that our types **G** and **R** are multicellular organisms, which we suppose for simplicity are made of undifferentiated cells. Let us suppose that the only difference between the two types is that the type **R** is less cohesive than type **G** (see Figure 2.5 for an illustration). Despite this difference the cells of both types reproduce with the same rate. As a result of this difference in cohesiveness when described from the higher level perspective, we observe one entity (or individual) that persists and grows over time, while the other disappears, leaving two offspring. Yet, this represents merely an epiphenomenon of cellular reproduction.

Although possible in some cases, the solution of explaining a process of natural selection from reproduction at a lower level will not always represent a good alternative for at least two reasons. First, a lower level description might not always be possible. This is especially true if we intend our conceptual tools to be used in other contexts than the biological context (e.g., cultural evolution, see Section 7.2). There could be, from a higher level perspective, no difference between an entity merely surviving or growing due to cellular reproduction and an entity surviving or growing for other reasons which everywhere we look would not involve any reproduction. Yet, in one case we would argue that because there is a commensurability with reproductive output at the cellular level, this entity participate in a selection process while it does not in the case where a description in cellular terms is not available. If ecologically the type growing by cellular divisions is similar to the type growing without reproduction being involved then this seems to be an arbitrary distinction. Growth, instead of being seen as a surrogate for reproduction or a kind of reproduction, should rather, in my view, be regarded as another equivalent form allowing entities to participate in a selective process that can itself produce ENS.

Second, in Van Valen's case (and many other cases), although there is production of new shoots over time, in the case where the tiller does not disappear, it is not clear that this represents a case of reproduction at the shoot level and growth at the higher level. At best it is ambiguous and thus this further fuels the view that reproduction and growth are commensurable and should be regarded as equivalent, rather than growth being seen as a function of reproduction. There is no fact of the matter as to whether a shoot represents a new individual or whether it simply is a new part of the larger organism. Attempting to make it a case of reproduction can be done for heuristic purpose, but it would be a mistake to claim that the seedling of grass does reproduce or does grow given a controversial notion of individuality. Rather, it should be remarked that in a large population where there would be different types of tillers, the most successful type will be the one that reproduce and/or developmentally grows at a faster rate due to its intrinsic-invariable properties. One can choose to define reproduction and growth according to specific definitions but in any case, both should be considered when examining the process of natural selection and ENS.

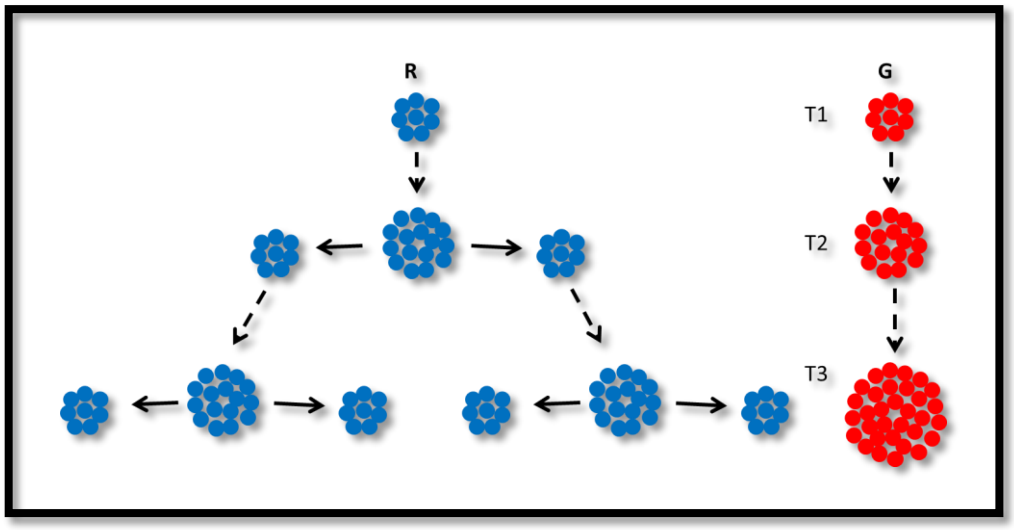


Figure 2.5. Illustration of the model with the types R and G, from a lower level perspective, over three units of time.

Missing this latter point can render some particular evolutionary phenomena hard to interpret under a classical view in which reproduction is seen as essential for natural selection to occur since two answers depending only on one's interpretation of 'reproduction' would be available. By noticing that developmental growth can be ecologically equivalent to reproduction, and that the number of individuals produced only represents a good proxy of fitness under some specific cases, we avoid this problem. This problem should be linked to the ontological interpretations (Michod 2005; Okasha 2006) of the concept of multilevel selection 2 initially developed by Damuth and Heisler (1988), which I criticise in Chapter 6.

Although my discussion has mainly concerned developmental growth, the same sort of argument can be run with respect to persistence. We can imagine a case in which every cell that dies is replaced by a new one, so that what we call persistence is in fact nothing less than a perpetual reproduction where the process of reproduction occurs *little by little* (see Figure 2.6 for an illustration of the equivalence of this mode of reproduction to a classical mode of

reproduction). As noted by Godfrey-Smith (2009b, 104), some metaphysicians argue that persistence of any physical thing is the result of earlier temporal stages causing later ones. This means that persistence *might* be seen as a form of reproduction. Yet, whether natural selection occurs cannot depend on the willingness of interpreting the same physical phenomenon as reproduction or persistence. This remark also gives us a good incentive to consider persistence as equivalent to reproduction for the same reasons developmental growth is.

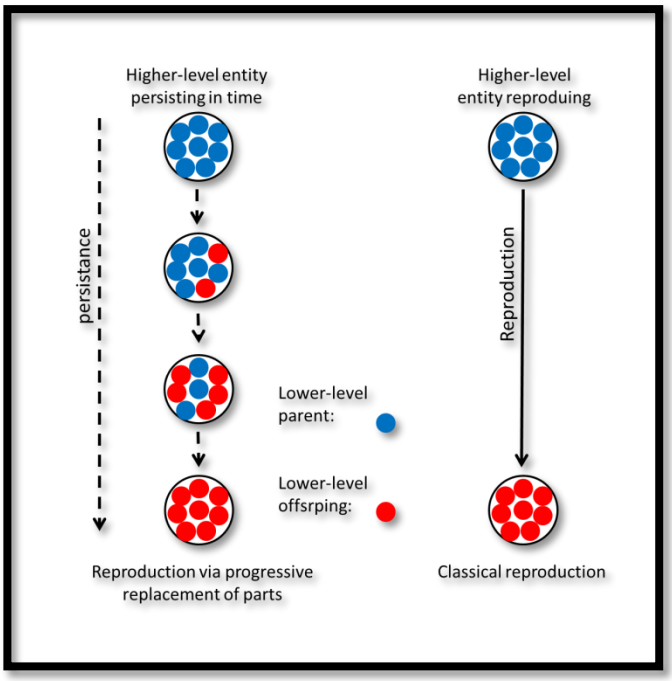


Figure 2.6. Illustration of the equivalence between the mode of reproduction involving perpetual replacement of lower level entities (parts) forming the higher level entities that persist and the classical mode of reproduction.

Mechanisms of reproductions in multicellular sexual organisms are often very complex and sometimes involve a germ-soma separation, but the simple examples above demonstrate that evolutionarily speaking reproduction, persistence and developmental growth should be regarded

as equivalent because differences between them are not necessarily ecologically relevant and more importantly because the concept of biological individual is ambiguous. Although Godfrey-Smith acknowledges a potential link between reproduction, persistence and growth, he defends the view that evolutionary significance of natural selection of entities that persist (without change) is very limited, because there is no *multiplication* to use Maynard-Smith's terminology (Maynard Smith 1988) of the substrate on which natural selection can act upon, nor even its *replacement* (Godfrey-Smith 2009b, 104-107). I fully agree on this point. Yet that does not mean however that the *process* of natural selection in and of itself requires multiplication or replacement. Replacement is a necessary conditions for *perpetual ENS* to appear (see Bourrat 2014 and Chapter 3, for more details as to why this is the case). By perpetual ENS, I mean here ENS that occurs indefinitely over time. Yet, the question as to whether *perpetual ENS* is possible in populations of entities that do not reproduce is a different one from the question as to whether the process of *natural selection* can occur at all in a given population, though the two questions are linked for ENS requires natural selection to occur. Similarly the question as to whether *perpetual ENS* is possible should be conceptually separated from the question as to whether *ENS* is possible at all. Multiplication, which entails replacement, is necessary for complex adaptation to appear in a population. But as I have already pointed out, natural selection and complex adaptations (as well as ENS) although related should be kept separate. Furthermore, it should be noted that reproduction is not the only way a population can multiply its substrate. In fact, developmental growth can also do the same thing.

It should also be noted that if the process of natural selection is possible without reproduction and that a process necessarily has an outcome, assuming that natural selection is the only difference maker in the population, ENS should be one possible outcome of a process

of natural selection without reproduction. What sort of ENS this leads to will be explored in the next chapter with the case of differential persistence. Okasha (2006, 214), acknowledges that differential persistence is a form of selection but quickly dismisses it as 'uninteresting'. Whether a process leads to interesting outcomes should not lead us to develop conceptual tools unable to account for those cases. Furthermore we will see in the next chapter that this form of selection can, under some circumstances, be interesting since reproduction and inheritance can evolve and, I will argue, be selected by this form of natural selection relying initially only on differential persistence. I will thus demonstrate the fundamental link between cases of ENS *qua* differential persistence and ENS *qua* differential reproduction.

2.6. Biased inheritance and natural selection: one or two processes?

Thus far, I have defended the view that reproduction is inessential for the process of natural selection to occur. If one takes all these considerations into account, the case for natural selection as being possible without reproduction is, in my view, quite strong. My arguments are largely based on the fact that reproduction, developmental growth and persistence are equivalent because there is no consensus on the definition of individuality and thus that the process of natural selection should occur whenever there is a difference in intrinsic-invariable properties between the entities of a population leading to a difference in any of those three outcomes. If one agrees with this view, it follows that inheritance and consequently heritability as classically understood (for more detail on the concept of heritability see Section 2.3 and Chapter 4) is also inessential to natural selection, since inheritance entails reproduction. I want nevertheless to make the point that, even in the case of a population of reproducing entities, *biased inheritance,* of which

I provide a definition below, should not be regarded as a requirement for natural selection to occur. I use here the notion of 'biased inheritance' rather than simply 'inheritance' because as long as individuals reproduce, they necessarily transmit something rather than nothing to their offspring and thus *some* inheritance is always entailed in an act of reproduction (I will come back to this question in detail in Chapter 3).

This vernacular notion of inheritance is thus not the one used by evolutionary biologists who prefer the concept of heritability. Under one interpretation, heritability, we saw, is a population-level measure of resemblance between parent and offspring.[62] But heritability can be linked to the concept of inheritance by the following conditions. To say that there is heritability on a trait in a population with discrete generations[63] is to say two things: Firstly, that there are different phenotypes in the population (variation); secondly, that parents produce offspring that on average resemble them more than they resemble other individuals of the parental generation. This second feature is what I call 'biased inheritance', where 'biased' is understood here as a bias towards the parental phenotype.[64] Thus, in my terminology perfect inheritance is a form of biased inheritance in which the offspring has the exact same phenotype as its parent(s). If one or both of these conditions are not met by a population, then there is no heritability in that population. Thus, assuming there is *variation* in phenotype in the population, if heritability is a requirement for the process of natural selection to occur, as claimed among others by Sober (1984), Endler (1986) and Haig (2012) (see Section 4), this is equivalent to the claim that a requirement for

[62] I refer here the regression approach to heritability.
[63] The reasons why the assumption of discrete generation is important will be treated in the next chapter.
[64] To see more clearly the difference between unbiased and biased inheritance with an example, see Section 3.3.

natural selection is biased inheritance when individuals reproduce (second condition) since variation, the first requirement for there to be heritability, is granted by assumption.

To test the validity of my claim that biased inheritance is not essential for natural selection to occur, let us take the case of stabilising selection proposed by Godfrey-Smith (2009b, 25-26) in a population made of three types of individuals: 'short', 'intermediate' and 'tall' (see Figure 2.7). In this population, short and tall individuals produce on average 0.5 offspring of their type (half produce one offspring, the other half do not reproduce), while the intermediate individuals produce on average 1.5 offspring (half of them produce one offspring while the other half produce two offspring). Two-thirds of the time the offspring is of the intermediate type, one sixth of the time it is tall and the remaining sixth of the time it is short. As a result, at the next generation (assuming discrete generations and a large population size), the population has the same frequency of each variant and thus no evolution is observed.

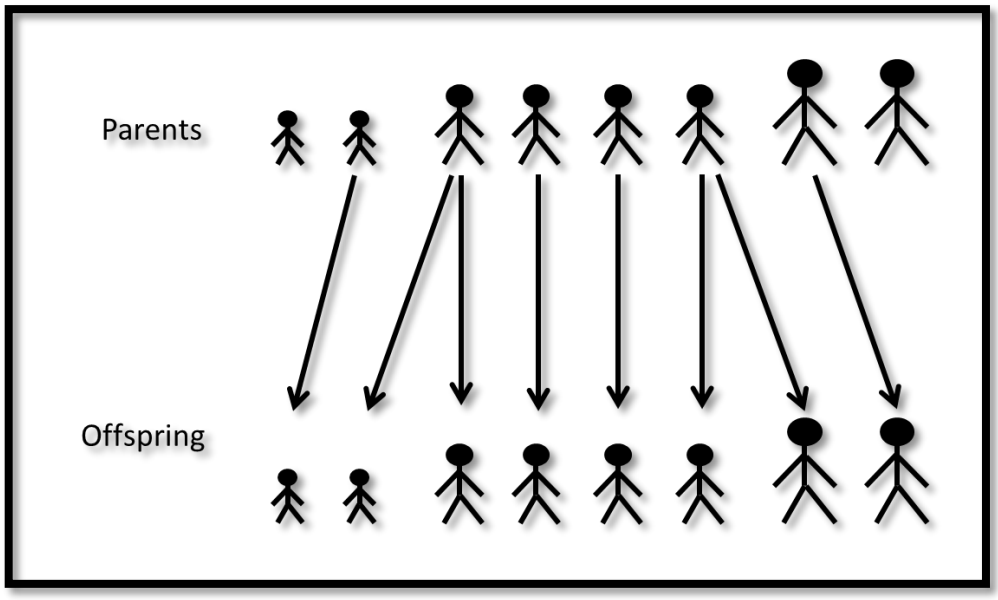


Figure 2.7. Illustration of a case of stabilising selection in an asexual population (inspired by Godfrey-Smith 2007, 2009).

Godfrey-Smith explains this phenomenon as the result of the pattern of inheritance and differences in fitness (that is, natural selection) 'pushing' in opposite directions. This is consistent with the interpretation of evolutionary change in the cases proposed by Godfrey-Smith and Lewontin (1993) and briefly discussed in Subsection 2.3.2 of this chapter. I agree with this way of representing the phenomenon for it naturally falls under the view in population genetics that there are different forces or causes of evolutionary change (which can be nil) – classically natural selection, drift, mutation and migration. If one defines natural selection as a process occurring when there is a difference in fitness between the types (or organisms) of a population, as is done in population genetics using the selection coefficient '*S*' (Crow & Kimura 1970, 179), then natural selection and the pattern of inheritance can be seen as two different causes of evolutionary change in Godfrey-Smith's example. Under this view "selection causes observable changes within a generation […] without recourse to the inheritance of characters. In contrast, evolutionary

response to selection, for example, the change in phenotypic mean from one generation to the next, certainly does depend on inheritance" (Arnold & Wade 1984, 709). Note that this view originates directly from the Breeder's equation, which is a classical equation of quantitative genetics and measures what is called 'the response to selection' (R) with two terms: one interpreted as the *selection differential* (S) and another interpreted as heritability (h^2) (Falconer 1981) as follows:

$$R = Sh^2 \qquad (2.1)$$

Following the Breeder's equation (equation 2.1), the response to selection is only the evolutionary product of a combination of the process of selection, measured through the differential selection, and heritability. Now, as I argued in subsections 2.3.2 and 2.3.3 of this chapter, with no knowledge of the underlying genetic determination of the pattern of inheritance in the population, as supposed in Godfrey-Smith's example, imperfect transmission of characters between generations can be formally seen as a (non-standard) form of mutation across generation or at very least as a distinct evolutionary process. When the genetic determination is known (if any), the term used when the genes of a population are shuffled between each generation is *recombination*. Recombination can be non-independent, that is, genes for different phenotypes can be found preferentially with other genes for different traits. In this case, recombination is a non-negligible evolutionary force (individual events of recombination make a difference on total evolutionary change). Recombination can also be independent, in which case it leads to no evolutionary change. Whether or not we know the genetic determinism, it is uncontroversial in evolutionary biology that mutation or recombination and natural selection are distinct evolutionary processes and that one cannot be seen as a sub-process of the other. Thus formally,

Godfrey-Smith's example is a mixed case of natural selection and mutation (or recombination) pushing in opposite directions or at any rate, if one refuses to consider imperfect inheritance as a form of mutation, a mixed case of natural selection and another evolutionary force.

In the above example, it is thus clear that natural selection should be regarded as independent from the concept of (biased)-inheritance. Natural selection, in this example, should thus be understood as the process by which individuals because of differences in phenotypes, end up producing a different number of offspring, irrespective of their offspring's phenotypes. If the pattern of inheritance leads to the production of offspring with a phenotype different from their parent, this does not mean that there was no natural selection going on. As I mentioned earlier, Godfrey-Smith's example is classically referred to as a case of stabilising selection in which the intermediate phenotype rather than one of the extremes is selected. But the fact that Godfrey-Smith used a case of stabilising selection rather than directional selection (that is, when one of the two extreme phenotypes is selected) should not mislead us into thinking that the lack of evolutionary change can only occur in the case of stabilising selection: the same phenomenon can occur in a case of directional selection as well.

To convince ourselves, let us imagine a case similar to the one proposed by Godfrey-Smith, in which tall individuals are fitter[65] (see Figure 2.8). Suppose they have on average 1.5 offspring (half of the tall individuals have one offspring while the others have two offspring). Among the offspring they produce, two-thirds are tall, while one-sixth are small and another sixth are intermediate. Small and intermediate individuals produce on average 0.5 offspring (one-half of the individuals produces no offspring while the other half produces one offspring) with the

[65] Other cases can be found in Godfrey-Smith (2007; 2009b, 169-172).

phenotype of their parent. In this case, as in Godfrey-Smith's case, there is no evolutionary change although this is a case of directional selection (we could also imagine cases of disruptive selection). Besides a difference in fitness between the individuals of the population, the only other difference between Godfrey-Smith's case and the present one is a difference in patterns of inheritance.

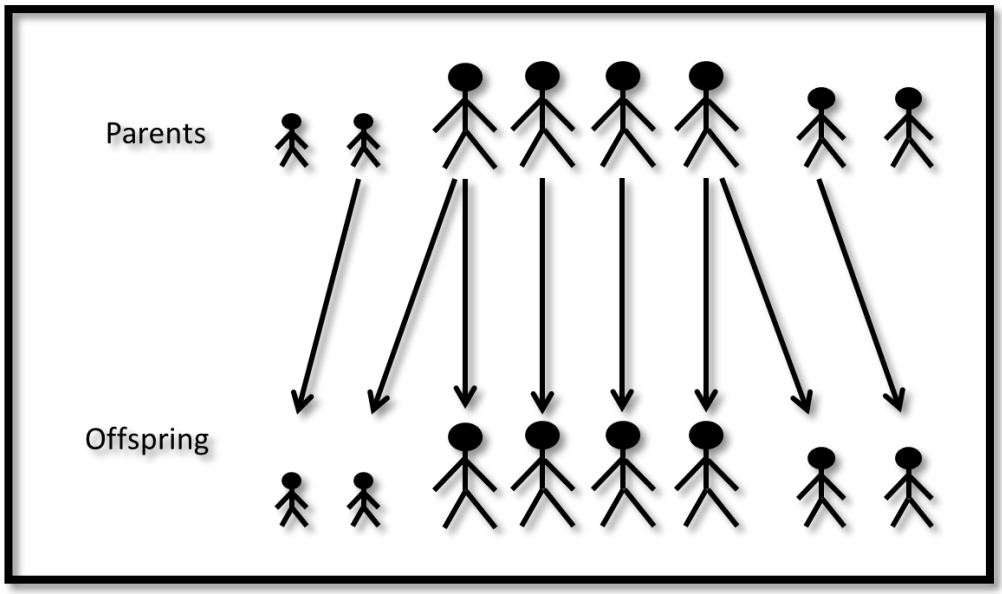


Figure 2.8. Illustration of a case of directional selection without evolution in an asexual population without evolution.

One can imagine an infinite number of cases, similar to the two above, in which natural selection and the pattern of inheritance in the population push in opposite directions, leading to no evolutionary change. In a similar way, one can conceive of an infinite number of cases in which the only difference between those cases is a difference in pattern of inheritance (no differences in patterns of fitness) that leads to different evolutionary changes. Compare for

example the two following cases (Figure 2.9*a* and *b*), once again inspired by Godfrey-Smith's example, in which short, intermediate and tall individuals produce consistently respectively one, two and three offspring but in one case (Figure 2.9*a*) inheritance of the trait is perfect for all types, while in the second (Figure 2.9*b*), there is a chance, increasing with the height of the parent, that the offspring produced will be short (in Figure 2.9*b*, the chance is 0.5 for intermediate individuals and 2/3 for tall individuals). In both cases, although the pattern of natural selection is the same (that is, same fitness differences), the evolutionary change resulting are vastly different. In the first case the average population size will tend toward the phenotype tall, going down from one-third of short individuals at the first generation to one-sixth at the second one (Figure 2.9*a*). In the second case, depending on the strength of the transmission bias, the population might tend toward the phenotype 'short' if the transmission bias toward short offspring for intermediate and tall individuals is strong enough. This is what is observed in the case presented in Figure 2.9*b*, where at the second generation there are two-thirds small individuals compared to one-third at the first generation.

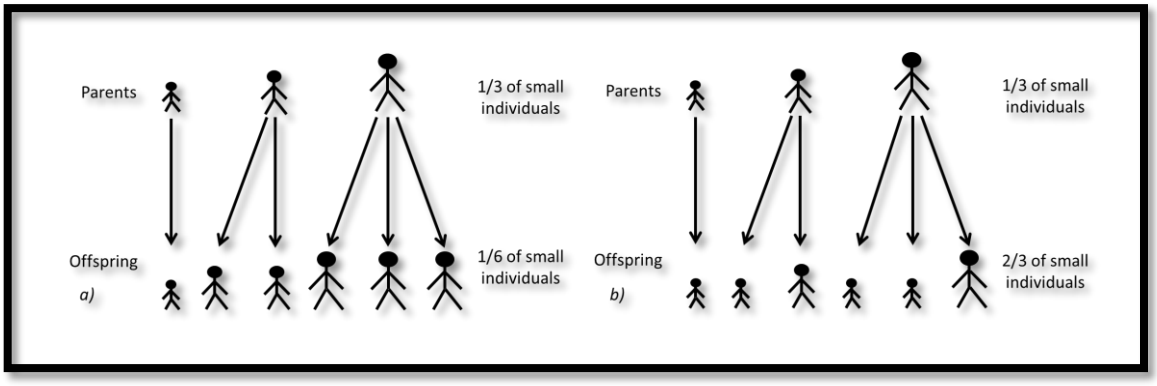


Figure 2.9. Illustration of two cases with the same pattern of selection for height but with different patterns of inheritance.
(Inheritance is perfect in a while it is not in b. This leads to different evolutionary changes after one generation.)

These different examples demonstrate that, unless inheritance is perfect, no predictions of intergenerational evolutionary change can be made from differences in fitness alone. This leads me to conclude that the process of natural selection should be defined independently from any reference to inheritance, since it appears conceptually distinct from it. Reference to inheritance in the cases described in this section is intergenerational evolutionary change due in part to natural selection (a form of ENS). But as I have emphasised in the previous section the question as to whether natural selection is occurring is different from the questions as to whether ENS will result and if yes, what sort of ENS (e.g., perpetual, intragenerational, intergenerational, etc.)

2.7. Conclusion of the chapter

I started this chapter by showing that the classical approach to ENS was insufficient to grasp what the process of natural selection in and of itself is. To achieve this, I have proposed a complementary approach that considers an idealised population in which there is nothing but natural selection. This has helped me clarifying the links between imperfect inheritance and natural selection. I have shown that when there is imperfect transmission and difference in fitness, at least two evolutionary processes are at play in the population, namely natural selection and another one I have suggested should be attributed to mutation (in a formal sense).

I have then argued that reproduction and inheritance should not be included in the process of natural selection but rather are, together with natural selection responsible for different forms of ENS. Yet some, such as Godfrey-Smith (2009b, 103-107) for example, consider cases

of persistence as 'marginal' cases of ENS, when, as I already pointed out, Okasha considers them 'not very interesting' because they will not lead to adaptations (Okasha 2006, 214). Whether uninteresting or marginal forms of ENS, those more basic phenomena should nevertheless be readily explained by our conceptual apparatus on natural selection. Van Valen's and Bouchard's examples, but also many other 'oddities' for the received view on evolutionary theory, should find easy interpretations. One startling fact is that they currently do not when classical interpretations are applied to them. Furthermore, in the next chapter, I will challenge the claim made by Okasha that ENS cases without reproduction lead necessarily to uninteresting cases of ENS. My conclusion is different from his. ENS without reproduction can be interesting! In fact, I will demonstrate that reproduction and inheritance might themselves be complex adaptations of a process of natural selection in which initially there is no reproduction using one of Okasha's own strategies which he calls the strategy of 'endogenisation'.

3. From persistors to replicators: Endogenising Reproduction and Inheritance

3.1. Introduction

In the previous chapter, I have argued that the process of natural selection, and consequently ENS, does not require reproduction and (biased) inheritance to occur. One minor aim of the current chapter is to give more weight to this argument, by showing *in silico* that ENS, when understood more broadly, can occur in populations of entities that do not reproduce. More particularly, I will show that ENS in non-reproducing and reproducing populations fundamentally results from one and the same process. A second, more ambitious aim of the chapter is to provide an evolutionary explanation of the pervasiveness of reproduction and inheritance in biological organisms and show that, despite not being essential components of ENS, they do have a special role in Darwinian processes. In fact, I show that they can be seen as primordial adaptations of non-reproductive Darwinian populations and that they represent a major channel by which complex adaptations can appear.[66] The take home message of the chapter will this be that cases ENS in which there is reproduction and high fidelity inheritance *can evolve* via a weaker form of ENS, from populations in which there is neither inheritance nor reproduction. I will call this phenomenon, 'endogenisation-of-inheritance'

[66] Another channel by which complex adaptation can appear is differential growth of parts as suggested by (Bouchard 2008, 2011).

The remainder of the chapter is divided into four sections. In Section 3.2 I start from the claim presented in the conclusion of the previous chapter, that ENS without reproduction is a genuine form of ENS. I call this less demanding process *weak* ENS. The fact that weak ENS does not require reproduction to occur represents a serious problem for the classical summaries of ENS presented in the previous chapter. However, I concur with Okasha (2006) that this form of ENS is evolutionarily uninteresting *in the long run,* since populations of non-reproducing entities cannot lead to adaptations as classically understood by evolutionary biologists. At this point there will be no remaining doubt that the summaries of ENS encountered in the previous chapters describe one particular form of ENS, namely, the form that allows for cumulative evolution and complex adaptations. Following Godfrey-Smith's terminology, I will call this *paradigmatic* ENS. With this distinction in mind, I then ask whether paradigmatic ENS can originate from weak ENS, or whether the two forms of ENS are due to two fundamentally different processes of natural selection.

In sections 3.3 and 3.4, I answer this question by showing, using individual-based models that paradigmatic ENS not only *can* but *is likely to* originate from weak ENS. These two sections present a diachronic account of ENS in which the third ingredient, inheritance (entailing reproduction), is not taken for granted, but rather is 'endogenised' into a broader account of ENS. I borrow the concept of endogenisation from Okasha's (2006) discussion on the levels of selection, in which he explains that multicellular organisms, once taken for granted by individual level selection theory, have been endogenised by the multilevel selection theory by viewing multicellularity as simply another product of natural selection**.** The origin of the term itself comes from the causal modelling literature. In Chapter 1, when I presented causal graphs, I distinguished two sorts of variables: namely, *exogenous* ones that are given in the causal model, and *endogenous*

ones that are explained by other variables in the model. With this in mind, the strategy of endogenising can be understood has explaining variables which have previously been taken for granted in a model (such as reproduction and inheritance), by reference to other, more fundamental variables present in the model.

Endogenising paradigmatic ENS, and more particularly inheritance, will necessitate four cumulative steps:

a) New variation must be introduced in the population over time (mutation). Without new variation, no cumulative evolution at all is possible. This will be postulated (or exogenised) in all my models.

b) At least some members of this population must be able to reproduce (that is, there must be *multiplication*, or at the very least, *replacement* of entities). If not, the population will become extinct.

c) Given that the population is able to maintain its size, *all* the entities of the population must be able to reproduce.

d) Differences in fitness must be transmitted from parents to offspring (memory).

I show how steps *a* and *b* can be realised in Section 3.3 using minimal assumptions, and how steps *c* and *d* can be realised in Section 3.4. At the end of Section 3.4 it will become clear that inheritance is actually a complex adaptation of Darwinian populations in which, originally, only weak ENS is possible. Most of my reasoning will be based on simulations of individual-based models using the software NETLOGO 5.02 (Wilensky 1999). Using simulations will have two virtues. First, it will help the reader to see the different concepts and processes I explore 'in

action'. Second, it will allow me to develop models that, although simple, are too complex to be expressed verbally in an efficient way.

In Section 3.5, I briefly assess the value of the endogenisation-of-inheritance hypothesis in three domains of relevance for the concept of ENS: origins of life, evolutionary transitions in individuality, and cultural evolution.

3.2. ENS without reproduction

As we have seen in Chapter 2, many definitions of natural selection postulate reproduction and inheritance (or more precisely, generational heritability) as being necessary conditions for a process of natural selection to occur. However, I have shown that these factors are not necessary, if one distinguishes natural selection from its products; namely, ENS and complex adaptations. In this section I will revisit this argument, but instead of focusing on the *process* of natural selection, I will rather focus on its product, namely, ENS.

In the previous chapter, I presented a quote of Van Valen, in which he challenged the view that differential reproduction and differential growth represent two fundamentally distinct processes from an evolutionary perspective. Here is another more extremist quote from him, where he challenges the view that ENS requires differential reproduction (or even growth at all):

> "[Granite] is composed mostly of grains of feldspars and quartz, with some mica and other minerals inserted among them. When granite weathers, the feldspars and micas become clays but nothing much happens to the quartz grains. They are most resistant and get transported down streams or along shores. Thus most beaches are the result of differentially eroded granite. This is an example of natural selection in the nonliving world. Quartz grains survive longer than feldspar grains, and there is a

> progressive increase in the average resistance to weathering, of the set of grains that have still survived […] The lack of reproduction imposes constraints on the flexibility of evolution here, but one shouldn't confuse that with the selection itself." (Van Valen 1989, 2)

Although grains of quartz, feldspars and mica do not reproduce, there is a form of selection on grains of minerals which leads to an overall change in the population. To use the terminology introduced earlier, 'hardness' is an intrinsic-invariable property of a grain of mineral, and it is different between the different types of grains. This difference in hardness causes a corresponding difference in persistence between types. Thus, after some time, we observe a change in average hardness of the population of remaining grains, as the proportion of quartz progressively increases.

From now on, let us call the entities that are unable to reproduce and only undergo this type of selection, 'persistors'.[67] The fact that a form of ENS can be observed in populations of persistors represents a serious problem for the classical account of ENS. As I argued earlier, the classical recipes supposedly capture the essence of ENS; yet it seems they cannot recognise what is arguably the simplest example of this process; namely, viability selection in populations of persistors. Why, then, is the classical account so widely endorsed?

I noted in the previous chapter that one main reason is that, for some authors, the concept of natural selection is *interesting*, and can lead to *adaptations*, only when it is applied to entities that reproduce (e.g., Okasha 2006, 214). Okasha uses the term *weak selection* when referring to processes of selection between different types of persistors. If selection in populations of persistors cannot lead to adaptations, it might be justified to distinguish cases of differential

[67] In Bourrat (2014) I called them 'survivors', but the term 'persistor' seems less ambiguous than survivor.

persistence (*weak* ENS,) from cases of classical or *paradigmatic* ENS. Although the process of weak selection seems to be fundamentally the same process as that which operates in paradigmatic cases of selection, the product might be different in a substantial way. But one thing is sure, whether the process is interesting or not cannot alone count as a substantial reason to consider it as of different nature.

Godfrey-Smith (2009, 40) provides another answer. Although, like Okasha, he recognises that an extension of the Darwinian apparatus to populations of persistors is possible, it is in his view, a very artificial one. By 'artificial', I take Godfrey-Smith to mean that historically, the concept of natural selection has only been applied to (living) entities that do reproduce. Applying these ideas to entities that do not reproduce is, in this sense, artificial because the theory was not elaborated with non-reproductive entities in mind. Although it might be historically artificial to extend the concept of ENS to populations of entities that do not reproduce, such artificiality, like interest, cannot be a strong argument against a foundational description of Darwinian processes; or against what Griesemer calls 'generalization by abstraction' of Darwinian evolutionary theory (Griesemer 2005). As I have already pointed out, if recipes for ENS are conceived as representing foundational descriptions, they should be able to account for the simplest cases. To be fair to Godfrey-Smith, it should be reemphasised that his approach to assessing the concept of ENS is to apply a prototype and similarity strategy to the notions of paradigmatic and marginal ENS, whilst my strategy is to look for complete generality; that is, to be able to distinguish between when ENS is occurring, and when it is not, using 'if/then' criteria.

Some people even refuse to grant differential persistence or viability of types the status of evolution, let alone ENS. I am inclined to grant it the status of ENS because changes in frequency of types or variants due to the sole action of selection are usually identified by

population geneticists as evolution; and as I discussed in the Introduction of the thesis, such distributive evolution is a reasonable sense of evolution. Essentially, the problem is a terminological one, and not what I am ultimately interested in here. The important question is whether weak ENS (or whatever one wants to call it) and paradigmatic ENS are fundamentally distinct.

As we saw, Okasha considers weak ENS to be uninteresting, but this is because, according to him, weak ENS cannot lead to adaptations. The ability to lead to adaptation seems to be the kind of fundamental difference between the two forms of ENS we are looking for. But is this proposition correct? Is it correct to claim that there are no adaptations under a process of weak ENS? This of course depends on what Okasha means by adaptation. Brandon (1990, 40) proposes distinguishing *ahistorical adaptations,*[68] which simply refer to properties that increase the fitness of their possessors, from *adaptation as products of the process of ENS*, produced once several steps of mutation and selection have occurred in a population (what I have so far referred to as *complex* adaptations). I take it that what Brandon means by a 'product of the process of ENS' is something similar to what Godfrey-Smith calls the 'positive role of natural selection' (Godfrey-Smith 2009, 49-50). In his discussion on origin and distribution explanations that I mentioned in the Introduction, Godfrey-Smith proposes that in spite of its 'negative' role (that is, eliminating variation over time) natural selection has a positive, creative role in evolution. It does so by "changing the population-level background against which new mutations can occur" (Godfrey-Smith 2009, 50; see also Godfrey-Smith's quote at the beginning of Section 2.4).

[68] A synonym for ahistorical adaptation is 'adaptiveness'.

The following is a schematic way to represent the process behind changes in population-level background against which new mutations can occur. In a population of entities that reproduce and in which there is variation with regard to one trait, selection (let us assume it is directional) will eliminate most of the variants, keeping only the fittest ones. These variants will then reproduce and transmit their phenotype to their offspring, so that the successful phenotypes become relatively more represented in the overall population. New advantageous mutations further increasing the fitness of their bearers will thus become more probable. Godfrey-Smith argues that this is the kind of phenomenon that leads to complex adaptations such as the human eye or brain. This is, I take it, the kind of adaptation that Okasha claims is not possible in populations of persistors. Ahistorical adaptations are, however, perfectly possible in these populations.

In some respects Okasha is right. Complex adaptations cannot appear in population of persistors, but not because of some fundamental distinction between a population of persistors and a population of reproducing entities. The reason why they cannot appear is because the probability of successive 'lucky' mutations is so low in those populations that they become nearly impossible. This is because there is no multiplication with memory in those populations. Going back to Van Valen's quote, a low probability of successive advantageous mutations is the constraint on the flexibility of evolution, he is talking about, that the lack of reproduction imposes on populations. Yet, in principle, with luck or with an enormous population size (assuming that some mutations occur), genuine complex adaptations for some entities could appear in a population of persistors. It should also be noted that the notion of 'complex adaptation' is also a continuum. Thus, the concept of adaptation cannot qualitatively separate the two forms of ENS. Is there any other way to separate them?

In the next two sections I argue that weak and paradigmatic ENS should in fact not be separated, because paradigmatic ENS is a special case of ENS that not only can, but is likely to originate in populations of persistors with the ability to mutate. The reason for this is that inheritance and reproduction can themselves *be* complex adaptations in those populations. In fact, moving from a population of persistors to a population exhibiting paradigmatic ENS will require four cumulative steps or conditions mentioned in the introduction of this chapter. The first step is *a*) that new variation is introduced in the population (mutation). Without it no cumulative evolution is possible. The second step is *b*) that the population is able to maintain its size, or if the size of the population is not limited, to increase (multiplication). This is not the cases in weak ENS, in which populations are ultimately condemned to extinction. But a population being able to maintain its size does not necessarily make it a population in which paradigmatic ENS occurs. Two other important phenomena occur in paradigmatic cases of ENS: *c*) advantageous phenotypes are able to be transmitted during reproduction, making population-level changes against which new mutations can occur possible; and finally *d*) reproduction is pervasive, therefore each entity in the population is in principle able to reproduce. Invasion of one variant in the population of reproducing entities under the sole action of natural selection is the hallmark of a process of complex adaptation. Because I will come back to these four conditions, I have regrouped them in the Table 3.1.

Table 3.1. The four necessary steps for moving from a case of weak ENS to a case of paradigmatic ENS.

Condition	Description
a	New variation is introduced in the population over time (mutation).
b	The population is able to maintain its size or, if the size of the population is not limited, to increase (multiplication).
c	Advantageous phenotypes are able to be transmitted during reproduction, making population-level changes against which new mutations can occur, possible.
d	Reproduction is pervasive, therefore each entity in the population is in principle able to reproduce

In the next section I build three individual-based models that will show how one can move from a population of persistors to a population of entities able to maintain its size and thus meet conditions *a* and *b*. I will distinguish the concept of *procreation*, which is reproduction without inheritance, from the concept of *minimal reproduction*, which is indefinite procreation; that is, the ability to procreate being transmitted from parents to offspring. I will then show that minimal reproduction, but not procreation, once introduced in a population of persistors through mutation (condition *a*), allows this population to maintain its size (condition *b*). Yet these conditions will be insufficient for paradigmatic ENS to occur. At that point I will use the distinction between biased and unbiased inheritance proposed in the previous chapter. In Section 3.4, using other models, I show how conditions *c* and *d* can be obtained from a population of entities able to maintain its size, by introducing biased inheritance of the ability to procreate (that is, reproduction) and biased inheritance of other Darwinian traits; and why biased inheritance represents a genuine adaptation of those populations. Towards the end of Section 4, I propose that the existence of replicators (Dawkins 1982; Hull 1980) can result from a similar process of adaptation and thus can be conceived as a product of natural selection rather than condition for it.

3.3. Reproduction and inheritance in ENS

How can we get a population which maintains its size from a population of persistors? It seems that the ability to reproduce is the right property to give to our persistors. But reproduction is a complex phenomenon that can be decomposed into (at least) two more fundamental phenomena: the first, which I name *procreation,* is reproduction without any property transmitted from parents to offspring, including the ability to procreate; the second is inheritance of the ability to procreate. Note that I distinguish procreation from 'progeneration' (Griesemer 2000) which is, in my terminology, procreation with material overlap (see more on material overlap in Section 6.3.2). A procreated entity of any given reference class inherits nothing from its parents but the facts of coming into existence and membership of that class. It might be difficult to imagine an act of procreation which would not involve the transmission of some properties other than the fact of coming into existence, and the choice of a reference class is a well-known problem in statistics (Hájek 2007, see also Chapter 5).

One way to choose the class in the particular case of ENS is to consider that the properties of an object defining the class should be properties that do not vary in the population. Rotten meat could be said to 'procreate' maggots if one had in mind a theory of spontaneous generation, with the reference class 'entities made of flesh'. This is because being made of flesh does not vary whether one is referring to meat or maggots. Although spontaneous generation theories are absent from modern biology, I gave this example to make clear that procreation in a given class of objects does not necessarily involve inheritance or transmission of traits, and that

it is therefore the right concept we need for our purposes, since we want to dissociate inheritance from reproduction.

With the distinction between procreation and reproduction in mind, we can now start to imagine the fate of a population of persistors in which there is variation in viability and some of these persistors spontaneously mutate into *procreators*: entities able to procreate. Does this mutation allow our initial population of persistors to maintain its size?

To answer this question, I built a set of individual-based models using the software NETLOGO 5.02. In their simplest form, individual-based models simulate the behaviour of individuals that are 'asked' every unit of time to behave in a certain way (e.g., to move, to reproduce, to die etc.). Two advantages of individual-based models, which represent a microscopic approach to modelling, over macroscopic approaches are their flexibility (which makes them appealing for testing new hypotheses), and that they allow for each individual in the simulation to be unique. These two features make individual-based modelling ideal for the purpose of this chapter. For more on individual-based models see Grimm & Railsback (2005) and Railsback & Grimm (2011). My models describe a population initially composed of persistors. Each persistor has only one property, its survival rate over each unit of time, which remains constant during its whole life and is fixed at birth. The survival rate of each persistor is chosen randomly between 0 and 0.99 (to avoid immortal persistors). Time is discrete. For every unit of time the population undergoes one selection event.

During the event of selection, each individual is selected according to its survival rate – also called viability – which is taken to be an intrinsic-invariable, as are all the properties of the entities in this chapter, unless otherwise mentioned, as in cases of mutation. Selection operates

as follows: if a given individual has a survival rate of 0.6 (recall that this value is chosen randomly at birth), this means that its chances of surviving this round of selection will be 60%. The lower the survival rate, the lower the chances of the individual surviving until the next unit of time.

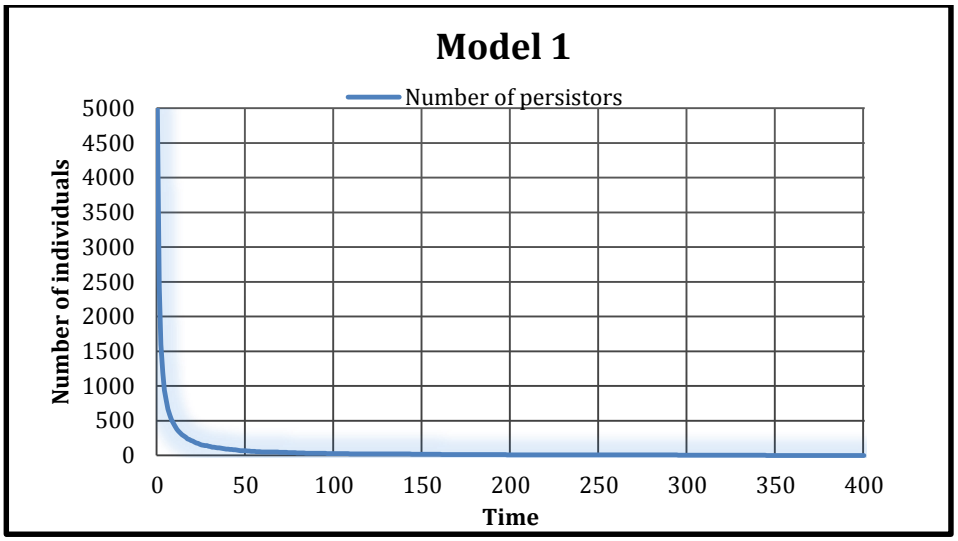


Figure 3.1. Simulation of Model 1 over 400 units of time.

Model 1 is essentially similar to ENS without reproduction, as proposed in Section 3.2 (weak ENS). There are only persistors in the population and no mutants. Figure 3.1 represents the graph of one typical simulation of Model 1 in which the number of persistors is recorded over time. At t=0, 5000 persistors are created. As expected the population rapidly goes extinct. Model 1 represents a null model in my framework, upon which further models are built. Table 3.2 summarises the different models and associated concepts used in Section 3.3 and 3.4.

What happens now to this population if procreation is introduced by mutation of one persistor into a procreator? A second model (Model 2) was created to answer this question. Model

2 is very similar to Model 1 but procreation is introduced in the population by including one mutant individual at the beginning of the simulation (see Table 3.2). When a persistor mutates into a procreator it acquires the possibility to procreate.[69] Condition *a* (that new variation is introduced into the population, see Table 3.1) is thus introduced into the model. Hence, procreators have two properties: a survival rate within each unit of time (as do persistors) and a fertility rate (that is, number of individuals procreated within each unit of time as long as the procreator is alive). The fertility rate of a procreator remains constant throughout its whole life unless further mutation occurs.

Before going on, I need to make clearer and use the distinction between unbiased and biased inheritance introduced in Section 2.6. Let me remind the reader that when there is unbiased inheritance of a trait, the only thing a parent transmits to their offspring is the trait itself. The offspring's phenotype associated with this trait is uncorrelated to the parent's phenotype and thus can take any possible value on that trait. When there is biased inheritance of a given trait, there is a correlation between the parent's phenotype and the offspring's phenotype on that trait. Thus, in a case of unbiased inheritance, no prediction of the offspring's phenotype from the parent's phenotype is possible while *some* level of prediction is possible in a case of biased inheritance.

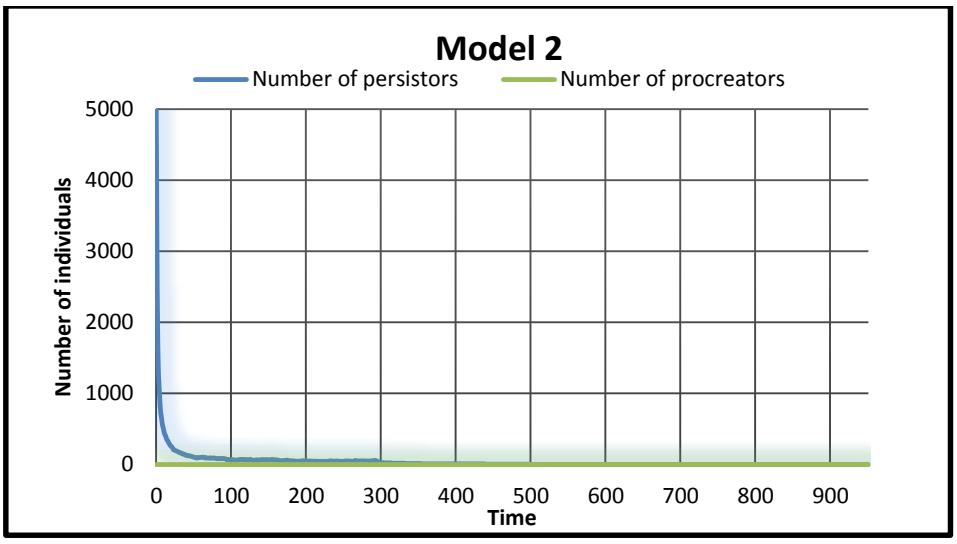
To make this clear imagine that an asexual entity of size *S* produces nine offspring, and that the scope of possibility on size of a given entity of the class of reference is between 1 and 9. If there is unbiased inheritance of *S*, the size of the offspring should be uncorrelated (whether linearly or otherwise) with the size of the parent (and thus unpredictable), and should thus be on

[69] In Model 2 we assume that the mutation has already occurred. Mutation is thus exogenous to the model.

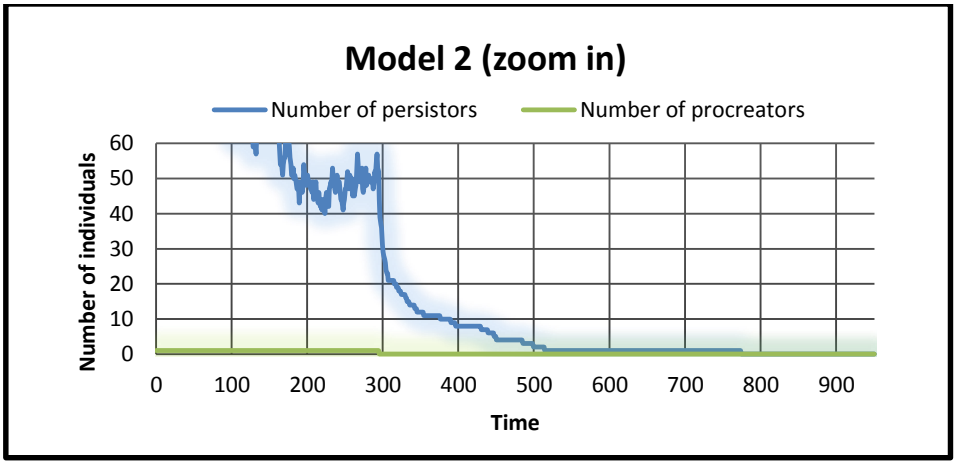
average 5, no matter what the size of the parent (any size between one and ten) and what the distribution. If there is biased inheritance of *S*, the offspring will be more likely on average to have size *S* than any other size between 1 and 10. Note that if the parent has a size *S* of 5, the average size of offspring will also be 5, as in the case of unbiased inheritance, but a correlation between parent and offspring should be observed.

In Model 2, as in Model 1, every individual undergoes selection. But after selection procreators are able to procreate.[70] During this stage, each procreator, if any exist, procreates according to their fertility rate. We suppose that there is no (biased) inheritance of viability so that the offspring's viability is chosen randomly between the minimal and maximal values allowed by the model (0 and 0.99). Because there is no inheritance of the ability to procreate, offspring of procreators are persistors. After the stage of procreation, another stage is introduced in the model. I call it the check-for-overcrowding stage. During this stage, the size of the population is assessed and if it is higher than its initial size, individuals are randomly chosen and killed until the population roughly reaches its initial size. The sequence of events undergone by every individual in Model 2 is *selection* → *reproduction* → *check-for-overcrowding*; this restarts every unit of time.

[70] In Model 2 and subsequent ones, procreation is always asexual.



a.



b. (zoom in)
Figure 3.2. Simulation of Model 2 over 900 units of time.

Figure 3.2 and 3.3 both represent the graph of one (and the same) typical simulation of Model 2 in which the number of persistors and procreators is recorded over time. At t=0, 4999 persistors and 1 procreator (which has mutated from a persistor) are created. For the purpose of my argument, the viability and fertility of the mutant procreators are the highest possible allowed by the model (respectively 0.99 and 10) in order to maximise its reproductive output and see

whether any change at all is possible when compared to a population of persistors only. As can be seen in Figure 3.2*b*, the population becomes extinct after 700 units of time while the procreator dies before the 300th unit of time. This is not a surprising result. Because procreators' offspring are persistors, the population is only able to produce new entities as long as the initial procreator is alive. Once this initial procreator dies, the population becomes similar to the population of persistors in Model 1. No matter how many procreators there are in the initial population, this population is ultimately condemned to extinction.

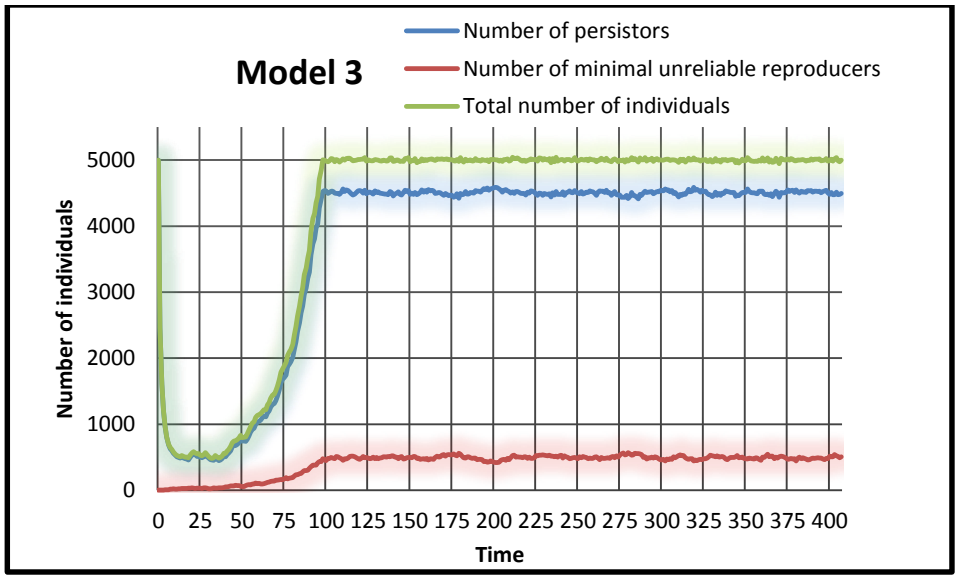
Thus, a population in which procreation has emerged is unable to maintain its size and the type of ENS observed in this population is very similar to weak ENS. This result is good evidence that inheritance *must* be part of the reason why a population is able to display paradigmatic ENS (and more proximally, to satisfy *b*) simply because, without inheritance of the ability to procreate, a population will become extinct. Let's now see what happens in our population if, instead of a procreator, the mutant introduced is a *minimal reproducer*. I define a minimal reproducer as a procreator which is able to transmit its ability to procreate to some, if not all, of its offspring. I call this entity a *minimal* reproducer because its ability to procreate is typically not transmitted to every one of its offspring.

In Model 3 (see Table 3.2), I start as in Model 2 with a population of 4999 persistors and one mutant (in this case a minimal reproducer). Minimal reproducers behave in a similar way as procreators except that some of their offspring are also minimal reproducers. Thus, the only difference between Model 2 and Model 3 is that in Model 3, the rules of transmission of fertility and of the ability to procreate between parent and offspring must be chosen (see Table 3.2). The rule of transmission for viability remains the same as in model 2. The sequence of events undergone by every individual each unit of time in this model is the same as in Model 2.

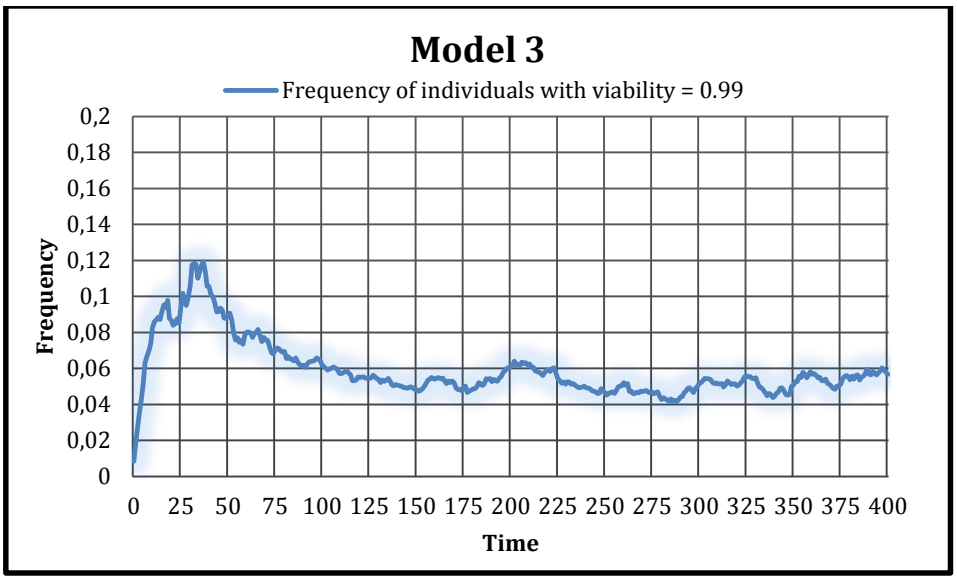
Because I am interested, for the moment, only in the effect of inheritance of the ability to procreate on the evolutionary dynamic of a population, the rule of transmission for fertility chosen for Model 3 is that each minimal reproducer produced receives a fertility rate chosen randomly, as with viability in Model 2 (and 3); that is, a fertility between the minimal (0) and the maximal (10) fertility rate set by the model. Thus there is unbiased inheritance of fertility rate. The rule of transmission for the ability to procreate is more difficult to choose. In fact, by introducing minimal reproducers in the model, the notion of unbiased or biased inheritance on the trait 'ability to procreate' seems inappropriate. This is because the ability to procreate is an evolutionary innovation that 'creates' the possibility of inheritance in a population. However, once some individuals have the ability to procreate and are able to transmit it to some extent (perhaps at a very low level, say one offspring for each five offspring produced), there can be either unbiased or biased inheritance of this ability to the next generation – this will be based on the proportion of offspring of a minimal reproducer that will be able to pass the ability to procreate to their own offspring. When this proportion is unpredictable within the reference class, there is unbiased inheritance. I call minimal reproducers with this ability, *unreliable* minimal reproducers. When there is biased inheritance of the ability to procreate (which entails the inheritance of the ability to procreate), not only can we predict the proportion of offspring that will receive this ability but we can also predict to some extent the level at which they will receive the same ability than their parent. This probability reflects the reliability of the channel transmission between parent and offspring on the ability to procreate. I call minimal reproducers with biased inheritance of the ability to procreate: *reliable* minimal reproducers.

In Model 3 unbiased inheritance of the ability to procreate is introduced in order to see whether this simpler form of inheritance is sufficient to make it such that our population is able

to maintain its size. I set inheritance of the ability to procreate at a maximal level of 0.20 so that a minimal reproducer cannot have more than 20% of its offspring as minimal reproducers. The level at which a minimal reproducer's offspring receives this ability is chosen randomly between 0 and 0.2. I have set the maximal level of inheritance very low in order see whether even low inheritance of the ability to procreate is sufficient for a population to maintain its size, and thus satisfy *b*.



a.



b.
Figure 3.3. Simulation of Model 3 over 400 units of time.

Figure 3.3*a* represents the graph of a typical simulation of Model 3 in which the numbers of persistors, minimal reproducers, and total individuals is recorded over time. At t=0, 4999 persistors and 1 minimal reproducer, which has mutated from a persistor, are created. As in Model 2, for the purpose of my argument in this simulation the survival rate, fertility rate, and transmission rate of the ability to procreate of the mutant minimal reproducer are the maximum possible allowed by the model (respectively 0.99, 10 and 0.20), in order to maximise the chance of observing a difference between this model and the previous ones. The graph shows that during the first 15 units of time (in this particular simulation) the population behaves in similar way as in the two previous models: the population size drastically decreases. However, in contrast to the previous models, the population does not go extinct. After some time, the population size starts to increase until it reaches the maximal population size, at around the 100th unit of time. After that point, the size of the population remains stable and is composed of around 10% of minimal reproducers.

We can conclude from Model 3 that even a low level of unbiased inheritance of the ability to procreate is sufficient for a population to maintain its size. It thus satisfies *b*, one of the two remaining requirements, aside from the requirement of production of variation, for a population to be able to exhibit paradigmatic ENS. We have not yet established whether it satisfies *c*, the requirement that there be change in population level background against which new mutations can occur; and *d*, the requirement that every individual of the population is able to reproduce, is not met.

One way to see whether *c* is satisfied is to check whether individuals with the highest fitness are able to invade the population. If they do, this will represent a change in the population-level background against which new mutations can occur. But measuring the fitness of entities in our population is problematic because we have two types of entities, persistors and unreliable minimal reproducers, with very different behaviours. In fact, we cannot use the classical proxy of fitness as the product of viability and fertility for persistors, because their fertility is nil. We are in exactly in the case depicted in Figure 2.4 in which comparing the fitness of the types **R** and **G** was problematic because they have two very different life cycles.

Yet, in Model 3, we can be sure that the fittest persistors are those with a survival rate of 0.99. This is because we suppose that viability is an intrinsic-invariable property and that all the persistors are in the exact same environment. Similarly, we can be sure that the fittest minimal reproducers are those with the highest viability (0.99). This because there is unbiased inheritance of both viability and fertility in the model so that during each event of reproduction, all the unreliable minimal reproducers produce, on average, the same number of offspring with the same intrinsic-invariable properties; namely, an average fertility of 5 for the offspring that are unreliable

minimal reproducers, and an average viability of 0.5 for all offspring (whether minimal reproducer or persistor).

Viability selection is thus the only form of selection in the population regardless of whether the population contains persistors or minimal reproducers. This means that in Model 3, viability is the only parameter relevant to estimate the fitness of entities of both classes (that is, viability is the only actual difference maker in the evolutionary dynamic). Therefore, to evaluate whether entities with the highest fitness are able to invade the population, we need to see whether the variants of each type (that is, persistor and minimal reproducer) with the highest viability allowed by the model (in this case 0.99) are able to invade the population.

Figure 3.3*b* represents a typical simulation of Model 3 over 400 units of time in which the proportions of entities with a viability of 0.99 is measured. The figure shows that the proportion of entities with a survival rate equal to 0.99, very low at t=0, increases and reaches an equilibrium of around 0.05, at around t=130. Thus, in this model the population is not invaded by the fittest variants of each type. This is explained by the fact that there is no biased inheritance of any properties in this set up, and successes and failures in terms of survival are recorded only during the lifetime of an individual. The memory necessary for the effects of natural selection to be recorded in this population is neither exhibited by persistor, nor by unreliable minimal reproducer. Each persistor procreated by a minimal reproducer has the same chance of having a low viability as a high one, in spite of the fact that most of the persistors procreated come from minimal reproducers with high viability. Essentially, the form of ENS observed in Model 3 is not very different from weak ENS apart from the fact that there is minimal reproduction in the model, which allows for multiplication of the substrate of the population. I name it *minimally-reproductive* ENS and it does not satisfy *c* (see Table 3.1).

Table 3.2. The different concepts explored in sections 3.3 and 3.4.

Model	Property introduced to entities, besides persistence	Description of the property introduced	Entities observed	Type of ENS observed
1	None	NA	Persistors	Weak, with complete exhaustion of the population
2	Procreation	Reproduction with no inheritance, including the ability to procreate	Persistors Procreator	Quasi-Weak, with complete exhaustion of the population
3	*Unreliable* minimal reproduction	*Unbiased* inheritance of the ability to procreate	Persistors Unreliable minimal reproducers	Unreliable-minimally-reproductive: self-maintenance of population size, but no changes in the background.
4	Darwinian minimal reproduction	As Model 3 + Biased inheritance of one or more traits with viability consequences	Persistors Darwinian Reproducers	*Minimally-reproductive and cumulative*: complex cumulative adaptations are possible
5	*Reliable* minimal reproduction / Reproduction	As Model 3 + Mutations increasing or decreasing the degree of transmission of the ability to procreate, and mutations increasing or decreasing the bias of transmission of this ability	Persistors Reliable minimal reproducers then Reproducers only	From *reliable-minimally-reproductive* to *reproductive*: as in Model 3 but with possible changes in the background against which new mutations can occur.
6	Reproduction / Replication	As Model 5+ Mutations increasing or decreasing the bias of transmission of one or more traits with viability consequences	Reproducers then Darwinian reproducers then Replicators	*reproductive and cumulative* i.e. *paradigmatic*: complex cumulative adaptations are possible

3.4. Biased inheritance and paradigmatic ENS

Unbiased inheritance of the ability to procreate coupled with unbiased inheritance of fertility and viability are thus insufficient for natural selection to sustainably record past successes and failures in reproducing populations, and therefore to change the population-level

background. In the following model (Model 4), taking the same set-up as in Model 3, I introduce biased inheritance of viability in order to evaluate whether it allows the entities with the highest viability to invade the population and thus change the population-level background against which new mutations can occur (see Table 3.2). There is no straightforward way to choose which level of biased inheritance of viability should be introduced. With perfect inheritance of viability, the same survival rate is transmitted from the parent to the offspring. With a very slightly biased inheritance, this viability is only slightly more likely to be transmitted than any other. For the sake of simplicity, I chose to introduce perfect (biased) inheritance of viability in Model 4. The mutant initially present in the population is still an unreliable minimal reproducer because the ability to reproduce is not transmitted reliably from parent to offspring; but it is also what I call a *Darwinian* reproducer because its viability – one of the classical proxies of fitness (hence the prefix 'Darwinian') – is transmitted with a higher probability than chance to its offspring.

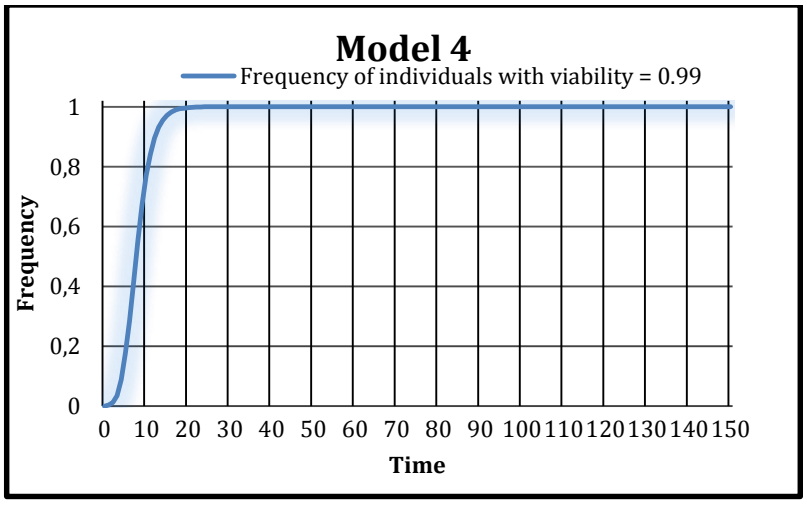


Figure 3.4. Simulation of Model 4 over 150 units of time.

Figure 3.4 represents the graph of a typical simulation of Model 4 over 150 units of time in which the proportion of entities with a viability of 0.99 is measured. The figure shows that the proportion of entities with a survival rate equal to 0.99, very low at t=0, increases and in contrast to Model 3, does not reach an equilibrium but rapidly invades the population. This is explained by the fact that, because there is biased inheritance of viability in this set-up, successes and failures are now recorded by natural selection. Each persistor procreated by a minimal reproducer has the same viability as its parent. Because more procreated individuals (whether persistors or minimal reproducers) have a viability of 0.99 and the population is able to maintain its size, the variants with the highest fitness invade the population.

Thus, Model 4 shows that biased inheritance of viability allows the population to satisfy *c*. In the population described by the model, individual changes brought about by advantageous mutations can be transmitted to the whole population. This is the mark of cumulative ENS and of populations able to display complex adaptations such as the human eye or brain. I call this form of ENS, *minimally-reproductive cumulative* ENS. Yet, minimally-reproductive cumulative ENS is not *reproductive cumulative* ENS (that is, paradigmatic ENS), in which there are no minimal reproducers, but reproducers only – all (or almost all) of which are able to transmit indefinitely their ability to procreate to all of their offspring. For a population to be able to exhibit paradigmatic ENS it must meet requirement *d*; that is, that virtually any entity in the population is in principle able to reproduce (see Table 3.1).

With Model 5 I will show that highly biased inheritance of the ability to procreate is a cumulative or complex adaptation of mixed populations of persistors and minimal reproducers (that is, initially exhibiting minimally-reproductive ENS), leading those populations to exhibit *reproductive* ENS (that is, ENS in which every entity reproduces) and thus satisfy *d*. Although

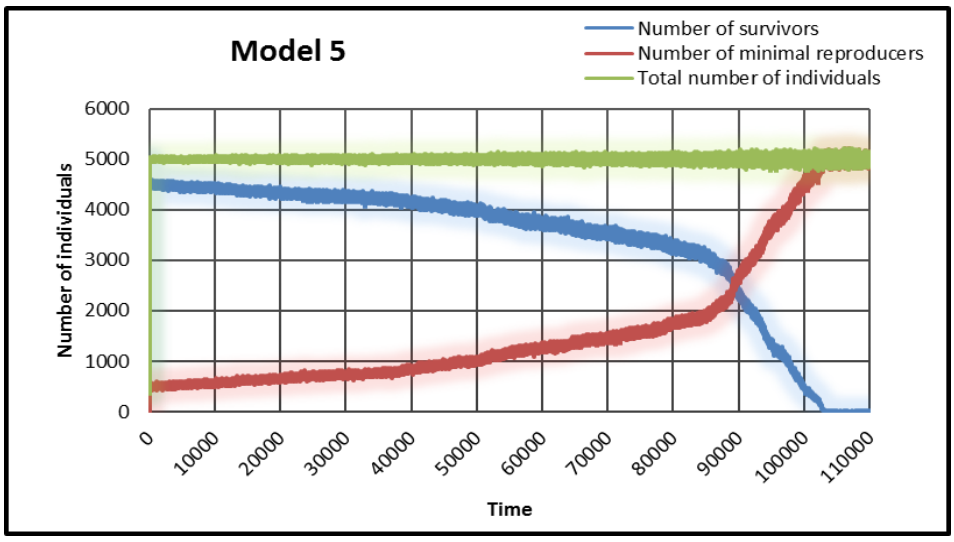
Model 4 showed that biased inheritance of viability is adaptive, it did not show how we can move from a population with unbiased inheritance of viability to a population with a high level of inheritance of viability; that is, the endogenising of biased inheritance of viability. That is what I will demonstrate with Model 6, in which both *c* and *d* (see Table 3.1) are satisfied within one and the same population.

How can we get reproductive ENS from a mixed population of persistors and minimal reproducers? To answer this question, I start with the set-up used in Model 3, in which there is unbiased inheritance of the ability to procreate (with a maximum level of 0.2). The result we observed was a population ultimately able to maintain its initial size but composed of a mix of persistors and minimal reproducers with very low levels of inheritance of the ability to procreate. To have a population composed of reproducers only, two things must happen. First, there needs to be inheritance of the ability to procreate at the maximal level (1) resulting in the population becoming a population in which there are *reproducers* and not solely minimal reproducers as in the in models 3 and 4. Second, each offspring procreated by a reproducer must receive the ability to procreate at the same level as its parent, that is, transmit this ability to all of its offspring. In other words, there must be perfect biased inheritance of the ability to procreate, the population thus becoming a population of *reliable* reproducers. In other words, we want to know how to move from a population similar to the population of Model 3 to a population with perfect inheritance on the ability to procreate at a maximal rate, that is, a population of reliable reproducers.

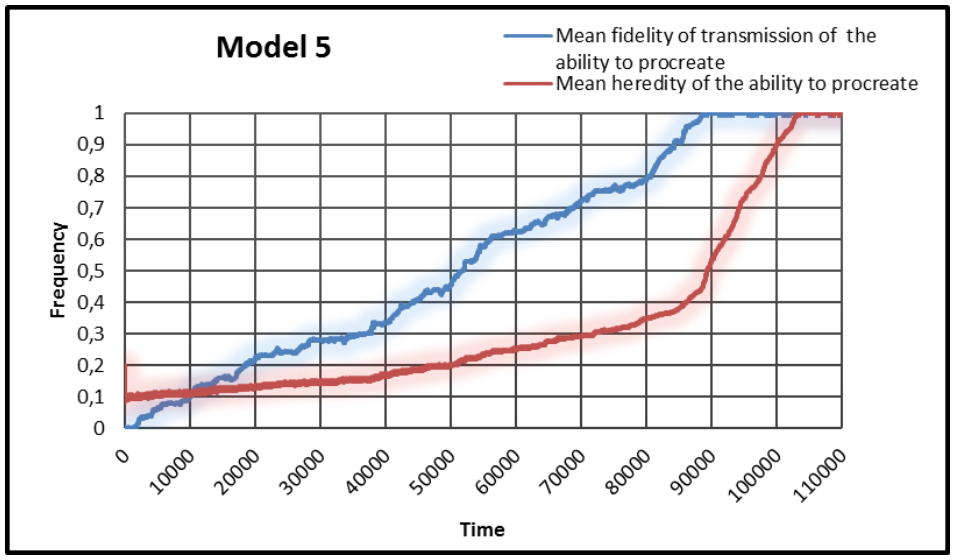
Model 5 shows how that is possible. There are two crucial differences between Model 3 and Model 5 (see Table 3.2). In Model 5, on top of the three stages (selection, reproduction and checking-for-overcrowding) undergone by each minimal reproducer, a fourth 'mutation' stage is added before the selection stage so that the sequence of events undergone by every individual

each unit of time is *mutation* → *selection* → *reproduction* → *check-for-overcrowding*. During the mutation stage, each minimal reproducer can mutate independently on two traits at a 10^{-3} rate. A mutation on the first trait increases or decreases, with the same probability and to the same degree, the maximal rate at which an individual is able to transmit its ability to procreate (that is, increases or decreases inheritance of this ability). If the maximal rate of transmission of a minimal reproducer is 1, it becomes a reproducer. A mutation on the second trait increases or decreases the fidelity at which the maximal ability to procreate of an individual is transmitted. If the fidelity of the transmission of the ability to procreate is nil, there is unbiased inheritance of the ability to procreate at a given level. If the fidelity is perfect, there is perfect (biased) inheritance of the ability to procreate and *unreliable* reproducers become *reliable* reproducers.

As with Model 3, I initially set the maximal level of inheritance of the ability to procreate at 0.2 for each minimal reproducer in the population. This means that at most, only 20% of the offspring of these minimal reproducers will also be minimal reproducers. The other 80% will be persistors. The fidelity at which each minimal reproducer can transmit its ability is initially nil, so that there is unbiased inheritance of the ability to procreate. This means that each procreated minimal reproducer will have initially a level of inheritance of the ability to procreate randomly chosen between 0 and 0.2. At each event of mutation on the maximal level of the ability to procreate, a minimal reproducer chosen randomly in the population sees its maximal level of procreation increase or decrease by 0.01, with the same probability. Mutatis mutandis, at each event of mutation on the fidelity of transmission of the ability to procreate, a minimal reproducer chosen randomly sees the fidelity at which it will transmit its ability to procreate increase or decrease by 0.01. All other parameters are kept identical to Model 3.



a.



b.

Figure 3.5. Simulation of Model 5 over 110 000 units of time.

Figure 3.5*a* represents a typical simulation of Model 5 in which the number of persistors and reproducers, as well as the total number of individuals is recorded over time. During the first units of time (which are not visible because Figure 3.5*a* is not detailed enough) the total number of entities in the population rapidly decreases. The explanation is strictly the same as for the

previous models: most persistors and minimal reproducers die because their survival rate is low. Only persistors and minimal reproducers with a high viability survive. Some of the minimal reproducers start procreating new minimal reproducers with a maximal rate of 0.2, which in turn are able to reproduce. This explains why the population does not go extinct and starts increasing in size: more and more minimal reproducers are procreated. Figure 3.5*b*, shows that both the maximal level of inheritance of the ability to procreate and the mean fidelity of inheritance on the ability to procreate, starting respectively with a value of 0 and 0.2, reach values superior to 0.99; after t=884 000 and t=104 000 respectively (in this simulation). This explains why at t=104 000, the population is composed of more than 99.5 % minimal reproducers (which are now effectively reproducers). At that stage almost no persistors are procreated because the maximal level of inheritance on the ability to procreate is nearly 1 and the fidelity with which this level is transmitted is also nearly 1 (as shown on Figure 3.5*b*). If 1 is not reached in both cases it is only due to the fact that some minimal reproducers mutate every unit of time (on average 5 entities per unit of time in a population of 5000 entities) and half of those mutations reduce the average fidelity and average maximal level of transmission of the ability to procreate.

Model 5 shows that high inheritance of the ability to procreate, as well as high fidelity on the level of transmission of the ability to procreate are both attractors of minimal Darwinian populations. In fact, a population composed mostly of persistors and minimal reproducers with a very low level of inheritance of the ability to procreate and very low fidelity of transmission of this ability is invaded by mutants with very high level of biased inheritance on this trait through successive mutations, and leads the population to become a population of *reproducers*. More interestingly, mutants that overall decrease the number of minimal reproducers produced each unit of time, which represent 50 % of all mutations, seem to play no or only a limited role in the

evolutionary dynamics of the population. There is here an asymmetry between mutations increasing and mutations decreasing the level of inheritance of the ability to procreate. This asymmetry is what makes changes to the population-level background against which new mutations can occur possible.

The explanation underlying the asymmetry is the following. Each time a new mutation increases either the maximal level of inheritance of the ability to procreate or the fidelity with which this maximal level is transmitted, this increases the rate of production of minimal reproducers because the minimal reproducers procreated are themselves able to procreate new minimal reproducers. If, on the contrary, a mutation increases the level of production of persistors (via a mutation decreasing inheritance of ability to procreate or the fidelity with which it is transmitted), this mutation will have long-term downstream consequences that will lead the bearers of this mutation to have overall less offspring than the rest of the population, and their type will be less represented in (and ultimately eliminated from) the population. This, I claim, represents a genuine case of cumulative ENS. Reproduction can thus be a complex (or multiple-step) adaptation in minimal Darwinian populations, which emerges as a result of cumulative ENS. Model 5 satisfies requirement *d*, but as such not requirement *c* because inheritance on traits other than the ability to procreate is nil.

In a population in which every individual minimally reproduces, paradigmatic ENS is now extremely easy to attain. The only requirement is biased inheritance on traits having consequences for viability or fertility, as in Model 4. When this occurs, reproductive cumulative ENS – that is, paradigmatic ENS – becomes possible, and structures like the human eye become in principle much more likely to appear because the number of trials made by natural selection is maximal.

To show this, I created Model 6 (see Table 3.1), which starts with the setup obtained at the end of the simulation of Model 5: a population of reproducers.

In a population of reproducers, as I defined them, the only trait reliably transmitted is the ability to procreate. All other traits are uncorrelated between parents and offspring (unbiased inheritance). Let us now imagine a population of reproducers in which each reproducer is able to mutate, as in Model 5, but on two different traits, one increasing or decreasing its viability,[71] the other increasing or decreasing the fidelity with which its viability will be transmitted to its offspring. If paradigmatic ENS is possible in such population, we should expect that ultimately the population will be invaded by individuals with the highest possible viability allowed by the model.

In Model 6, I start with a population of reproducers with viabilities and fertilities all chosen randomly between 0 and 0.99 and 0 and 10 respectively. As in Model 5 there is initially unbiased inheritance of viability and fertility. However, unlike Model 5, individuals can mutate at a 10^{-3} rate on two traits, namely their viability and the fidelity with which they transmit their viability to their offspring. Each mutation will lead to an increase or decrease on the value of the trait by 0.01 (unless the traits already have their minimal or maximal value). As in Model 5 the initial population is composed of 5000 individuals. For the sake of simplicity, viability is transmitted as if it were only dependent on one trait (e.g., size), but more complex simulations could be used with viability depending on more traits.

[71] A similar mutation could appear and affect fertility.

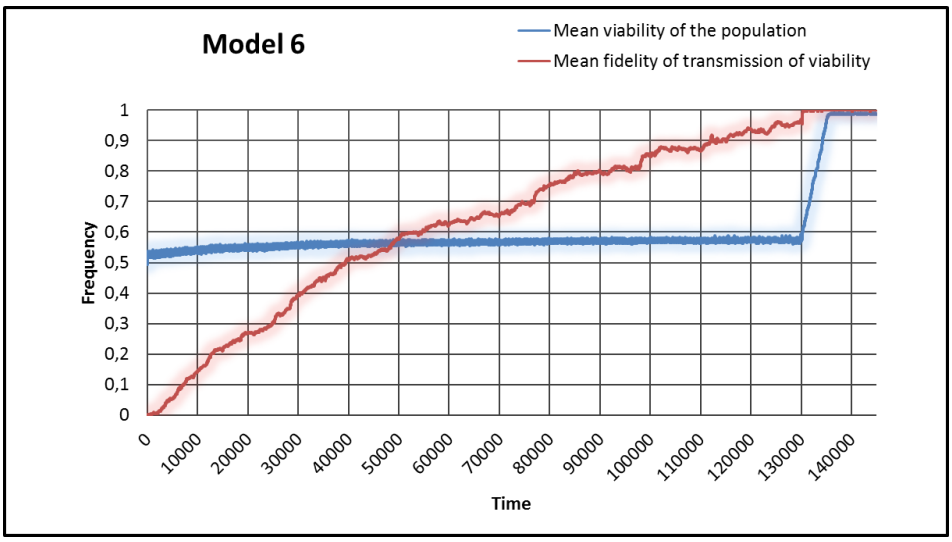


Figure 3.6. Simulation of Model 6 over 145 000 units of time.

Model 6 shows two things. First, high inheritance of viability is obtained and can be regarded as a complex adaptation of populations of reproducers analogous to reproduction as an adaptation of populations of minimal reproducers. In fact, as can be seen in Figure 3.6, although half of the mutations decrease the fidelity of transmission of viability from parent to offspring, after t=133000, fidelity is maximal. This is explained by the fact that reproducers with higher viability have more offspring than reproducers with lower viability when they are able to transmit their viability with a probability higher than chance. This also explains why the mean viability of the population initially slowly increases (from 0.5 at t= 0 to 0.575 at t=130 000 in this simulation). As biased inheritance of viability reaches its maximal value, the dynamic of the population radically changes: the mean viability of the population rapidly increases until it reaches the maximal value allowed by the model and remains at that level. The explanation here is exactly the same as previously. The only difference is that reproducers are now *replicators*. Replicators are able to transmit Darwinian traits to *all* their offspring with very high fidelity. When that happens, *any* difference increasing fitness leads to *large* differences in the population-level background against

which new mutations can occur. Indeed, at that point, any mutation increasing the viability of replicators is perfectly recorded by natural selection. Increases in viability are also recorded before the appearance of replicators, but the recording is imperfect and thus changes in population level background do not have the dramatic effects they have in populations of replicators.

I will make a further distinction between *minimal replicators*, *replicators* and what I call *ultimate replicators*. Minimal replicators only transmit faithfully *one* of their traits (aside from the ability to procreate) to their offspring. This definition of replicators leads to a form of replication similar to the one argued for by Nanay (2011), who claims that replication is about properties rather than entities. Replicators transmit faithfully more than one trait to their offspring, while ultimate replicators are replicators in a classical sense (Dawkins 1982, Hull 1980) and transmit all their traits to their offspring. Further models could show that the selection of a global mechanism of transmission of many, if not all traits, is more advantageous than a successive selection of particular mechanisms of transmission for each Darwinian trait. Thus structures like DNA with 'unlimited' heredity (Maynard Smith & Szathmáry 1995, 43) and proofreading during transmission, are expected products of natural selection. I leave this demonstration open to further work.

Using different models I have shown that reproduction and inheritance, usually conceived of as requirements for ENS, are actually predictable products of it, given that new variation is introduced in a population of persistors. By doing so, I have endogenised reproduction and inheritance in a broader conception of ENS.

3.5. Assessing the endogenisation-of-inheritance hypothesis

The six models I have proposed demonstrate that under very flexible assumptions both reproduction and inheritance are likely to appear and spread by natural selection in populations of entities in which those properties are not initially present. It follows that they embody a more abstract conception of ENS than standard conceptions in which reproduction and inheritance are taken as necessary conditions for selection to occur. Griesemer (2005) argues that abstractions of Darwinian evolutionary theory take two forms. One consists in abstracting the level at which ENS can occur. This is the strategy adopted by Lewontin (1970) in his classical paper "The Units of Selection". The other strategy consists in abstracting away from the contingent mechanisms of biological entities, keeping only the functions that are necessary and sufficient for ENS. Where do the endogenisation-of-inheritance hypothesis resulting from the 6 models I have proposed stand with regards to Griesemer's two forms of abstraction? In none of the 6 models did I make any assumptions concerning the level at which the mechanisms of viability, inheritance of and inheritance occur. Nor did I give a precise mechanism of inheritance. Thus, these models represent both an abstraction from any specific level of selection and an abstraction away from specific biological mechanisms. They employ both of Griesemer's strategies of abstraction.

Earlier examples of the second strategy of abstraction have concluded that replication, or at the very least reproduction, is necessary for ENS. I criticised this view in the previous chapter and the models presented in this chapter further prove the point. My endogenisation-of-inheritance hypothesis represents a plausible Darwinian explanation of the origin of inheritance starting with persistors and assuming some random events of mutation.

Some will regard the abstraction instantiated by the six models as justified simply by the fact that it allows the statement of the endogenisation-of-inheritance hypothesis. Others will wonder what role such a very abstract generalisation of evolutionary theory can have in understanding real biological phenomena. In the remainder of this section I briefly explore three areas in which the endogenisation-of-inheritance model I have proposed could potentially be useful.

The first area in which the endogenisation-of-inheritance model can be applied is obviously origins of life studies, and especially in the study of the origins of ENS. Here is not the place to discuss the different theories on the origins of life and the strengths and weaknesses of each of them (for reviews see Cleland 2013; Dyson 1999; Fry 2011; Maynard Smith & Szathmáry 1995). I simply note that the origins of life and the origins of ENS, although often treated simultaneously, should be distinguished – for a population might be composed of entities considered as not alive and yet undergoing ENS. To a large extent the series of models I have proposed represents a detailed scenario of Dawkins's hypothesis in the second chapter of *The Selfish Gene* (Dawkins, 1976) where he notes in passing that "Darwin's 'survival of the fittest' is really a special case of a more general law *of survival of the stable"* (1976, 13, my emphasis), and that life originated from selection of molecules for stability. But Dawkins jumps very quickly to the idea that, through some improbable but possible event, a molecule or a complex of molecules must have acquired the ability to self-replicate using small building-blocks available in its surroundings and thus become the first 'replicator'. The same leap is made by Michod (1999) and by Maynard Smith and Szathmáry (1995)**,** among others. These authors presuppose that evolution becomes an interesting phenomenon (and that life emerges) only when an entity is able to self-replicate with high fidelity. One virtue of the endogenisation-of-inheritance model I have

proposed, is that it shows that ENS can occur and lead to a complex adaptation such as high-fidelity inheritance even without high-fidelity replication initially and without a one-step large mutation. The replicator-first view of the origins of ENS is thus unwarranted. Whether the scenario I have proposed for the origins of reproduction and inheritance is the correct one is not of the first importance, since my goal was primarily to clarify the space of possibilities with respect to reproduction and inheritance. And in fact, procreation, inheritance of procreation, and the mechanisms rendering this inheritance biased might have all originated simultaneously, rather than sequentially. But the value of recognising that there are different concepts behind what is usually referred as 'inheritance' is demonstrated by the fact that it allows us to see that there are different alternative scenarios for this key step in evolution.

Another domain for which the endogenisation-of-inheritance hypothesis might be applicable is in the broader study of evolutionary transitions in individuality (see Chapter 6 for more details on evolutionary transitions in individuality). In the models proposed above, the agents could represent any entities at any level. Procreators and minimal reproducers could represent early multicellular organisms for example. Each primordial multicellular organism would initially fission and would pass on its traits at the collective level more or less randomly from one generation to the other. Collectives would, even at that stage, be selected for and evolve by natural selection, because those collectives that passed on their traits with more certainty to their offspring would have an advantage, and thus spread.

Finally, the endogenisation-of-inheritance hypothesis might be useful in other domains in which the Darwinian apparatus has been deployed, most notably with cultural evolution (Claidière, Scott-Phillips, & Sperber 2014; e.g., Godfrey-Smith 2009b; Godfrey-Smith 2012; Laland & Brown 2002; Mesoudi, et al. 2004; e.g., Richerson & Boyd 2005; Sperber 1996). If a

replicator is not necessary for ENS in the biological realm, then efforts to find a cultural analog may be unnecessary. This conclusion should be welcomed by cultural evolutionists, since forms of 'memetics' relying strongly on the notion of replicator (Blackmore 1999; Dawkins 1976; Dennett 1995) have failed to produce many significant results and their theoretical foundations have been heavily criticised (e.g. Sperber 1996, 2000). I will have more to say on cultural evolution in the general conclusion of this thesis.

The mere fact that the endogenisation-of-inheritance hypothesis raises questions about what counts as Darwinian evolution is important and might lead to further development of the idea.

3.6. Conclusion of the chapter

In this Chapter I have shown four things about ENS. First, strictly speaking, reproduction is not needed for ENS to occur, as I have already emphasised in the previous chapter. Viability selection is sufficient to produce evolutionary changes through the action of natural selection. If reproduction is not necessary, the same can be said about inheritance, since the former is a prerequisite for the latter. This is also consistent with the ideas developed in Chapter 2. Second, I have shown that paradigmatic ENS can arise through successive episodes of adaptation in populations of persistors, provided that suitable mutations can arise in these populations. In doing so, I have endogenised reproduction and inheritance into a broader conception of ENS. Third, I have briefly shown that replicators, and specific replicating structures such as DNA, can arise as adaptations in populations of entities that do not reproduce with high-fidelity. Finally, I have shown potential applications of the endogenisation-of-inheritance hypothesis in three domains of current evolutionary research. One implication of these ideas that I have not

developed here, but has been a recurrent theme of chapters 1 and 2 and will continue to be in chapters 4 and 5, is the links between inheritance and the concept of fitness. We have seen that even without differences in viability or/and fertility the evolutionary success of entities with different level of inheritance on the ability to procreate could be predicted. This clearly is problematic if a classical definition of fitness based only on viability and fertility is used without references to type identity.

Most of the work done so far was made without direct reference to formal evolutionary theory and the reader might doubt that what I have argued for is in line with it. In the next chapter, using the Price equation, a classical tool of evolutionary theory, and comparing it with the classical approach used so far, I vindicate the view that ENS does not require reproduction in order to occur. I focus my analysis more specifically on the concepts of heritability and fitness in this equation.

4. Moving from the Classical to the Formal Account of ENS: Heritability and Fitness reconsidered

4.1. Introduction

We have seen in the previous chapters that a generic recipe for ENS can roughly be delineated as follows. For ENS to occur, a population of entities must exhibit 1) variation, that leads to 2) difference in fitness among the variants and that 3) is heritable. We have also seen that in spite of the different meanings the term *evolution* in the expression *evolution by natural selection* can take (see the Introduction of the thesis), one legitimate meaning is changes in the frequency of variants in a population over time. This is a classical definition in population genetics. We have also argued that this recipe is supposed to apply not only to biological populations of organisms but to *any* population of entities, as well as any level of organisation and any temporal scale.

The classical approach to ENS represents one way to abstract the Darwinian apparatus. But in recent years, evolutionary theory is undergoing another form of abstraction that relies more heavily on mathematics. I call this approach the *formal* approach to ENS. The works of Okasha (2006), Rice (2004) or Grafen (2009), to cite only few examples, represent this tendency. All of them use the Price equation which they regard as an exact and highly general formulation of the evolutionary change occurring between two times (often two generations). In this chapter I follow suit and will use the Price equation to characterise the formal approach. More precisely, I show that the view that natural selection is independent from reproduction, which has been vindicated in the two previous chapters from a classical approach, finds a natural interpretation

in the Price equation. More precisely I challenge the classical interpretations of fitness as being related to reproductive output, and heritability as being a population level measure of parents-offspring transmission.

The remainder of the chapter will be divided into six sections. In Section 4.2, I present one possible derivation of the Price equation. In Section 4.3, I present the different approaches to heritability, followed by the formulation of the Price equation proposed by Okasha in his 2006 book which makes explicit the term h^2 of heritability. This formulation will be useful later on to bridge the two approaches to ENS for, in its classical formulation, the Price equation has no term referring to heritability while heritability is one of the three ingredients of the classical approach.

In Section 4.4, I examine what looks to be a potential problem for the classical approach to ENS identified by Earnshaw-Whyte (2012). Although this chapter will mostly be concerned with the formal approach to ENS, this step back to the recipe approach will be necessary to show how the two approaches can interact and enlighten each other, and also that there is fundamentally no disagreement between them in spite of an apparent one identified by Earnshaw-Whyte. According to Earnshaw-Whyte, one can find cases of ENS in which there is no heritability, yet ENS is observed if one refers to the classical meaning of evolution in population genetics, that is, distributive evolution in my terminology (see Introduction of the thesis). He proposes a verbal model in which there is ENS but no heritability as classically interpreted and therefore that the recipe approach seems, under his reading, unable to account for evolutionary change. This leads him to conclude that the classical concept of ENS is irremediably flawed and should be abandoned. I claim that his conclusion is too strong. Under an uncharitable reading of Earnshaw-Whyte's model, an intuitive interpretation of the recipes leads to the correct prediction. Under a more charitable reading of his model, this intuitive

interpretation leads to wrong predictions of evolutionary change as Earnshaw-Whyte claims. I show however, that the only difference between the two readings of Earshaw-Whyte's model depends on an interpretation where individuals in the model are considered as parents and one where they are considered as offspring. This will substantiate the claims I made in Chapter 2 that the notion of individuality in evolutionary theory is a fuzzy concept. I will reaffirm that a difference in interpretation as to whether an individual reproduces or persists should not be the basis for changing our claims about heritability and ENS.

In Section 4.5, in light of the two possible readings of Earnshaw-Whyte's model, I propose that a correct interpretation of the notion of heritability includes both the offspring of the parental population and the parents that have survived after reproduction (if any) and present in the offspring population. I call this interpretation *temporal* heritability. Further on, I show that this interpretation is in accordance with the Price equation since the Price equation can describe the evolutionary change of populations of individuals over an arbitrary period of time and in which selection involves differential persistence instead of differential reproductive rates (Rice 2004, 169). Finally, starting from the formulation of the Price equation given by Okasha and presented in Section 4.3, I show that temporal heritability can be separated into two different components: one I name *generational heritability* which corresponds to the intuitive concept of heritability and the second is *persistence heritability*, usually neglected in the recipe approach. I show that both persistence heritability and generational heritability can be nil in some cases (e.g., non-overlap of generations in the case of persistence heritability, no reproduction in the case of generational heritability), but that in most real biological populations they both need to be taken into account to accurately predict evolutionary change.

In Section 4.6, I make a similar move to the one made for heritability in Section 4.5 with the Price equation, but focus instead on the term w, that is fitness. I show in formal terms that ENS does not requires differences in reproductive output to occur and that differences in persistence and developmental growth can lead to ENS in a substantial way. In light of these demonstrations, I propose towards the end of the chapter the correct interpretation one should give of the terms 'fitness' and 'heritability' in the recipes for ENS. I will conclude that there is no fundamental opposition between the classical approach to ENS and its formal counterpart if one generalises the concept of fitness and heritability to cases that only involve differential persistence.

4.2. The Price equation

In 1970 George Price published what would become the Price equation (Price 1970; see also Robertson 1966) that was extended to its hierarchical form in 1972 (Price 1972a). The Price equation is a simple algebraic result describing the evolution of a character in population between two times, generally two generations. The success and widespread use of this equation by theorists owes to its generality. Price, in his 1970 seminal paper was well aware of that when he writes that his equation

> gives frequency change under selection from one generation to the next for a single gene or for any linear function of any number of genes at any number of loci, holds for any sort of dominance or epistasis, for sexual or asexual reproduction, for random or non-random mating, for diploid, haploid or polyploid species, and even for imaginary species with more than two sexes (Price 1970, 520).

Price uses here the language of genetics but the Price equation can be applied to all sorts of entities, whether under genetic dependence or not (see for example El Mouden, André, Morin, & Nettle 2014 who apply the Price equation to cultural transmission; see also Helanterä & Uller 2010 who apply the Price equation for different inheritance systems from the genetic one)

There are many ways to derive the Price equation.[72] Below I chose one, inspired from Okasha (2006), that is formulated in phenotypic rather than genetic terms. In his book, Okasha is concerned primarily with the issue of the levels of selection. If it is true that the recursiveness[73] of the Price equation makes it particularly handy to treat multilevel selection, the simple form of the Price equation is also very useful to think about other problems in evolutionary theory. I will mostly be concerned with the simple form, that is, non-recursive, in this chapter and more generally in this thesis. Application to the multilevel case can be found for example in Hamilton (1975), Okasha (2006) or Frank (1998, 2012a).

Let us suppose a population of entities (for example of biological organisms) of size n, in which we want to know the evolution of a particular trait in the population, say the height of the organisms. For simplicity, we suppose discrete generations (each time an organism reproduces, it immediately ceases to exist and all organisms reproduce simultaneously). These assumptions will be partially relaxed in later sections. In any case, these assumptions do not restrain the validity of the Price equation only to asexual populations and with discrete generations.

[72] For different derivations of the Price equation see for example Okasha (2006), Rice (2004), Frank (1997, 1998, 2012b), Grafen (2000), Hamilton (1975), West, Griffin, & Gardner (2008), Gardner & Foster (2008), Queller (1992).
[73] 'Recursiveness' means here that one term of the Price equation can be decomposed using the same form of the Price equation. For instance the set of equations $a = b - c, c = d - e, e = f - g$, etc., exhibits recursiveness since the term c of the first equation can be decomposed in two terms following the same form as the first equation.

To study the evolution of a trait is equivalent to study the change in the average value of this trait in the population between two times. Each entity i of the population has the trait z in a certain state z_i. Here, z_i is the height of the individual i but any trait with more than one possible state will do, so that even so called 'discrete traits' can be accounted by the Price equation (see Okasha 2006, 24-25). For instance, if one wants to measure the evolution of the presence or absence of a particular phenotype they can choose to give the convention 0 for the absence and 1 for the presence of the phenotype.

Let us call the average value of the trait z in the population at the first generation $\bar{z}$ so that:

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$$

Each entity also has a reproductive output w_i. Most people, when referring to the Price equation, call the value w_i the fitness of individual i. I see no problem with this terminology, but it should be clear that fitness here does not represent the same notion of fitness we presented in the previous chapters, that is, it does not represent the concept of fitness associated with natural selection, but what we have called in Chapter 1 a proxy of fitness or realised fitness. In fact since different individuals can have a different reproductive output that can be attributed to differences in intrinsic-invariable, intrinsic-variable or extrinsic properties, not all differences should be attributed to natural selection. By measuring reproductive output indiscriminately, the term w could be associated with for instance, deterministic drift rather than natural selection. Thus, by

calling the term w_i 'fitness', one entertains an ambiguity which I regard at the origin of a number of issues surrounding the concept of fitness. For that reason in the context of a population with discrete generation, I prefer calling w_i the reproductive output of the individual i in order to avoid any ambiguity. Let us call the average reproductive output of the parent population $\bar{w}$ so that:

$$\bar{w} = \frac{1}{n}\sum_{i=1}^{n} w_i$$

For each individual i, we can also define the average value of the trait z after one generation as:

$$z_i' = z_i + \Delta z_i$$

were Δz_i is called the transmission bias. If reproduction is on average perfect we have $z_i' = z_i$.

With these basic definitions we can now calculate $\bar{z'}$ which represents the average trait of the character in the population after one generation:

$$\bar{z'} = \frac{1}{n}\sum_{i=1}^{n} \frac{w_i}{\bar{w}} z'_i$$

$\frac{w_i}{\bar{w}}$ represents the relative reproductive output of entity i. If an entity has more offspring than another, its weight on $\bar{z'}$ in the offspring population will be stronger. From there, we can calculate the average change in character z in the population between the two generations as:

$$\Delta\bar{z} = \bar{z'} - \bar{z} \qquad (4.1)$$

Replacing $\overline{z'}$ by the formula given above, we can rewrite equation (4.1) as:

$$\Delta\bar{z} = \frac{1}{n}\sum_{i=1}^{n}\frac{w_i}{\overline{w}}z'_i - \bar{z}$$

We can now replace z'_i by $z_i + \Delta z_i$ so that (4.1) can be rewritten as:

$$\Delta\bar{z} = \frac{1}{n}\sum_{i=1}^{n}\frac{w_i}{\overline{w}}(z_i + \Delta z_i) - \bar{z}$$

which once developed becomes:

$$\Delta\bar{z} = \frac{1}{n}\sum_{i=1}^{n}\frac{w_i}{\overline{w}}z_i + \frac{1}{n}\sum_{i=1}^{n}\frac{w_i}{\overline{w}}\Delta z_i - \bar{z}$$

If we multiply by $\overline{w}$ on both sides we get:

$$\overline{w}\Delta\bar{z} = \frac{1}{n}\sum_{i=1}^{n}w_i z_i + \frac{1}{n}\sum_{i=1}^{n}w_i\Delta z_i - \bar{z}\overline{w}$$

which once rearranged leads to:

$$\overline{w}\Delta\bar{z} = \frac{1}{n}\sum_{i=1}^{n}w_i z_i - \bar{z}\overline{w} + \frac{1}{n}\sum_{i=1}^{n}w_i\Delta z_i$$

We know that the covariance $cov(X,Y)$ between two random variables X and Y is defined as:

$$Cov\,(X,Y) = E(XY) - E(X)\times E(Y)$$

where $E(X)$ is the expectation of X.

Assuming n is very large we have:

$$E(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

With these two definitions, we can now rewrite (4.1) as:

$$\bar{w}\Delta\bar{z} = \ Cov(w_i, z_i) + \ E(w_i \Delta z_i)$$

Dropping the indices for convenience we obtain:

$$\bar{w}\Delta\bar{z} = \ Cov(w, z) \ + \ E(w\Delta z) \qquad (4.2)$$

Equation (4.2) represents the Price equation in its classical formulation. The first term of the RHS: $Cov\ (w, z)$, is the covariance the between absolute reproductive output w and character z of each individual. It is often interpreted as the evolutionary change due to *natural selection*. The second term of RHS: $E(w\Delta z)$ is the average, or expected value of the quantity $w_i \Delta z_i$, that is, the product of absolute reproductive output and transmission bias. It is often interpreted as the evolutionary change due to the *transmission bias*. Note that 'transmission bias' should not be confused with the notion 'biased inheritance' that was introduced Chapter 2, used extensively in Chapter 3, and will be used again in the present chapter and Chapter 4. The bias in 'biased inheritance' is in reference to the type of the parent(s) (biased toward the type of the parent), while the bias in 'transmission bias' refers to a departure from an event of perfect transmission. If reproduction is on average perfect, $\Delta z_i = 0$ that is, when z_i and z_i' are equal, there is no

'transmission bias' when referring to an event of perfect transmission, but inheritance, in my terminology, is strongly biased.

We can see in equation (4.2) that there is no mention of heritability while there is mention of variation in reproductive output. For that reason it renders this equation difficult to compare with the classical approach to ENS in which there is an explicitly mention of heritability. In the next section, after providing and discussing in more details the classical interpretation of the concept of heritability than I did in Chapter 2 (Section 2.3), I use one formulation of the Price equation proposed by Okasha, in which the term of heritability appears, to show a link between the recipe and formal approach to ENS.

4.3. Heritability in the Price equation

4.3.1. Different approaches to inheritance and heritability

As we have seen in Chapter 2, the classical approach to ENS requires inheritance between parent(s) and offspring as a necessary condition. I have mentioned one problem with the notion of inheritance, namely that it is a qualitative concept that does not correspond to what evolutionary theorists mean by inheritance. I proposed instead to use the notion of *biased* inheritance which better corresponds to this usage of inheritance. Basically, biased inheritance allows for *memory* in the population which is necessary for natural selection to have traction and lead to lasting evolutionary changes, what I called in Chapter 2, following classical terminology of quantitative genetic, the response to selection. But, as such, biased inheritance does not allow for quantitative predictions of evolutionary changes because it is an individual notion of

inheritance when what is required for measuring evolutionary change is a population level measure. To make these predictions, different statistical measures of inheritance, regrouped under the term *heritability*, have been developed. I briefly presented the two main approaches of heritability, namely the variance and the regression approach, in Section 2.3, and claimed in Subsection 2.3.2 that the variance approach might be superior in some cases. It is now time to give a more thorough treatment of the concept of heritability as it is used in evolutionary theory.

Heritability is usually defined as "a ratio of variances, specifically as the proportion of total variance in a population for a particular measurement, taken at a particular time or age, that is attributable to variation in additive genetic or total genetic values" (Visscher, Hill, & Wray 2008, 255). In a genetic context it can thus either refer to the part of the variance of a phenotype in a population attributable to total genetic variance *or* to additive genetic variance. In the first case heritability does not presuppose any inheritance mechanism of phenotypes. This is the notion of heritability preferred by behavioural geneticists and psychologists who are interested in knowing the contribution of genes to psychological traits (Downes 2009a). It is usually labelled broad-sense heritability: H^2. But broad sense heritability, as I mentioned in Chapter 2, is not the notion employed in the context of natural or artificial selection. In fact, in those later contexts, for heritability to be positive, there must be the transmission of the trait over time (usually from parents to offspring).

But not all genetic effects are transmitted between generations. In, fact H^2 includes genetic effects that, once meiotic segregation and recombination as well as fecundation have occurred in many sexual organisms, will not be transmitted to the next generation. These effects are classically called gene interaction (or epistatic) and dominance effects (Downes 2009a). Gene

interaction effects are effects on phenotypes that are due to the interaction of the combination of genes an organism happens to have. Dominance effects are a specific kind of gene interaction effect in which the interaction occurs between genes at the same locus (one gene dominates the other in its phenotypic expression). Because gene interaction effects broadly understood depend on the genetic context of the organism, the only effects on phenotype relevant for evolutionary theorists are effects that are kept between generations. These are called additive genetic effects. Narrow sense or realised heritability h^2 is the concept of heritability that measures these effects. It is defined as the proportion of phenotypic variance due to additive genetic variance.

Within evolutionary theory another, more abstract definition of realised heritability is often preferred. Heritability can be defined as "the linear regression of average offspring character on parental character, or mid-parental character if reproduction is sexual" (Okasha 2006, p. 34). This is the notion of heritability used among others by Godfrey-Smith (2007, 2009b), Okasha (2006, 2010), Rice (2004) and Roughgarden (1979).

Heritability can thus correspond to three distinct notions (Jacquard 1983; Sesardic 2005): 1) broad-sense genetic heritability, 2) narrow-sense or realised genetic heritability using the ratio of variance approach, henceforth the variance approach, and 3) heritability as a measure of parent-offspring phenotypic resemblance, henceforth the regression approach. Importantly, both the second and third notions of heritability correspond to realised heritability, the only form of heritability that matter for my purpose. One difference between these two notions of realised heritability is that the third notion can be applied to cases different from the biological ones, such as non-genetic inheritance (e.g., cultural evolution). Yet, both notions are used in evolutionary theory and there are disagreements over which one of them, if any, is the most general one. Some authors consider that the regression approach is more fundamental than the ratio of additive

genetic variance approach (Okasha 2006, 2010; Rice 2004), while others disagree. For example, Downes (2009b) proposes that the notion of heritability to be used should depend on one's agenda. In some cases, he argues, one of the two definitions might be more appropriate.

In Chapter 2, I claimed that in some respects the variance approach was superior to the regression approach, because the regression approach captured more than effects due to differences in intrinsic-invariable properties and thus could not be linked as straightforwardly to natural selection as the variance approach can (since differences in genes are a form of difference in intrinsic-invariable properties). But there are three main reasons advanced by the proponents of the regression approach to consider the regression approach as more general than the genetic variance approach. One is that genotypes or genes can be treated as phenotypes (Rice 2004, 165) which means that the regression is intrinsically more general than the variance approach. Another is that the Darwinian apparatus can be applied to substrates other than genes / genotypes (e.g., culture). Using a definition of heritability that relies on additive genetic variance will only be problematic in those cases. Finally, Rice (2004, 204-205) and Godfrey-Smith (2007, 510) both argue that in some significant cases of evolution, heritability can be negative or greater than 1. These values cannot be obtained if one uses the genetic variance notion of heritability where the values of heritability are comprised between 0 and 1 (inclusive).

I accept the first two reasons as being valid but as I will show, they can easily be accommodated by the variance approach. However, I believe that the third one is mistaken, if heritability is used to only capture ENS in a strict sense[74] and not any other part of the total evolutionary change. In fact, there are cases where regression heritability might be negative, but

[74] This is also valid for probabilistic drift (see Chapter 1).

that but that is due, I argue, to an evolutionary process conceptually distinct from natural selection. Suppose that regression heritability in a population between two generations is negative. One might conclude misleadingly that the evolutionary change observed is due to the traction given by heritability to natural selection when in fact what explains the evolutionary change observed is some particular mechanism of imperfect transmission between parent and offspring which conceptually distinct from selection – what I called 'transgenerational mutation' (or non-independent recombination in cases where the genetic determinism is known) in Chapter 2. In fact, if intrinsic-invariable factors are transmitted perfectly between two generations and their effects are additive, it is impossible that the slope of regression of average offspring character on parental character is superior to 1 or inferior to 0 without invoking another factor.

Another problem was mentioned in Subsection 2.3.1 where a correlation between the environment of the parents and the offspring in the population that is not due to the causal role of entities, leads to a positive covariance between parent and offspring character, but one that should be associated with ENS. This represents a second possible factor that would lead to a slope of regression superior to 1 or inferior to 0. Using the variance approach does not lead to this kind of problems (at least in theory) and if one intends to give a causal interpretation of heritability, as I do, this is a substantial difference between the two approaches.

It seems that a good balance between the variance and regression approach to heritability can be found if one recognises that the right concept of heritability to be associated strictly with natural selection in evolutionary theory should include cases in which inheritance between parents and offspring is not genetic and at the same time finds a natural causal interpretation. The concept of gene is understood here in its molecular sense (Griffiths & Stotz 2013), that is roughly speaking DNA based, yet intrinsic-invariable, such as for example the cortical pattern of some ciliates

(Beisson & Sonneborn 1965). Remember, as we saw in chapters 1 and 2, that only intrinsic-invariable properties can be the basis for the process of natural selection. Yet, it seems quite improbable that only genes would be intrinsic-invariable in biological systems.

One way to delimit this more general concept is to use the regression approach with two additional requirements. The first requirement is that the environment of the parent(s) and the offspring should be kept constant or if not that any correlation between them should be accounted for. This is to avoid the problems of correlation between parent and offspring environments (or differences in intrinsic-variable and extrinsic properties) that are not accountable by differences in intrinsic-invariable properties (see Subsection 2.3.1). The second requirement is that the mode of transmission between all the members of the population is the same, so that the rate of production of intrinsic-invariable properties between parent and offspring is the same across all members of the population.

Another solution, maybe simpler, is to use the variance approach while recognising that *any intrinsic-invariable character*, whether genetic or not,[75] transmitted from parent(s) to offspring should be accounted for by this definition of heritability and thus use a definition of heritability as the part of the phenotypic variance due to the additive *intrinsic-invariable* variance instead of *additive genetic variance*. I have no particularly strong views on which one, if any, of these two possibilities is the best and will leave it open for further work.

There are thus different concepts of heritability used in the literature with only one that is of interest for this issue, namely narrow-sense (or realised) heritability which can be estimated

[75] 'Genetic' is used here in its molecular sense, not its evolutionary sense (see Griffiths & Neumann-Held 1999 for the differences between these two concepts of the gene).

in two different ways. In the rest of the chapter, I will use the regression method, since this is the one privileged by Okasha and I will follow some of Okasha's formalism throughout the remainder of the chapter. Keeping in mind the two limitations mentioned earlier, I will suppose the same mode of transmission between all the members of the population and that there is no covariance between parent and offspring environments.

4.3.2. Heritability: from recipes to the Price equation

I will now introduce a formulation of the Price equation different from the one presented in Section 4.2 that will help me dispute the idea that the existence of parents and offspring is necessary for ENS. The approach to heritability I will use will be based on the regression approach.

In the first chapter of his 2006 book, Okasha proposes a bridge between the classical and the formal approach to ENS. His aim is to formulate the Price equation in a way that will make apparent the three ingredients for the recipes for ENS. As we have seen in Section 4.2, the classical formulation of the Price equation is given by equation (4.2):

$$\bar{w}\Delta\bar{z} = Cov(w, z) + E(w\Delta z) \qquad (4.2)$$

But for my purpose I will start, following Okasha (2006), with another formulation of the Price equation by dividing both sides of equation (4.2) by $\bar{w}$. This yields:

$$\Delta\bar{z} = Cov(\omega, z) + E_w(\Delta z) \qquad (4.3)$$

where ω is the relative reproductive output ($\omega = w/\bar{w}$) and $E_w(\Delta z)$ is the reproductive-output-weighted average of the quantity Δz.

As with equation (4.2), there is no heritability term h^2 in (4.3). Starting from equation (4.3) Okasha proposes a formulation of the Price equation that makes h^2 explicit. Okasha proceeds as follows. He starts by noting that the term $E_w(\Delta z)$ in (4.3) can be decomposed as follows (Okasha 2006, 26):

$$E_w(\Delta z) = E(\Delta z) + Cov(\omega, \Delta z) \qquad (4.4)$$

In fact we have $Cov(\omega, \Delta z) = E(\omega \times \Delta z) - E(\omega) \times E(\Delta z)$.

Since:

$$E(\omega) = 1 \text{ and } E(\omega \times \Delta z) = \frac{1}{n}\sum_1^n \frac{w_i}{\bar{w}} \Delta z_i = E_w(\Delta z)$$

we have:

$$Cov(\omega, \Delta z) = E_w(\Delta z) - E(\Delta z)$$

and thus:

$$E_w(\Delta z) = E(\Delta z) + Cov(\omega, \Delta z).$$

Rewriting equation (4.3) with (4.4) we obtain:

$$\Delta \bar{z} = Cov(\omega, z) + E(\Delta z) + Cov(\omega, \Delta z)$$

Noting that $\Delta z = z' - z$, this equation can be simplified. This yields:

$$\Delta \bar{z} = Cov(\omega, z) + E(\Delta z) + Cov(\omega, z') - Cov(\omega, z)$$

$$\Delta \bar{z} = Cov(\omega, z') + E(\Delta z) \qquad (4.5)$$

$Cov\ (\omega, z')$ is the covariance between an entity's relative reproductive output and the average trait value of its offspring.

From equation (4.5), we can now make heritability h^2 apparent. In fact, we know that with the regression line on a graph representing the average offspring values of character z on parental values (or mid-parental in the case of sexual reproduction) of each individual of the initial population, for each point i on the figure we have (see Figure 4.1):

$$z_i' = a + h^2 z_i + e$$

where a is the intercept and e is what remains unexplained by the regression.

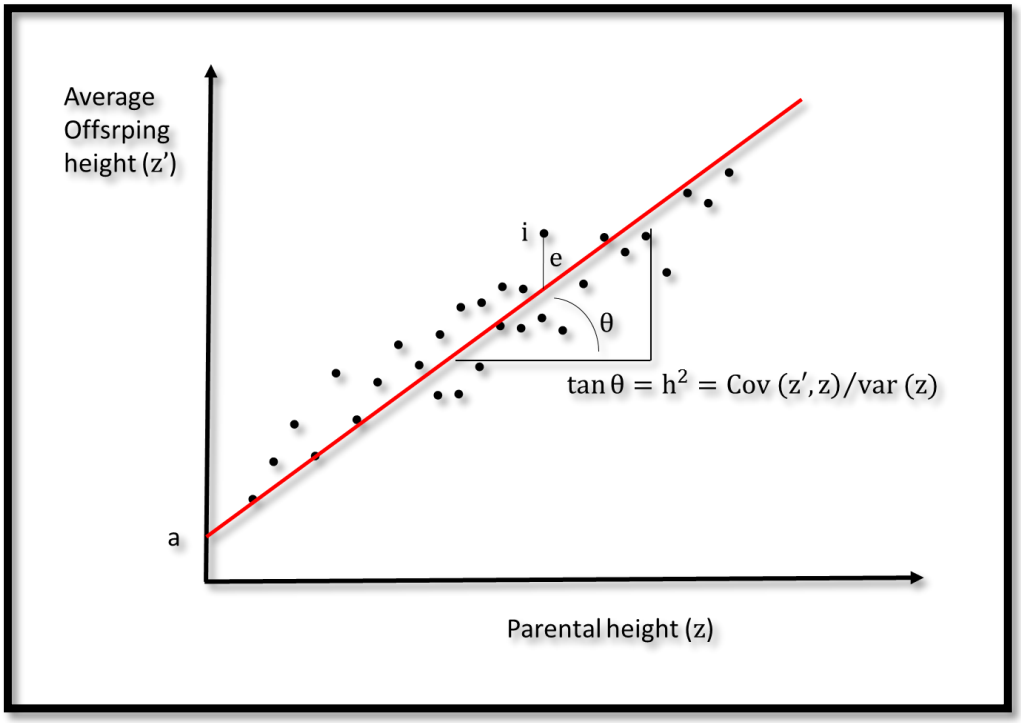


Figure 4.1. Offspring-parent regression for z using least-squares method after Okasha (2006, 35).

Replacing this decomposition of z' in equation (4.5) we obtain:

$$\Delta\bar{z} = Cov(\omega, a + h^2 z + e) + E(\Delta z)$$

$$\Delta\bar{z} = Cov(\omega, a) + h^2 Cov(\omega, z) + Cov(\omega, e) + E(\Delta z) \quad (4.6)$$

Because $Cov(\omega, a) = 0$, since a is a constant and the assumption $Cov(\omega, e) = 0$ is a reasonable one (Okasha 2006, 35) we can simplify this result into:

$$\Delta\bar{z} = h^2 Cov(\omega, z) + E(\Delta z) \quad (4.7)$$

where h^2 is heritability defined as the linear regression of average offspring trait $\mathbf{z}'$ on parental trait z. According to Okasha (2006), the term $h^2 Cov(\omega, z)$ can be interpreted as the change in character z due to selection, and the term $E(\Delta z)$ as the change on character z that would have occurred anyway without selection. This is the interpretation I follow here. As noted by Okasha, there is an obvious link between the term $h^2 Cov(\omega, z)$ and the Breeder's equation ($R = h^2 Cov(\omega, z)$) briefly presented in Chapter 2, since the response to selection (R) is nothing less than some evolutionary change and $Cov(\omega, z)$ corresponds to the selective differential (S). If the term h^2 is nil, then no evolutionary change due to natural selection will result. The term $Cov(\omega, z)$ embeds the two other ingredients of recipes for ENS namely, differences in expected reproductive output (fitness) and variation (Okasha 2006, p. 36), so that in accordance with the recipes if $Cov(\omega, z) = 0$, then no ENS is observed.

Before moving further, one difference between (4.2) and (4.7) that should be noted is that in (4.7) something must be said about Δz in the second term $E(\Delta z)$ in the case a given individual has no offspring (reproductive output nil). In such a case, it is not immediately clear

what value z' should be given in the second term of the RHS. Equation (4.2) is immune to this problem because the second term is weighted by fitness. By convention, in equation (4.7), when an individual does not have any offspring, the value given to the character z' will be the value of the average character of the offspring population ($\overline{z}'$). A possible justification of this convention is that whether an individual has some offspring with, on average, the same value on the character than the average value on this character in the whole offspring population or whether it has no offspring, the evolutionary consequences at the population level are perfectly equivalent, excepted of course that the production of offspring leads to a larger offspring population size. Leaving aside equation (4.7) for the moment, I now turn to what Earnshaw-Whyte (2012) has claimed to be a major flaw of the recipe account of ENS. I will then come back to equation (4.7) to show that a solution to Earnshaw-Whyte's problem naturally lie in it once the regression approach to heritability is generalised.

4.4. A problem for the recipes?

Let us suppose a population of entities of two types **A** and **B** that reproduce synchronically with the peculiarity that each entity whether it is of type **A** or type **B** has a 50% chance to produce an offspring of type **A** and 50% chance to produce an offspring of type **B**. This means there is no heritability of the type as it has been defined so far under the regression approach,[76] that is, the slope of the linear regression of average offspring type on parent type is 0. Let us postulate that both types have the same fertility of 1 offspring every unit of time if they

[76] Some complex inheritance mechanism is supposed to produce this pattern.

survive until reproduction, but that type **A** has on average a viability of 0.7 every unit of time whereas **B** has only a viability of 0.3. This is the model proposed by (Earnshaw-Whyte 2012, 398).

Under an intuitive reading of the classical account, no ENS should be observed because there is no heritability of the type. However, that is not what is observed. Figure 4.2 represents a simulation I have made of this model (Model 1) using the software NETLOGO 5.02 (Wilensky 1999), which, as we have seen in the previous chapter, is classically used for individual-based modelling. At t=0, 5000 individuals (2500 of type **A** and 2500 of type **B**) are created. Then each individual is asked its type. If the type is **A**, the individual has a 70% chance of survival until the next unit of time. If the type is **B**, the chance is only 30%. This represents the *selection* stage of the model. From there, all the surviving individuals are asked to reproduce 1 individual with a 50% chance of producing an individual of type **A** and a 50% chance of producing an individual of type **B**. This represents the *reproduction* stage of the model. Finally, after reproduction, if the population size is larger than 5000, individuals are killed randomly until the population reaches its initial size. I call this stage *check-for-overcrowding*. This last stage represents the only difference with Earnshaw-Whyte's verbal model in which the population size is unlimited. Computational limitations do not allow for simulation with unlimited population size. This has no important consequence for this purpose since unlimited population size is an unrealistic assumption and the large size of the population makes drift (in this particular model probabilistic drift, see Chapter 1, Subsection 1.3.2.2) extremely weak. Once this stage is over, one unit of time is counted and the sequence *selection* → *reproduction* → *check-for-overcrowding* restarts indefinitely for each surviving entity as well as newly created entities, unless the population goes extinct.

The graph on Figure 4.2 shows that, contrary to what is expected, upon a classical interpretation of the recipes for ENS (that is, that the frequency of **A** and **B** remain stable around 0.5), the frequency of type **A** increases up to an equilibrium which is reached after t=7, roughly equal to 0.62 while the frequency of type **B** decreases to 0.38.

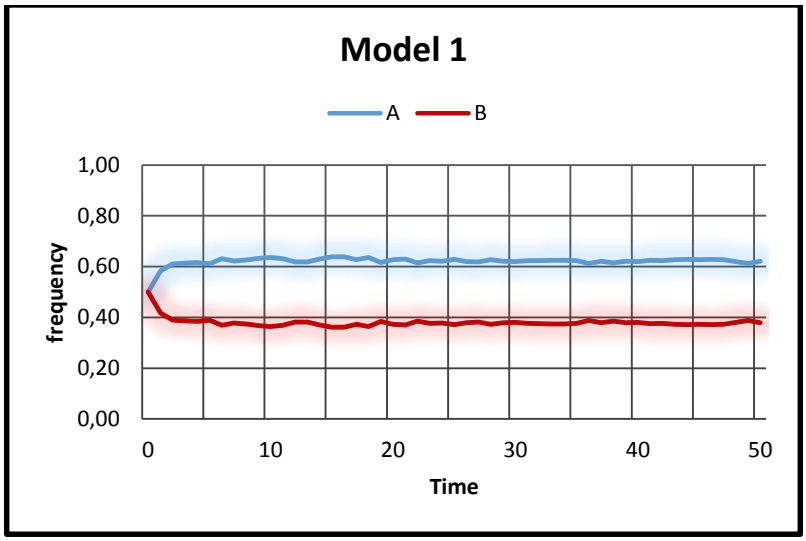


Figure 4.2. Simulation of Model 1 over 50 units of time.

The evolution observed is limited, but we can confidently assume that it is a genuine case of ENS, since the only difference between the two types is a difference in intrinsic-invariable property (here the viability of the entities) within the same environment overall and probabilistic drift is very weak. Since there is no differences extrinsic and intrinsic-variable properties the evolutionary change observed cannot be attributed to deterministic drift. Furthermore, there is no intragenerational mutation. However, there is transgenerational mutation, but it cannot explain the directional evolution observed because the mode of reproduction is the same between the two types which both reproduce with unbiased inheritance. This case, as per my definitions

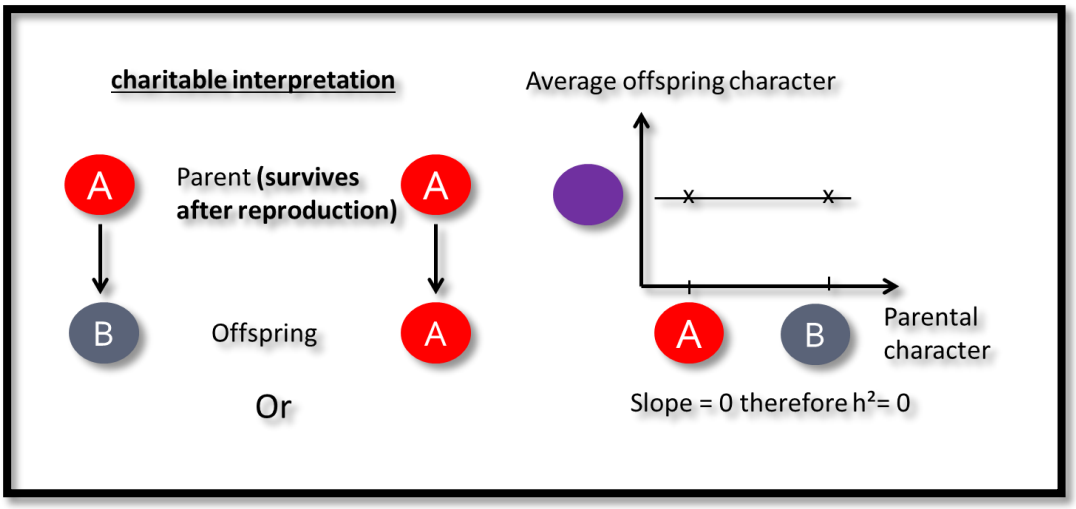
in Chapter 2 is thus not a case of pure ENS, not even one that is close to pure ENS, but the force of transgenerational mutation pushes in the same direction and with the same intensity for the two types so that they cannot explain why type **A** has a higher frequency. Thus, the evolution observed is attributable only to natural selection. Let us recall once more that a textbook definition of evolution in population genetics is a change of frequency of a given variant in the population (e.g., Futuyma 2006) which corresponds to my notion of distributive evolution (see Introduction of the thesis).

This very simple model shows therefore that there can be ENS despite the absence of heritability as classically interpreted under the regression approach. I need to emphasise here that the phenomenon observed in the model does not in and of itself represent an oddity. Indeed, as it will become clearer in the next section, the same phenomenon can be observed in populations of real organisms that do not reproduce such as those discussed in Chapter 2 and a similar effect will be observed in any population with overlapping generations.

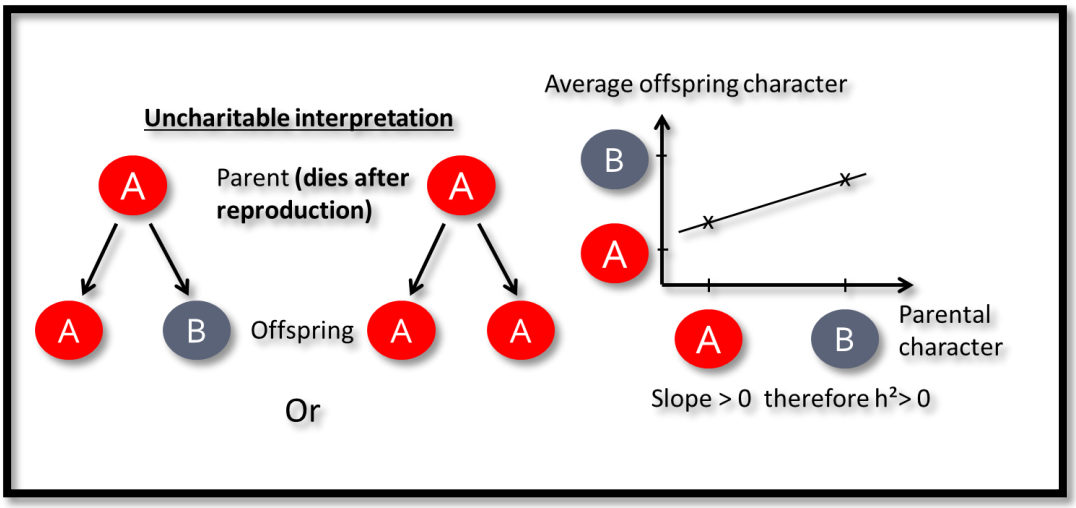
If ENS is observed in this model lacking heritability as classically defined, does this mean that the classical formulation of ENS is flawed? This is what Earnshaw-Whyte claims. In his own words, "th[e] classical account of ENS is flawed and should be abandoned" (2012, 97). In my view his conclusion is far too strong, since as I will demonstrate, it is the *interpretation* of the recipes for ENS given by Earnshaw-Whyte (and many others) that is faulty, not the recipes.

For one thing, Earnshaw-Whyte's model can be formalised in two different ways. These two ways depend entirely on what is understood by 'parental' and 'offspring' entities. Yet, from an evolutionary perspective these two interpretations are perfectly equivalent. We have seen in Chapter 2 that there was no clear cut way to define what an individual is. Here is an example why

individuality and the production of new individuals once they have been defined, should not matter as much as it does currently in evolutionary theory. Under a charitable interpretation of Earnshaw-Whyte's model, every individual produces one offspring if it survives until the reproduction stage (see Figure 4.3*a*). Under such interpretation, there is a clear parent, a clear offspring and no heritability defined as the slope of regression of average offspring character on parental character. Importantly under this interpretation, there is an overlap of generations. Under a less charitable interpretation for Earnshaw-Whyte's agenda, the fertility of one offspring with survival of the parent and subsequent same probability of selection is understood as an entity reproducing asexually such as a bacterium for instance (see Figure 4.3*b*). The parental entity ceases to exist when the two offspring are produced. There is therefore no overlap of generations. From one bacterium of type **A**, we can get either two bacteria of type **A** or one bacterium of type **A** and one bacterium of type **B** with equal probability. With this second interpretation, there is heritability in the population as classically interpreted since from one parent of a given type the two offspring resulting from reproduction are on average more similar to their parent than two individuals chosen randomly at different generations in the population. This demonstrates clearly that the classical concept of ENS is not flawed in light of Earnshaw-Whyte's model. Rather it suggests that the concept of heritability is not as straightforward as typically assumed in the recipes and that the notion of parent and offspring can be problematic for the classical approach if not interpreted carefully.



a.



b.

Figure 4.3. Illustration of Earnshaw-Whyte's model uncharitably interpreted.

A more convincing and realistic model for the point Earnshaw-Whyte is attempting to make would be a case of age-structured populations with individuals of different ages with different viabilities and fertilities (Charlesworth 1994; see also Godfrey-Smith 2007 for a short introduction). Thus, let us suppose, for the sake of argument, that there is an obvious distinction

between parent and offspring such as in parthenogenetic organisms or those that reproduce by budding, as in the case of yeasts, so that we therefore refer to a clear-cut case of a population with overlapping generations and can charitably interpret Earnshaw-Whyte's model. In such cases the notion of heritability used so far does not seem to capture what is indubitably a case of ENS. Yet, I do not think that the tripartite recipe description of ENS should be abandoned even in these cases. Instead, I will argue that in light of the equation (4.7) given earlier, we should reinterpret the concept of heritability in recipes for ENS to make it more rigorous.

4.5. Reinterpreting Heritability for Recipes

The Price equation, as I mentioned in Section 3, predicts the evolutionary changes of a given character over time. One of its important features is that although one generation is usually the pace of time over which evolutionary change is measured, that is not at all necessary. In fact, any arbitrary period of time can be used (Rice 2004, p. 169). This remark is extremely important because it means that evolutionary change can be measured over a period of time smaller than one generation, or even with entities that do not reproduce. Thus when viability selection is the only type of selection present in a population, a *temporal* interpretation of the Price equation can be used (Okasha 2006, 24).

The possibility of a temporal interpretation of the Price equation when the entities do not reproduce immediately raises the question as to what account of heritability should one give in these cases[77] using equation (4.7). Recall that I defined heritability as the linear regression of

[77] This is also obviously true for the term w, but I said earlier that I leave this problem for the next section.

average *offspring* character on *parental* character. But in these cases, we do not have any parents or offspring. We only have the same individuals over time, minus those which, if any, died at a later time. There seems thus to be some inconsistency between the view that the existence of heritability, as classically understood under the regression approach, depends on the existence of generations and the view gained from the Price equation and true by definition, that heritability can be defined without the existence of generations, as displayed in equation (4.7).

At this point, one could either deny that the Price equation provides a valid description of evolutionary change or assert that the concept of heritability used classically is not general enough or inexact. The first option seems unlikely. In fact, not only is the Price equation true by definition, but making the claim that it does not represent a valid description of evolutionary change would fly in the face of some of the most recent developments in evolutionary theory. I will thus adopt in this section the second option and aim at developing a more inclusive notion of heritability that will accommodate Earnshaw-Whyte's model as well as the notion of ENS in populations of non-reproducing entities (Bouchard 2008, 2011; Bourrat 2014; Doolittle 2014, Chapters 2 and 3) and ENS in populations in which overlap of generations is undeniable.

Starting from the Price equation, to predict the change of a given character over time in a population that does not reproduce between t_1 and t_2, we must consider for each individual at t_1 that survives until t_2, that they are their own offspring. Once this is done, the slope of linear regression between the character z of all the surviving entities between t_1 and t_2 obtained becomes positive (provided that individuals retain their character z for their whole life). From an evolutionary perspective, this means that persisting can be seen as formally equivalent to reproducing. A correct interpretation of heritability to be used for recipes for ENS should thus

include a component of persistence. For biological organisms, persistence means survival. But I prefer to use the more general term of persistence to account for cases of ENS in domains other than the biological one.

Up to this point I have only shown that persistence should be taken into account in our definition of heritability in populations that do not reproduce. Yet, some readers might remain unconvinced that a component of persistence should be incorporated to our interpretation of the concept of heritability since they remain sceptical, in spite of the arguments I provided in the previous chapter, that any population of entities that does not reproduce is nothing more than an odd or negligible case of ENS. They might argue that the classical interpretation of heritability as the slope of the regression line of the average offspring on parental character is perfectly valid as long as the entities reproduce. Although, I disagree that non-reproducing populations should be dismissed as negligible or unimportant, to make my account work, I nonetheless need nonetheless to show that persistence can also be an important component of heritability even in populations that do reproduce.

If our goal is to describe accurately the evolutionary change resulting from the reproduction of the members of a population, persistence will only be relevant if there is an overlap of generations. In fact, some parents need to survive after they reproduce if one wants to talk about persistence after reproduction. Interestingly, I have, following Okasha, made so far the assumption of no overlap of generations. But Okasha, following a consensus, clearly states that it is made for simplification purposes only and that the Price equation can be derived even when this assumption is relaxed (2006, 20, footnote 11). Making this assumption along with the assumptions of simultaneous reproduction, asexuality and perfect transmission is a common move made in the philosophy of biology and more generally in evolutionary theory. Yet, one

should be aware of the limitations they impose on the interpretation of the concept of heritability (but also fitness as we will see in the next section) in an evolutionary context. The demonstrations I provide below can be regarded both as an extension of Okasha's work when the assumption of non-overlap of generations is relaxed, as well as an insight on what our concept of heritability should look like once less simplifying assumptions are made about the population. To visualise the problem more clearly, let us turn back to Earnshaw-Whyte's model in which entities do reproduce and in which there is overlap of generations (under the charitable interpretation).

As we have already seen, in Earnshaw-Whyte's model, if heritability is interpreted as a case of non-overlapping generations, the slope of the regression between the parent character and the offspring character is nil. This is because both types produce 50% of offspring of each type. But this leaves out parental entities which have survived one or many events of reproduction and which 1) are still present in the population and 2) have the same chance of survival and reproduction than the offspring entities produced at the previous generation. The fact that the surviving parents and offspring have the same chance of survival and reproduction at the next generation is a peculiarity of Earnshaw-Whyte's model. Different assumptions could be made without invalidating the points I am making (as it is the case in aged structured populations). These surviving entities must therefore be incorporated in the linear regression in the same way that offspring are since, if that is not done, we cannot pretend that the predictions made at the population level are correct. This is because one part of the population is not taken into account after evolution has occurred. Once the surviving individuals are taken into account in our measure of heritability between the population measured before and after reproduction (assuming there is synchronicity in reproduction between all the members of the population and that the surviving individuals do not mutate), the slope of the linear regression becomes positive,

and the correct predictions of evolutionary change between the two times can be made. This shows that the classical interpretation of heritability is insufficient to accurately predict evolutionary change even in populations of reproducing entities.

In practice, in population genetics, in a case of a population with overlap of generations, heritability would be estimated using a population between two times and would thus include parents that survived over time. Thus everything would be done as if the surviving parent were their own offspring from an earlier time. As I argued in Chapter 2, it does not make any difference, evolutionarily speaking, whether an entity produces one offspring or simply persists over time. However if our definition of heritability depends on the concept of reproduction and if this concept is fuzzy or to some extent arbitrary, our notion of heritability should be revised to eliminate these elements of arbitrariness.

Let us sum up what has been argued thus far. ENS can be observed in a number of cases. One of them is described accurately by the intuitive interpretation of heritability in the recipes. This is the case when there is inheritance and variation in fitness in a population of reproducing entities with non-overlapping generations. In this case the classical and intuitive concept of heritability defined as the slope of the linear regression between the average offspring character on the parental one will yield the correct prediction of ENS. I relabel the classical and intuitive concept of heritability as *generational* heritability. In other cases, there is ENS but the intuitive interpretation of recipes either misses that there is ENS at all or makes inexact predictions about the extent to which there is ENS. The intuitive interpretation of the recipes will miss ENS when there is no generational heritability in the population in spite of reproduction[78] but there is

[78] I assume for simplicity that reproduction is synchronic between all the entities of the population.

persistence of the trait in the surviving entities (this is Earnshaw-Whyte's model). It will also miss it when there is no reproduction. Finally it will make inexact predictions when there is both persistence of the parental trait between generations (via surviving parents) and generational heritability in the population. In those cases heritability must be interpreted with a component of persistence instead of only reproduction. I call this component persistence heritability, which I define as the linear regression of average parental character in the offspring generation – or at a later stage if entities do not reproduce – on parental character in the parental generation. Note that persistence and persistence heritability are two different concepts. The first refers to an individual while second one refers to a population in the same way inheritance refers to an individual and (generational) heritability to a population. While the value of generational heritability between two times depends on mechanisms of inheritance, the value of persistence heritability depends on mechanisms preventing the mutation of entities into entities of other types.

To give an example of how inexact predictions can be made by using generational heritability and neglecting persistence heritability, we can imagine a population of two types, **A** and **B** reproducing with perfect inheritance. Under the intuitive interpretation of heritability there is a heritability of 1 on the character 'type' in this population. As in Model 1 presented in Section 3 of this chapter, there is the same difference in viability between **A** (0.7) and **B** (0.3). However in the present case, both **A** and **B**, if they survive after reproduction, mutate with a probability of 0.5.[79] If that happens, the entity changes its type (that is, becomes **A** if it was **B** and **B** if it was **A**) and takes all the properties of its new type. The prevalence value of 0.5 for the mutations is

[79] This case is thus as with Earnshaw-Whyte model (Model 1) not a pure case of ENS, but in this case too if any directional evolutionary change is observed, it will not be attributable to mutation because the pattern of mutation is the same for both type.

deliberately unrealistic, since my goal here is only to make a general point on the role of persistence for ENS, not to match a specific target system.

An intuitive interpretation of recipes for ENS predicts that because there is perfect heritability, the type **A** should invade the population. However that is not what is observed if this model is run. Figure 4.4 represents a simulation of this model (Model 2).[80] When compared to model 1, there are two differences. First, the transmission of types between parent and offspring is perfect. Second, a *mutation* stage is added just after the reproduction stage. During that stage each parental entity has a 50% chance to change its type and thus become a member of the other type, taking all the properties of its new type.[81] Thus the sequence undergone by each entity every unit of time is the following: *selection* → *reproduction* → *mutation* → *check-for-overcrowding*. At t=0, 2500 individuals of each type are created.

[80] NETLOGO 5.02 has been used to run this simulation.
[81] Importantly newly produced offspring are not concerned by the mutation stage. They will only be concerned by it if they survive the next selection stage.

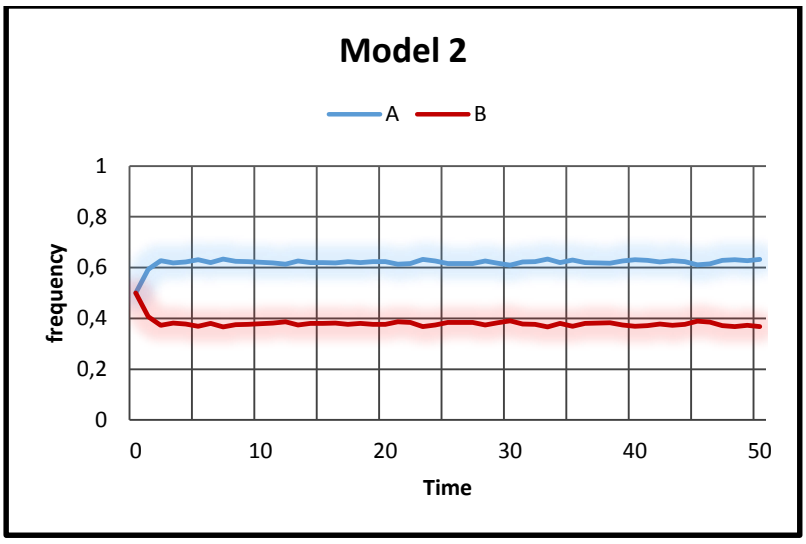


Figure 4.4. Simulation of Model 2 over 50 units of time with a mutation rate of 0.5.

As can be seen on the graph in Figure 4.4, the prediction made by the intuitive interpretation of the recipes that relies on generational heritability is wrong simply because type **A** does not invade the population. In fact, the pattern of evolution observed is strictly the same as the one observed in model 1. Why is that the case? Because persistence heritability and generational heritability play the same role for ENS. Simply put, in Model 1 persistence heritability is maximal while generational heritability is nil. In Model 2 the contrary holds: persistence heritability is nil while generational heritability is maximal.

This demonstrates that a correct interpretation of the concept of heritability for measuring ENS must take into account the transmission over time of a trait in a given population without discrimination of the origin of the transmission, that is, either coming from persistence or/and from reproduction. I propose to label this concept: temporal heritability. Thus (temporal) heritability should not be defined as "the linear regression of average offspring character z' on

parental character z" but as "the linear regression of the average character *in the offspring population*[82] z' on the character in the parental population z" or again as "the linear regression of the average character z' at time t2 subsequent to t1 on the character z at time t1." In order to formally ensure this claim and anchor it in the existing literature, I finish this section by providing a formulation of the Price equation in which I decompose h^2 presented in equation (4.7) in two components that makes the links between temporal, generational and persistence heritability apparent.

We start with the form (4.7) of the Price equation given by Okasha (2006, p. 35)[83] and presented in Section 4.2:

$$\Delta\bar{z} = h^2 \, Cov\,(\omega_i, z_i) + E(\Delta z_i) \qquad (4.7)$$

h^2 being defined as the regression slope of parental character z' on average offspring character z' it follows that:

$$h^2 = \frac{cov\,(z_i', z_i)}{Var(z_i)} \qquad \text{(Okasha 2006, 34)}$$

For Okasha, h^2 is defined as generational heritability, that is, "the linear regression of average offspring character z' on parental character z". As I have argued, if there is overlap of generations or no reproduction, this represents an error of interpretation of what heritability means in the context of the Price equation. Rather, the definition to be used should correspond to temporal heritability, that is, "the linear regression of the average character *in the offspring population* z' on the character in the parental population z". At generation *n+1*, the offspring population is

[82] Which refers to both the potential surviving parental entities and the potential offspring of parental entities.
[83] For the purpose of the demonstration I have added the indices which were dropped in Section 2.

composed of both the offspring of the parent(s) of the generation *n* and the surviving individual(s) (if any) of the generation *n* at the generation *n+1*.

We can decompose z' which represents the average character value for each individual i in the offspring population as follows:

$$z'_i = \frac{(1-m_i)z_i + r_i z'_{i_o}}{1-m_i+r_i} \qquad (4.8)$$

$(1-m_i)z_i$ is the part of the average offspring-population character value of individual i due to the possible survival of this individual, where $1-m_i$ is the persistence between the two times over which evolutionary change is measured and m_i the mortality.[84] $r_i z'_{i_o}$ is the part of the average offspring-population character value of individual i due to the offspring produced by i, where r_i is the number of offspring produced by i and z_{i_o} the average character value of the offspring.[85] If an individual i has no offspring and does not survives between the two times ($1-m_i+r_i=0$), by convention the value of the character z' will be the value of the average character of the offspring population ($\bar{z}'$) comprising both the surviving individual of the parental population and/or the offspring produced by the parental population.[86] The justification of this convention is the same as the one given at the end of Section 3.2: whether the contribution of an individual at a later time on character z' is nil or amounts exactly to the average value of character at the (offspring) population level, does not make any evolutionary difference.

[84] The value of the character is assumed here to remain constant in one given individual, but that is not a necessary assumption.
[85] Reproduction is supposed to be asexual.
[86] The same convention will be used in equation (4.7) for the term $E(\Delta z)$.

If we replace (4.8) in the formal definition of heritability given above, dropping the indices i for convenience, we find that:

$$h^2 = \frac{cov\left(\frac{(1-m)z + rz'_o}{1-m+r}, z\right)}{Var(z)} = \frac{cov\left(\frac{(1-m)z}{1-m+r}, z\right) + cov\left(\frac{rz'_o}{1-m+r}, z\right)}{Var(z)} = \frac{cov\left(\frac{(1-m)z}{1-m+r}, z\right)}{Var(z)} + \frac{cov\left(\frac{rz'_o}{1-m+r}, z\right)}{Var(z)} \quad (4.9)$$

The term $\frac{cov\left(\frac{(1-m)z}{1-m+r}, z\right)}{Var(z)}$ corresponds to the regression coefficient of the parental character in the offspring population $(1-m)z$ on parental character z in the parental population, which is the definition of persistence heritability ($h_p{}^2$). The term $\frac{cov\left(\frac{rz'_o}{1-m+r}, z\right)}{Var(z)}$ corresponds to the regression coefficient of the average offspring character z'_o in the offspring population on parental character z, which is the definition of generational heritability ($h_g{}^2$).

Each of these two coefficients can be nil. When we replace h^2 by the formula given by (4.9) in (4.7), we get:

$$\Delta\bar{z} = \left(\frac{cov\left(\frac{(1-m)z}{1-m+r}, z\right)}{Var(z)} + \frac{cov\left(\frac{rz'_o}{1-m+r}, z\right)}{Var(z)} \right) Cov\,(\omega, z) + E(\Delta z)$$

$$= (h_p{}^2 + h_g{}^2)\, Cov\,(\omega, z) + E(\Delta z) \quad (4.10)$$

This is a formulation of the Price equation separating h^2 into generational and persistence heritability.

In a case of discrete generations which is the case proposed by Okasha (2006), we have $m = 1$. In that case equation (4.10) becomes:

$$\overline{\Delta z} = \left(\frac{cov\left(\frac{(1-1)z}{1-1+r}, z\right)}{Var(z)} + \frac{cov\left(\frac{rz'_o}{1-1+r}, z\right)}{Var(z)} \right) Cov\,(\omega, z)\; + E(\Delta z)$$

which once simplified is:

$$\overline{\Delta z} = \frac{cov\left(\frac{rz'_o}{r}, z\right)}{Var(z)}\; Cov\,(\omega, z) + E(\Delta z)$$

$$= {h_g}^2\, Cov\,(\omega, z) + E(\Delta z)$$

which shows that there is no persistence heritability in this case and thus that all the heritability is generational heritability.

The demonstration provided here shows how generational heritability and persistence heritability are commensurable. By 'commensurable', I mean here that it is meaningful to add them up and thus that they can play the same role evolutionarily speaking. $E(\Delta z)$ is still interpreted as the change on character z that would have occurred without selection, but this change can occur both by a transmission bias of the character from parent to offspring or a transmission bias of the character from one and the same individual over time by some form of mutation.

4.6. Interpreting w in the Price equation: Persistence and Developmental Growth

The possibility of a temporal interpretation of the Price equation does not challenge solely the concept of heritability made apparent in equation (4.7) by Okasha. It also challenges the classical interpretation of the term w (whether in equation (4.2) or equation (4.7)) which as we have seen is often called 'fitness' but that I preferred to call 'reproductive output' in the context of an asexual population reproducing discretely. May I reiterate that I made this choice because any difference in w between two entities of the population could be due in the Price equation either to differences in intrinsic-variable, extrinsic or intrinsic-invariable properties and thus should not necessarily be associated with natural selection. Yet, if $\Delta\bar{z}$ is measured in a population in which there is no reproduction or within a generation, as with heritability, that w cannot be the reproductive output of entities. In fact, if one makes this assumption, they will find that for each entity of the population – the reproductive output is invariably 0. A covariance with a constant being 0 and the product of 0 with another number being 0, we should predict, according to equation (4.7), that there is no evolutionary change in those populations or over a period of time shorter than one generation. But this would be obviously wrong since the frequency of the different types of a population would change over time. This means that differential persistence of entities will lead to a change in character ($\bar{z}$) at the population level as long as there is variation in this character, even when the population inexorably goes to extinction as argued in chapters 2 and 3.

Another problem arises when we relax the assumption of discrete generations. As I argued with heritability, when individuals survive and are present in the offspring generation, they should be counted as part of the offspring population. By considering that w is only the reproductive output of the entities of the parental generation, we will make wrong predictions about the overall evolutionary change of the population in a number of cases. This is especially

important if the difference between the entities, although biologically insignificant such as the one presented in Chapter 2 (that is, different degrees of persistence of a tiller over time, different level of production of glue etc...), lead to drastic differences as to what counts as growth and persistence on the one hand, and reproduction on the other hand.

Thus, how shall we interpret w in a general way that takes into account these cases? One way to do that is to assume that w does not represent the reproductive output of entities, but more generally the absolute growth they are causally connected to over a certain period of time. Because in evolutionary theory we are interested in the evolution of trait at the population level, this means that in spite of an *absolute growth* rate inferior to 1, the *relative growth* rate at the population level could still be positive. There are three ways under which relative growth of a type can increase. The first one is naturally by reproducing more offspring of the same type than the other types. This corresponds to the version of the price equation presented in with equation (4.2) in which there are discrete generations and thus no overlap of generations or lack of reproduction. Another way is by persisting longer than the other types. Finally a third way is by growing faster than the other types.

I am now going to develop a general version of the Price equation that takes those different modes of growth into account. From there I show how by making the assumption of discrete generations one can go back to equation (4.2). Let us start by decomposing w_i using the three components of growth mentioned earlier. We have w_i that depends on: r_i, the reproductive output produced by an entity between the two times during which evolutionary change is measured; m_i, the mortality over this same period; and finally g_i, the average developmental growth of i during the time over which $\Delta\bar{z}$ is calculated (we assume any offspring produced

cannot grow between the two times of the measure of evolutionary change). m_i is a real number with the value 0 or 1; r_i is a real number with a value between 0 and infinity; finally g_i is a real number between -1 and infinity where -1 corresponds to the limit case in which developmental growth is negative and leads to the death of the entity. We can now rewrite w_i once defined as growth as follows:

$$w_i = (1 - m_i) \times (1 + g_i) \times (1 + r_i),$$

which once developed gives:

$$w_i = 1 - m_i + g_i + r_i - m_i g_i - m_i r_i + g_i r_i - m_i g_i r_i \qquad (4.11)^{87}$$

Supposing that the time interval during which the evolutionary change is measured is small enough so that m_i, g_i and r_i are small, we can neglect the four last terms of the RHS of equation (4.11). We thus obtain:

$$w_i = 1 - m_i + g_i + r_i$$

We can now rewrite equation (4.2) by replacing w_i by its formulation in equation (4.11). We start with:

$$\bar{w}\Delta\bar{z} = Cov(w_i, z_i) + E(w_i \Delta z_i) \qquad (4.2)^{88}$$

[87] Thanks to Arnaud Pocheville for insightful discussions on this point.
[88] Indices have been added.

which becomes:

$$\bar{w}\Delta\bar{z} = Cov(1 - m_i + g_i + r_i) \; + \; E((1 - m_i + g_i + r_i) \times \Delta z_i) \qquad (4.12)$$

where $\bar{w}$ is defined as:

$$\bar{w} = \frac{1}{n}\sum_{i=1}^{n}(1 - m_i + g_i + r_i)$$

If we develop (4.12), we obtain:

$$\bar{w}\Delta\bar{z} = Cov(1 - m_i, z_i) + Cov(g_i, z_i) + \; Cov(r_i, z_i) + E((1 - m_i)\Delta z_i) +$$

$$E(g_i\Delta z_i) + \; E(r_i\Delta z_i) \qquad (4.13)$$

The six terms of the RHS in Equation (4.13) can seem hard to grasp intuitively, but they find a quite natural interpretation. The first term of the RHS, $Cov(1 - m_i, z_i)$, represents the evolutionary change due to selection corresponding to different persistence between the entities of the parental population. The second term of the RHS, $Cov(g_i, z_i)$, represents the evolutionary change due to selection corresponding to different developmental growth rates between the entities of the parental population surviving during the period of time over which evolutionary change is measured. The third term of the RHS, $Cov(r_i, z_i)$, represents the evolutionary change due to selection corresponding to different reproductive outputs between the entities of the parental population. The fourth term of the RHS, $E((1 - m_i)\Delta z_i)$, represents the evolutionary change due to the bias of transmission if an entity was surviving parental entities between the parental and the offspring population. It can be seen as some form of 'somatic' mutation. The fifth term on the RHS, $E(g_i\Delta z_i)$, represents the evolutionary change due to the bias of

transmission of the developmental parts of the surviving parental entities between the parental and the offspring population. It can also be seen as a form of somatic mutation on the growing parts of the entities. Finally, the sixth term on the RHS, $E(r_i \Delta z_i)$, represents the bias of transmission between parent and offspring. If transmission is perfect, this term is nil.

Equation (4.13) represents a general formulation of the Price equation that can account for the cases of lack of reproduction, differential growth, or non-overlap of generations in a population. If we now assume discrete generations and no developmental growth as we did earlier in this chapter, we have $m_i = 1$ and $g_i = 0$. Thus equation (4.13) becomes:

$$\bar{w}\Delta\bar{z} = Cov(1-1, z_i) + Cov(0, z_i) + Cov(r_i, z_i) + E((1-1)\Delta z_i) + E(0\Delta z_i) + E(r_i \Delta z_i),$$

which once simplified is:

$$\bar{w}\Delta\bar{z} = Cov(r_i, z_i) + E(r_i \Delta z_i).$$

We find the same form as equation (4.2), which shows that equation (4.2) is a special case of equation (4.13).

4.7. Conclusion of the chapter: why should we care?

In this chapter, I have presented a different, less intuitive concept of heritability and vindicated formally two other proxies of fitness different from reproductive output, namely persistence and developmental growth, as of potential evolutionary importance. With respect to heritability, both persistence heritability and generational heritability (that is, the intuitive

conception of heritability) are entailed in the concept I have termed temporal heritability. I have argued that temporal heritability is the right interpretation of heritability to use when talking the most generally about ENS and thus the one to be used both in the classical approach and the formal approach to ENS. If only generational heritability is used, this leads to some inconsistencies in the Price equation once the assumptions of reproduction and discrete generations are relaxed. Similarly, I have shown that with respect to w one should not consider it as representing only the reproductive output of entities but more generally growth which entails reproductive output, persistence, and developmental growth. If it is interpreted as reproductive output only, it will lead to wrong predictions of evolutionary change once the same assumptions of reproduction and overlap of generations are relaxed. The different arguments I have provided here are in perfect adequation with the three previous chapters, and because they rely on more formal arguments than in those chapters, strengthen the case that evolutionary theory can be dispensed from reproduction without falling apart.

One might object at this point, that the benefits of interpreting heritability as temporal heritability instead of generational heritability and w as growth instead of reproductive output, do not out-weigh the costs of doing so on the grounds that these difference are not important in most cases of ENS. I disagree. As I have argued in the previous chapter, it is true that persistence alone does not straightforwardly allow the same kind of complex adaptations observed in Darwinian populations in which there is multiplication. This point seems intuitively obvious and I provided a detailed explanation of why in the previous chapter (see also Bourrat 2014). But reproductive output is not the only way a population can multiply and the number of entities present in the population is sometimes not particularly important for the possibility of evolution and ENS. In fact, with a high degree of developmental growth and a high degree of persistence

only a few individuals would suffice for the possibility of multiplication. And for complex adaptation to occur it would only require that each part of an entity is modular enough to be able to mutate independently. This is precisely what is observed in the case of the plant mentioned earlier in Chapter 2 and described in more details in Bouchard (2004, 2008, 2011) or Godfrey-Smith (2009b).

Evolutionary biology and philosophy of biology have for too long focused solely on organisms that resemble us and marginalised cases that do not fit them (O'Malley & Dupré 2007). But there are many species for which a persistence account of heritability and/or a persistence/growth account of fitness can be given. This is usually not done because, although in practice the concepts I defined as temporal heritability and fitness as growth are the one used in evolutionary theory (the formal equations are exact), everything is done *as if* the concepts of generational heritability and reproductive output were the one used. In fact, the existence in botany of the ramets and the genet (Harper 1977) is only one of the many demonstrations of the inadequacy of fitness as defined in terms of reproductive outputs only. Finally, Earnshaw-Whyte and to some extent Okasha are not the only ones to use a generational interpretation of the concept of heritability when they should use a temporal one. I have found at least four other instances of this in the literature I have already cited namely: Gillespie (2004), Godfrey-Smith (2009b), M. Hamilton (2009) and Jacquard (1983).

5. Natural Selection and the Reference Grain Problem

5.1. Introduction

Chapters 1 and 2 revealed that the philosophical literature on natural selection is very dense and often entangled with issues surrounding the concept of fitness. We saw that most of this literature revolves around two related topics. The first one amounts to establishing whether natural selection can reasonably be considered as a 'force' or cause of evolutionary change. The second one amounts to distinguishing natural selection from drift. I noted that there is much disagreement over these two questions. To remind the reader, some authors argue that natural selection is a force or cause of evolutionary change (e.g., Bouchard & Rosenberg 2004; Millstein 2006; Reisman & Forber 2005; Rosenberg & Bouchard 2005; Shapiro & Sober 2007; Sober 1984; Stephens 2004) and that natural selection and drift are distinct causal processes, while the statisticalists disagree (e.g., Matthen & Ariew 2002, 2009; Walsh 2007, 2010; Walsh, et al. 2002) and argue that natural selection is fundamentally a statistical process and that drift and natural selection should not be separated.[89]

I provided my view in detail on these issues and will not discuss them further in this chapter. Suffice to say that to be consistent with the view I have developed so far and more particularly in the two first chapters, I distinguish natural selection from other evolutionary processes that might occur in a population of entities, namely mutation and drift. I consider all

[89] Some also argue that drift is an error term (e.g. Brandon 2005) or that it is a useful fiction (e.g., Rosenberg 1994).

three phenomena to be different causal processes which are responsible for the differential persistence and production of types (either through developmental growth or reproduction) in a population. I consider natural selection to be conceptually separate from the other two as I have shown in Chapter 1 with drift, and Chapter 2 with mutation. For those reasons and other reasons provided in the Introduction of the thesis, I do not equate natural selection with evolution as it has been proposed by some statisticalists (e.g., Matthen & Ariew 2002).

In this chapter, I turn to a different sort of problem associated with the concept of natural selection, which as far as I am aware, has hardly been noticed in the literature. A striking fact is that among the protagonists of the debate over whether natural selection is a force or cause and whether it can be distinguished from other evolutionary processes, all seem to accept that natural selection is a process independent from the grain of description at which it is described. In this Chapter I challenge this assumption. I show that because natural selection is associated with the notion of type-identity, then according to the grain of description one uses, one should typically expect different answers as to whether natural selection occurs in a population. I call this problem the 'reference grain problem' and compare it to the reference class problem in probability theory (Hájek 2007). Although I do not provide a complete and definitive solution to this problem, which seems to be *prima facie* insoluble, I suggest one way it could be addressed.

To unfold the reference grain problem, the remainder of the chapter is divided into three sections. In Section 5.2, drawing on the ideas developed in Section 2.3, I provide a simple *operational* definition of fitness applicable to populations of entities reproducing asexually and without overlap of generations. I argue that the operational concept of fitness relevant to evaluate the evolutionary change due to natural selection as opposed to other causes of evolutionary change, crucially depends on the notions of type. In Section 5.3, I expose the reference grain

problem. I show that because there are an indefinite number of types an entity can be categorised in by merely changing the grain of description, this leads to some problems for the ontological status of natural selection. I draw the similarity with the reference class problem in probabilities and show that the origin of the two problems is the same and I argue that the reference grain problem, if it might not be only found in evolutionary theory, takes a particular form in it. Finally, in Section 5.4, I briefly propose one way in which the problem could be addressed. I suggest that natural selection could be considered as a model of a process rather than a physical process and that this model might bear some similarities with its target system (a population) at a particular grain of description.

5.2. Operationalising fitness: evaluating differences in fitness associated with natural selection

We saw in Section 2.4 that there are many different definitions of natural selection in the literature, some of which include reproduction and inheritance, others which do not. I argued that a definition consistent with the idea that natural selection is a general principle should mention neither reproduction nor inheritance, especially since we saw in Chapter 4, using the Price equation, that reproduction with inheritance, developmental growth and persistence without mutation were formally equivalent. I also argued in Section 2.3 that when transmission is imperfect during an event of reproduction, the evolutionary change due to imperfect transmission should not be associated with natural selection but with an evolutionary process conceptually distinct from it (I suggested a non-standard notion of mutation). Although it is correct to claim, in such case, that ENS can occur even when there is imperfect inheritance

between parents and offspring, the evolutionary change resulting or the lack of such change, I argued that it results in a mixed case of ENS and evolution due to another evolutionary process.

With these preliminary remarks in mind, let us restate the principle of natural selection (PNS) I defined in Chapter 1:

> **(PNS)** The process of natural selection will occur in a population if and only if there is at least one difference in intrinsic-invariable property, in an environment **E**, between two or more individual entities of the population that can be causally linked to a downstream difference in reproductive output (or more generally relative growth).

The PNS leads to a view of fitness as a relational property that measures a difference in intrinsic-invariable properties between two or more entities. This is quite different from the way fitness is routinely *measured*. In fact, in many cases fitness is assessed via the average reproductive output after one generation. Yet, for well-known reasons reproductive output after one generation is problematic (see for instance Ariew & Lewontin 2004; Beatty & Finsen 1989; Godfrey-Smith 2009b, 23) and thus cannot be considered as a definition of fitness. Two major criticisms are the following. Sometimes, reproduction after one generation is insufficient to understand the dynamic of the population (see Chapter 6; Godfrey-Smith 2009b, 26). Second, there seems to be no one single parameter capturing the concept of fitness.

One advantage of the PNS is that it can be applied to any temporal or physical scale and any intrinsic-invariable property or set of intrinsic-invariable properties can serve as a basis to define fitness. Second, the PNS does not assume the concept of fitness to be only about difference in reproductive output. A large part of the three previous chapters has been dedicated

to substantiate that claim. I have shown that reproductive output is only one possible end-product of such difference and that persistence and developmental growth are also legitimate end products of differences in intrinsic-invariable properties. They should therefore be attributed to natural selection (Chapter 2). This view has been reinforced by the demonstration that formally, reproductive output is perfectly dispensable from the concept of fitness (Chapter 4). Finally, when a diachronic perspective is used and reproduction is 'endogenised', reproduction appears as an evolved product of natural selection rather than a condition for it to occur (Chapter 3).

If the reproductive output[90] of an entity does not constitute its fitness, under some particular assumptions there may nevertheless be a direct link between the two. The fitness of an entity for one particular intrinsic-invariable trait, over more than one generation is the difference in this property that *leads* to differences in reproductive output between this entity and (at least) another one being compared (or the average value of the population) within in given environment **E**.[91] I assume here a population reproducing asexually in discrete generations, in which there is no interaction between the entities of the population. The fact that a difference in reproductive output between two entities of different types is only *a downstream consequence of* differences in fitness, rather than differences in fitness *being* differences in reproductive output, constrains the circumstances under which reproductive output represents an accurate proxy for fitness (this will also be a major topic of Chapter 6).

We explored the reason why this is the case in Chapter 1, but to make it clear, I propose the following summary. In a deterministic setup, if differences in reproductive outputs are only

[90] I give here an analysis for the case of differences in reproductive outputs, but the same could be said, *mutatis mutandis*, with differences in developmental growth and persistence.
[91] Assuming there is no interaction between the entities of the population.

consequences of difference in fitness, this means that confronted by two different probability distributions of reproductive output in a deterministic setting for two given types of entities say **A** and **B** (see Figure 5.1), there are at least two generic non-mutually-exclusive causal interpretations leading to these probability distributions. I assume here that one does not know whether these two types are or have been overall in the same environmental conditions (that is, that there is no difference in extrinsic and no difference in intrinsic-variable properties between the entities of the two types). The first interpretation is that the two types are in different environmental conditions and these differences solely led to the differences in reproductive outputs observed. In this case, we should not assume that the evolutionary change observed is due to natural selection in spite of a difference in reproductive output, but rather, as I argued in Chapter 1, to a form of drift I called deterministic drift. The second one is that some difference in intrinsic-invariable properties between the two entities is causally responsible for the differences in distribution of reproductive outputs observed. In that case, we can assume that the difference in reproductive output observed results from natural selection. Only in this case is the relation between reproductive output and natural selection straightforward. This is because the number of offspring produced by one entity, once relativised over the average number of offspring produced by the population, is directly proportionate, in this case, to the difference in intrinsic-invariable properties between this entity and the average value in the population.

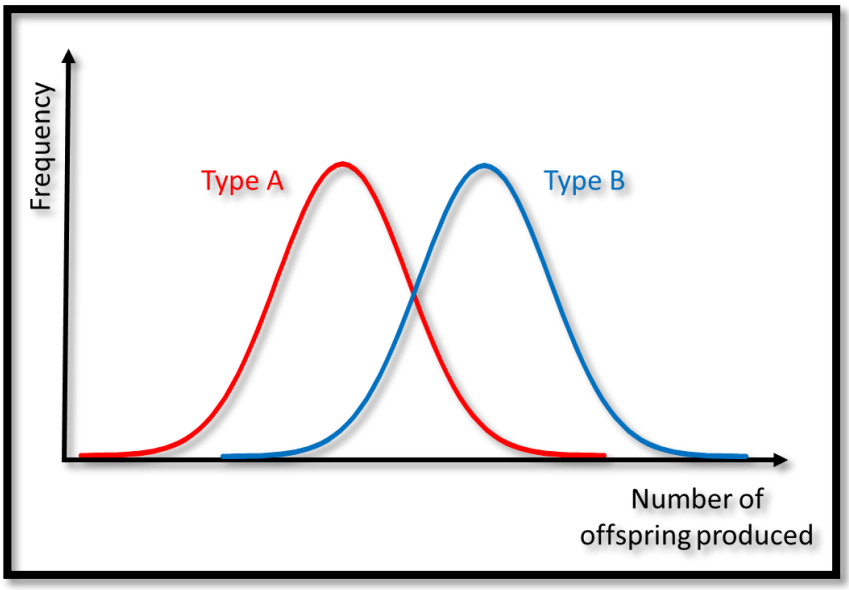


Figure 5.1. Two hypothetical probability distributions for two types of entities in a deterministic setup which can be the product of two causal processes: natural selection and deterministic drift.

So far so good, but there is a complication with this picture. Let us suppose that entities do not reproduce perfectly so that type **A** and **B**, while they produce some offspring of their type, also produce some offspring of a different type, say type **D**. Let us assume that **A** and **B** produce offspring of type **D** with a different rate in a given environment supposed constant over time. For type **A**, for each four offspring produced, three are of type **A** and one of type **D**. For type **B** two are of type **B** and two are of type **D** (see Figure 5.2). Is the fitness of type **A** and type **B** the same? If not, how should we infer the fitness after one generation of these two types from their reproductive outputs?

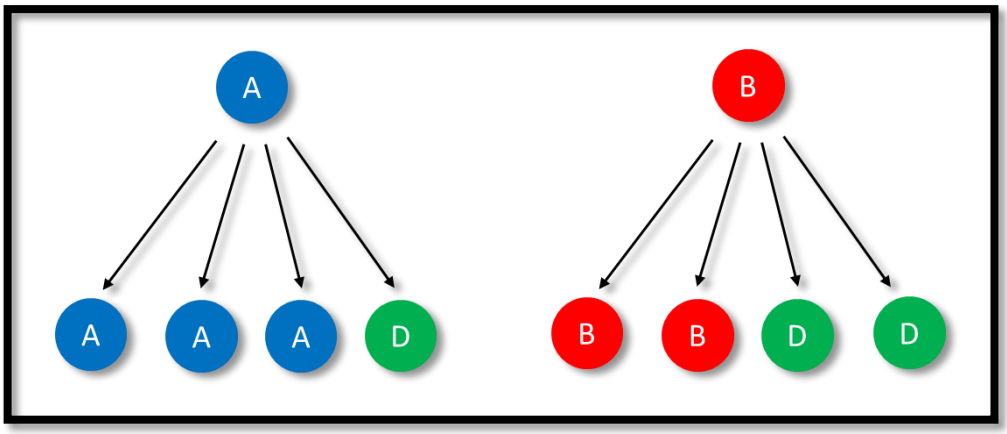


Figure 5.2. Two types of entity with different transmission bias.

In this example, if fitness was inferred before the production of offspring for these two types, say by using Bouchard and Rosenberg's (2004) definition of fitness as solution to design problems, we would not be able to tell which one of the two types is fitter. Both types seem to solve the solution posed by their environment to the same extent and this results in them producing the same number of offspring. However, type **A** transmits its type over generations with a higher frequency than type **B**. Does it matter? In Chapter 2, I argued that when offspring are of a different type than their parent, then the resulting evolutionary change should be associated with another evolutionary process distinct from natural selection (I suggest a non-standard notion of mutation). If the concept of fitness is intimately related to natural selection, then *intergenerational* fitness should be inferred from the difference in offspring of the same type as their parents, and not the overall number of offspring produced, which represents only an accurate proxy of *intragenerational* fitness.

A measure of intragenerational fitness is insufficient when one wants to predict evolutionary change over more than one generation because it does not take into account the effect of natural selection over the whole 'life cycle' of the entities. Thus, if one wants to know

whether the intergenerational change in frequency on a particular trait is the result of natural selection after more than one generation, one cannot simply measure the reproductive output of the entities of the population without knowing anything about the inheritance system of the entities. This is because the inheritance system will determine the pattern of transmission of the type across generations (or more generally over time assuming there are not intragenerational events of mutation). Thus, missing the difference between intra and intergenerational fitness would lead to wrong inferences about the role of natural selection in the evolutionary dynamic.

For example, let us suppose that type **D** does really badly in the environment and always dies briefly after having been produced. By measuring only the reproductive output of **A** and **B,** one would miss that fact and would believe both type have equal intergenerational fitness when in fact **A** is fitter than **B**. In fact, after many generations, the type **A** will be predominant in the population only because **A** is able to produce more offspring of its type due to some differences in intrinsic-invariable properties. It is thus important to measure offspring of the *same type* as the parent if one wants to link reproductive output to natural selection periods of time equal to or longer than one generation.

Although important for some of the systems used in this thesis, this consideration does not really matter when two entities with the same transmission bias are compared, such as for example in cases of sexual reproduction with fair meiosis. To see why, suppose the case as depicted in Figure 5.3 where three asexual types **A**, **B** and **C** reproducing in discrete generations, produce respectively 2, 4 and 6 offspring of their type and type **D** with the same bias: half of the time they produce offspring of type **D**, while the other half of the time, they produce offspring of their own type. In this case, whether one takes the total number of offspring or only the offspring of their type, the relative difference in the production of offspring of their type will be

the same. The fitness of type **B** is twice the fitness of type **A** and the fitness of type **C** is three times higher than the fitness of **A**.

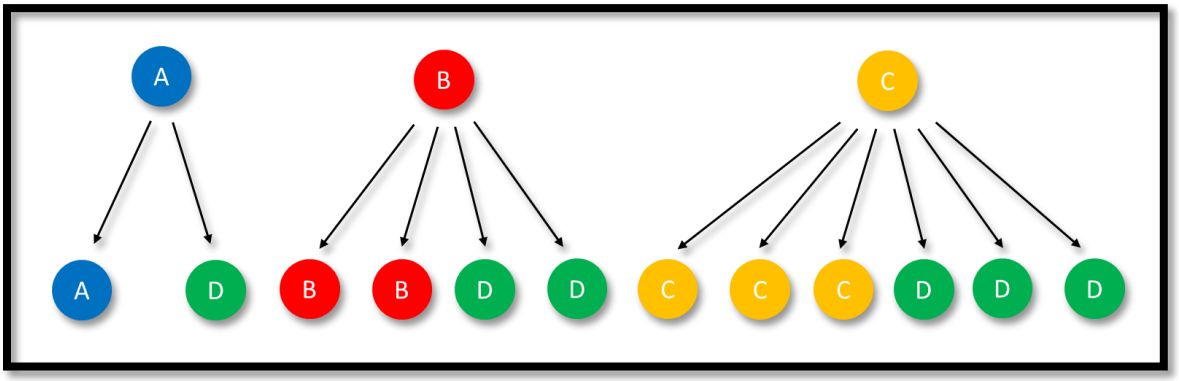


Figure 5.3. Three types of entity with the same transmission bias.

For this reason, in the case of sexual organisms with fair meiosis, assessing intergenerational fitness with reproductive output, does not constitute a problem in spite of the fact that their type are transmitted to their offspring only half of time during reproduction. However, had the meiosis of sexual organisms been unfair for some of them, this estimation would be inadequate for assessing intergenerational fitness and ultimately assessing the strength of natural selection in a population. This provides arguments against the view that reproductive output is an appropriate measure of fitness in cases where the inheritance system is not stable and/or not the same for all the members of the population, such as in the cases proposed in chapters 2 and 3. A classic example in which the inheritance system is not the same for all the members of the population is meiotic drive, in which the transmission of one allele is biased (see Burt & Trivers 2006 for more on the concept of meiotic drive). In this case, the classical interpretation is that selection, instead of occurring at the level of the individual organism (as in

the case of fair meiosis), occurs at the level of the gene (Okasha 2006). For reasons that will become apparent at the end of this chapter and Chapter 6, I do not endorse this claim. Suffice to say for now that the ontological interpretation of levels of selection is a very confused matter.

With these distinctions made, as well as those made in Chapter 2, it will becomes very clear in this chapter that the notions of fitness and natural selection once linked to total evolutionary change, crucially depend on the notion of type. Thus far, I have assumed that the notion of type is unproblematic in the context of evolutionary theory and there is, as far as I am aware, no dispute in the philosophy of biology on this matter. However, I will show now that this is far from being the case.

Intuitively a type is defined as a set of entities that share at least one identical property which is intrinsic and invariable over time and usually gives the label of the type. For instance all the red objects that remain red over time belong to the type 'red'. 'Red' is thus an intrinsic-invariable property (at least as long as the object remains red). When there are no differences in intrinsic-variable and extrinsic properties leading to differences in reproductive output between the entities of a population the fitness of an entity of this population can straightforwardly be assessed by counting the number of offspring or descendent of its type. Yet, a practical complication arises when one realises that belonging to a particular type presupposes that one has chosen a particular *grain of description*, that is, the amount of zooming in or out made when the system is observed. At a fine-grained description for instance, two replicated sequences of DNA are never perfectly identical and thus not of the same type. If this is true for DNA, this is even more relevant for two organisms or two groups of organisms. And this fact is not a biological one only. At a fine grained description there are no two identical snow-flakes. Depending on

which properties are measured and with what precision, two individual atoms of carbons of the same isotope might have different properties (e.g., different spin on their electrons).

This means that ultimately belonging to a particular type and producing offspring of the same type as one-self are not independent properties from the grain of description. By 'grain of description I mean here both the metric that is used to measure a particular variable (e.g., a scale of 'red') but also the temporal scale used to make to measure that particular variable (e.g., how long has the object remained 'red'), so the zooming in or out can also be made temporally.[92]

Scientists typically use different typologies to categorise the world which are driven by the question they want to answer. Take for example ecology and taxonomy. An ecologist might consider that two plants belong to the type 'primary producer' because they produce organic matter from inorganic matter. And in many cases whether they belong to the same species will be unimportant for the question at hand and thus not necessarily incorporated in a model. However, this difference will be of prime importance for a taxonomist. Another example is whether all the mutations leading to a particular phenotype are identical. At a certain grain of description and for some explanations it might not be relevant whether different human populations have different mutations to digest milk, and the exact biochemical pathway seems unimportant. Tishkoff et al. (2006), for instance, have shown that different mutations code for lactose persistence in different populations. There might be some small differences between each one of these mutations leading to a slightly different phenotype with some unknown pleiotropic effects. Yet, these differences seem irrelevant to explain why the tolerance of lactose is beneficial

[92] See Pocheville (2010, Chap. 2) which provides a nice treatment of this question.

in some environments. However, these differences might become relevant in some other contexts or if one intends for example to use these mutations as phylogenetic markers.

Although having different typologies at hand (either as a result of different grains of description or not) for one and the same object is very useful in science, it poses a problem for the concept of natural selection. In fact, whether natural selection occurs as a physical process in the world (in a population) should be independent from the particular scientific question under investigation and thus to a large extent independent from the grain of description used. Yet, as I demonstrate in the next section this is far from being the case.

5.3. Fitness and the reference grain problem

So far, we have seen that measuring fitness[93] crucially depends on type-identity, and that belonging to or producing a particular type rather than another one depends on the grain of description employed. In this section, I argue that if one accepts these premises, it follows that differences in fitness are not independent from the grain of description one uses and can lead to conflicting views as to whether a particular evolutionary change is due to natural selection or another evolutionary process.

One immediate consequence of the grain-of-description dependence of type-identity in relation to fitness is that one given entity, once examined from different grains of description, should be considered as having a different fitness. Following our definition of natural selection resulting from differences in fitness, this has an important consequence for our conception of

[93] From now on, by *fitness* I will mean *intergenerational fitness*.

the process of natural selection. This is because, natural selection, as I have argued, results from differences in fitness. The problem here is that assuming each grain of description would yield the same level of informational content,[94] nothing in evolutionary theory seems to be able to guide one's choice about what grain of description is the most appropriate to describe a particular situation or phenomenon.

This problem is unwelcome for several reasons. First, a scientific realist might want to argue that one desideratum for a property such as fitness to qualify as a *real* property is that it should remain independent from the grain of description. Second, arguably, obtaining a finer grained description, leaving aside problems of temporal scales, will in many case lead to a higher level of empirical content about a system and should allow us to identify a higher number of difference makers, which is one desideratum of predictive science. Yet, this leads to the worrying conclusion that, if one was to choose the lowest possible grain of description to describe evolutionary change, they would most likely have to claim that the difference in fitness between the difference entities is nil. This is because from the finest possible grain of description none of the parental entities would be able to produce perfectly identical offspring with respect to the focal property and thus the evolutionary change observed could not be attributed to natural selection.

Furthermore, according to the supervenience assumption which is widely accepted by scientists and philosophers of science, a physical property (and any process(es) depending on this property) existing at a coarser grain of description should also exist at the finest possible grain of description (even if it might be very difficult to describe) assuming that both descriptions are

[94] Which will not be true in many cases (see below).

faithful ones from their respective grain of description. This is because any two descriptions supervene on the same target system. However, it is also possible that some missing information from a coarse-grained description makes one physical property only detectable from a finer-grain description. Following the supervenience assumption, it is perfectly rational that different grains of description lead us to change our mind about whether an entity has a given property and whether a process occurs, but this should only happen when one moves from a coarser to a finer grain description not from a finer to a coarser grain of description (once again and this is important, assuming descriptions are faithful). This is because coarser-grained descriptions will in many cases have less empirical content than finer-grained descriptions, not the contrary. Thus, any physical properties and process 'discovered' at a given grain of description should always also be found at finer grains of description. Yet this is not necessarily the case for natural selection as I will illustrate below (see Figure 5.4 and Table 5.1 and main text below).

The different problems outlined above seem to be, if not specific to the process of natural selection, at least absent in some other physical and biological processes. To see this, take the property for a piece of ice melting at 0°C at normal pressure and the corresponding process of melting. Although this property/process couple can be described from different perspectives, assuming supervenience of all the descriptions on the same facts (with the same amount of empirical content), all these descriptions should lead to the conclusion that ice melts at that temperature and pressure (assuming temperature and pressure do not vary locally) even when some details are abstracted for explanatory salience. One could *in principle* describe the melting process from a molecular point of view, by describing how molecules of water become suddenly more agitated during what is referred to as the process of melting. This change in molecular agitation will simply be translated as 'the process of melting' at a macroscopic grain of description.

Although the two descriptions would involve a different terminology they would both be formally equivalent. This is because both descriptions supervene on the same objective facts of the world: the piece of ice is melting.

Take now a case of evolutionary change. Assuming a population of entities that can be categorised into distinct types from a given perspective in a perfectly homogeneous environment, while abstracting away some properties, we come to the conclusion that the different types have different fitnesses and thus that natural selection occurs in this population. As a result we claim that the population evolves by natural selection. But this conclusion might be altered if we change the grain of description. To see why, imagine now that you decide to be less abstract and introduce more details in your description. To do that, you separate each type you previously delineated into subtypes. And because of this different way of characterising the entities of the population, the conclusion is now that each subtype has the same fitness. For that reason natural selection cannot be invoked as the cause of evolution.[95]

To make this idea more concrete, Figure 5.4 and Table 5.1 provide an example. Imagine a population in which depending on what perspective we take, there are either two types of entities RED and BLUE or four types RED1, RED2, BLUE1 and BLUE2. From a coarse-grained perspective (P1) RED entities deterministically[96] produce 4 RED offspring in at each discrete generations, while BLUE entities produce only 3 BLUE offspring. Everything else being equal, from that perspective, it seems that natural selection occurs in this population because there is a difference in fitness between the two types (4 for RED and 3 for BLUE). Now, taking another perspective (P2) we categorise entities with finer nuances of colours and observe that the

[95] This assumes that types and subtypes are non-metrical types, that is, that they cannot be given a numeric value.
[96] The same demonstration could be made without the assumption of determinism.

RED and BLUE parents, when we were assessing their colour from P1, are respectively of the type RED1 and BLUE1 when assessing their colour from P2. Furthermore we notice that half of the offspring of RED1 are RED1 and the other half are RED2 while two third of the offspring of BLUE1 are BLUE1 while the remaining third are BLUE2 (see Figure 5.4). From that perspective, the fitness of RED1 and BLUE1 are identical since only the offspring of their types are incorporated into the measure of fitness and thus there is no natural selection in the population because types do not vary in fitness. Incidentally, transmission is non-systematically biased in the population, which is an analogous case to the case of meiotic drive mentioned earlier.

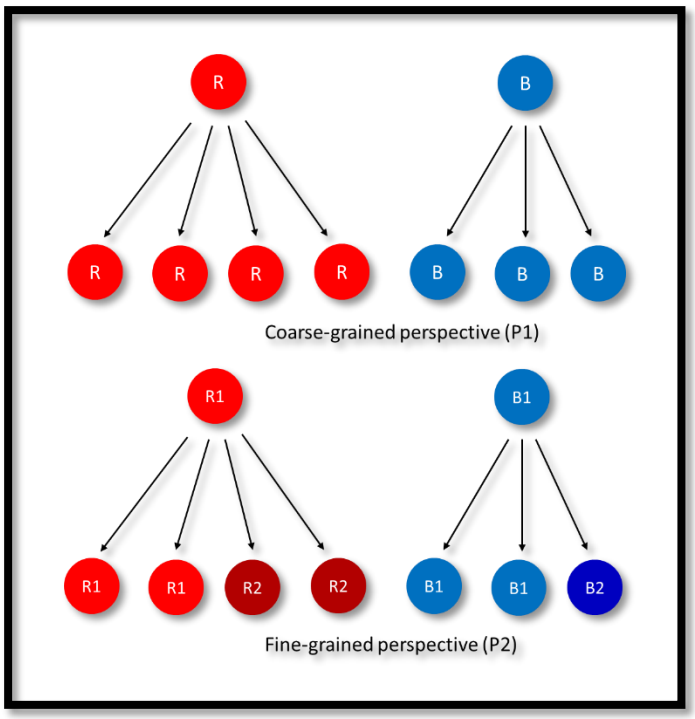


Figure 5.4. Reproductive output of two types of entities seen from two different perspectives with different grains of description.

Table 5.1. Conflicting views on whether there are differences in fitness between two entities and consequences for natural selection.

Perspective	Parental types	Reproductive output of the same type	Is natural selection occurring on the phenotype colour?
(P1) Coarse-grained	**RED**	**4**	**Yes because of differences in fitness**
	BLUE	**3**	
(P1) Coarse-grained	**RED1**	**2**	**No because no difference in fitness**
	BLUE1	**2**	

Contrary to the case of the melting ice, with the same initial amount of information about the world and because *you* decided to include more or less information in your description, conflicting views about whether natural selection occurs in this population are obtained. And the problem here is that each description is a legitimate one: there is no fact of the matter about which description is the correct one. In the case of the melting ice, changing the amount of details you decide to ignore does not lead you to the conclusion that ice melts in one case and not in the other. This is because holding background conditions constant, the melting point of ice is independent from the grain of description. This is different in this evolutionary case: Fitness is a property that is a function of a type when the melting point is not. This matters, because a type is a conceptual construct that is *not* independent from the grain of description employed. As a result fitness also depends on the grain of description employed, and the same goes for natural selection.[97]

[97] An analogous example could have been used using a difference in temporal grain of description in which **A** and **B** have a tendency to mutate to type **D** with a different rate. In the short term it seems that one type is fitter than the other, but because it mutates more often, on the long term there is no difference.

The problem described above represents an analogous problem to the well-known reference class problem in probabilities. If someone tells us that "we" have nine chances out of ten to die before the age of 90 years, this information will not make any sense if you do not know which class of reference has been delimited to obtain this probability. Is the class 'all humans on Earth', 'humans in our country', 'humans of our gender', 'humans living in the same city', 'humans with the same life-style', 'humans of the same age', 'living organisms', etc.? Because we belong to an infinite number of classes, one could say that we have an indefinite probability of dying before the age of 100, which represents a problem if probabilities are 'unconditional guides to life' (Hájek 2007) where 'unconditional' means 'independent from any reference class'. Although the reference class problem is classically thought to be only a problem for a frequentist interpretation of probability, Hájek (2007) convincingly demonstrates theory of probability theory has its own reference class problem. The problem I have presented with fitness has a common origin with the reference class problem, which as far as I am aware, is the first time this is recognised in the literature.

The common origin of these two problems lies in the fact that as recognised by Venn in 1866 "every single thing or event has an indefinite number of properties or attributes observable in it, and might therefore be considered as belonging to an indefinite number of different classes of things" (Venn 1876). In the case of the reference class problem in probability, because an event can be considered as belonging to many different classes, it is problematic to know what class should be used to predict it. In the case of natural selection, since different grains of description will lead to one class of objects (say RED) to be made of a multitude of subclasses (say RED1, RED2, RED3, etc.), it will be problematic to know which one of these different (sub)classes should be used to describe an event of evolutionary change. This problem widens

the scope of the reference class problem beyond probability theory and independently from it, since in my example, I have mentioned no probabilistic event. To distinguish the reference class problem in probability which occurs at any given grain of description from the reference class problem in evolutionary theory, I name the latter the reference *grain* problem.

One should note that the reference grain problem in evolutionary theory and the reference class problem in probability although overlap are nevertheless slightly different which justifies a different label. In fact, in the traditional reference class problem, the heart of the problem is to assess which one of an infinity of classes is the correct one for a given event. But the different possible classes an object belongs to *can be* independent from the grain of description used to delimit each class and depend only on different properties of the object (e.g., male, French, 30 years old). Contrast this with the reference grain problem in evolutionary theory. The problem here is also to choose the relevant class but not because one and the same object belongs to different classes that can represent different properties. Rather the problem here is that one and the same object can belong to many classes by merely changing the grain of description on one and the same property (e.g., intolerant to lactose and intolerant to lactose because of a particular metabolic pathway) and there seems no way to decide from evolutionary theory which grain of description and thus which class is best to use beyond our intuitions. To a large extent, our intuitions will depend on what question we want to answer. But if 'colour', to take the theoretical example in Figure 5.4, is causally linked to fitness, claiming that natural selection occurs or has occurred in a population between entities of different colours is not more meaningful than claiming you have 9 chances of 10 to die before 90. One must first stipulate which reference class and thus which grain of description one has used.

Because in the traditional reference class problem the different classes one and the same object belongs to can either represent different properties (e.g., red, tall, light) or one and the same property seen from different grain of description (e.g., red, crimson, Spanish crimson, etc.), it seems that the reference grain problem and the reference class problem represent subclass of a more general kind of problem that has to do with reference.

At this point, it should be noted that although the reference grain problem might not be limited to evolutionary theory and is certainly encountered in some other physical and biological processes, it takes a very particular form in evolutionary theory. In fact, the concept of fitness involves an assessment of similarities and differences between entities over time (such as between parent and offspring) and as I argued in the two first chapters, if offspring are different from their parents, the evolutionary process invoked to describe the evolutionary change should in this case differ from natural selection. But how much different is a parent from its offspring (or in the case of persistence, an individual between two times) depends from what grain of description one uses. It follows that from a different choice of grain of description, different evolutionary processes could be invoked for one and the same event. This type of phenomenon, where similarities and differences must be assessed over time to know to which evolutionary process the change should be attributed to, is not encountered in the case of the melting ice, in which it is assumed that if the ice has melted, one and the same mechanism is responsible across the whole ensemble of objects independently from assessing the similarities and differences over time of the objects forming the ensemble.

For instance, for the transition from a solid state to a liquid state for a block of ice, it is not supposed that there are different possible processes that can lead to the same outcome. In evolutionary theory, for one and the same evolutionary change say moving from a population

composed of 50% of **A** and 50% of **B** to a population composed only of **A**, there seems to be, at least two ways to get there. In fact, by merely changing the grain of description, assuming parents and offspring are not perfectly identical (so that there are subtypes of **As** and **Bs**), one could conclude either that the population has evolved by natural selection (at a coarse-grained description) or that it has evolved by mutation (in some cases of fine-grained descriptions).

5.4. Solving the reference grain problem?

As already noted Hájek (2007) demonstrates convincingly that the reference class problem is pervasive in probability theory and that all major probability interpretations face the problem in one form or another. Attempts to solve this problem have been made by arguing that one should use probability only over the narrowest class(es) for which reliable statistics can be compiled, or that one should use the maximal evidence one has at their disposal. However, these solutions are unsatisfactory. In fact, it is not clear what 'reliable statistics mean' (Hájek 2012), and using maximal evidence (which is equivalent to narrowing the reference class) will lead to use a reference class made of one single event, which means probability will become dispensable, unless the world is fundamentally indeterministic. One might be tempted to give a solution to the reference grain problem using an analogous solution to the ones proposed with the reference grain problem, that is, narrowing the reference class or using maximal evidence. It would corresponds here to use the finest possible grain of description one has at its disposal. Yet, as in the case of the reference class problem, this solution seems unsatisfactory. In fact, if one uses the smallest possible grain of description every entity is unique and thus any mention of natural

selection will degenerate into mutation. Furthermore most biological explanation use a rather coarse grain of description.

Because the solutions of narrowing down the reference class or using the maximal evidence at one's disposal are unsatisfactory, Hájek (2007) proposes to abandon the idea that events have unconditional probabilities and rather embrace the idea that any probability is relative to a reference class that must be specified to be meaningful. Although this solution is somewhat difficult to reconcile with the view that unconditional or single-event probabilities are 'guide to life', this might be the only solution to what seems otherwise to be an insoluble problem.

Concerning natural selection and the reference grain problem, an analogous solution might also be the only one available thus far. One could in fact argue that because any notion of fitness involving the notion of type will face a reference grain problem and that a notion of fitness without any (implicit or explicit) reference to type will be hard to reconcile with evolutionary theory, we should abandon the idea that fitness is an unconditional property of the entities forming a population. Rather, fitness should be seen as a property that is relative to the grain of description chosen and natural selection, mutation and drift the corresponding processes at that grain of description.

At this point one could interpret this proposal in two different ways. One could consider that fitness is just an instrumental property given the degree of knowledge one has and that entities do not literally have any fitness at all independent from a given grain of description. This moves would be a similar one to the one made by Rosenberg (1994) concerning biology as a whole. Another is that a given entity has an indefinite number of fitnesses at the same time. I have no particular view on which one of these two options is the best. The main issue here is to

reconcile the fact that the concepts of fitness and natural selection are very useful and successful tools for thinking about evolution (especially with respect to adaptations) with the idea that the presence/absence of natural selection will depend on what grain of description is used and thus does not represent objective facts about the world.

One way this could be done using Godfrey-Smith and Weisberg's account of models (Godfrey-Smith 2006, 2009a, 2009c; Weisberg 2012, 2013) would be to consider that natural selection is not a physical process in the same ways phase transitions are, but rather a model of a process that bears some resemblance with a particular target system at a particular grain of description. Or more precisely that the process of natural selection is a more idealised model that of the melting of ice, because it is less robust to changes in grains of descriptions.

5.5. Conclusion of the chapter

In this chapter, I have shown that the concept of fitness suffers from a problem analogous to the reference class problem in probability. I labelled this the *reference grain problem*. This problem originates from the fact that the notion of fitness and consequently the process of natural selection both depend on the types delimited in a population. Yet, different grains of description lead to the view that one and the same entity belongs to different types and produces offspring which either are of the same type or of different type from their parent. This can lead to conflicting views as to whether natural selection or (an)other evolutionary process(es) are responsible for the evolutionary change observed. Far from having proposed a definitive solution to this problem, I have suggested that one way to dissolve the problem would be to consider natural selection as a model of a process rather than a physical process.

In the next chapter I turn to a topic that has received a lot of attention in philosophy of biology and evolutionary theory more generally, namely the levels of selection and more particularly one of its contemporary avatars known as evolutionary transitions in individuality. In some models on evolutionary transitions in individuality it is assumed that an ontologically new level of selection arise once the transition is complete, that is, when a new biological individual has emerged at a higher level of organisation (e.g., a multicellular organism from unicellular organisms). I will show however that this claim does not resist close scrutiny. At best, levels of selection can be thought as one and the same process carved over different time scales or while referring to different substrates in a given population and not as ontologically distinct selection processes.

6. Time, Levels and Fitness in Evolutionary Transitions in Individuality

6.1. Introduction

Multilevel selection is the view that selection can act simultaneously at different levels of organisation. The simplest case of multilevel selection is a case with two levels: one I refer to, following Okasha (2006), as the *particle* level and the other I refer to as the *collective* level, with collectives being constituted of particles. The concept of multilevel selection has been one of the most contentious in evolutionary biology over the last 50 years, especially because it is related to the notion of group selection (for an overview of the debates surrounding this concept see Okasha 2006; Sober & Wilson 1998; Williams 1966, 1992; Wilson & Wilson 2007). Group selection is the view that groups made of individuals organisms can be selected independently from the individuals that compose them and thus represents a specific case of multilevel selection. Although largely rejected by the mid 1960's, especially due to the influence of George Williams' *Adaptation and Natural Selection* (Williams 1966), the notions of group selection and multilevel selection have progressively been made respectable (Okasha 2006; Sober & Wilson 1998; Wilson 1975; Wilson & Wilson 2007).

There are several reasons why the notions of group selection and multilevel selection have been particularly debated. One reason, maybe the most important one, is that in a multilevel setting, it can be very tricky to determine the extent to which each level causes a trait to spread in a population. Many disagreements remain as to which criteria should be used to do so. One of

the clearest, but also earliest discussion explaining why causality at different levels of organisation matters so much in multilevel settings can be found in Williams (1966, 16-17). Williams famously contrasted a 'herd of fleet deer' with a 'fleet herd of deer' to illustrate that some collective traits (in his example the fleetness of a herd of deer) leading *prima facie* to differences in particle fitness, can actually be reduced to differences in particle character (the fleetness of each deer composing the herd) leading to differences in particle fitness. The collective character and collective fitness, in this case, result merely from the summation of particle character and fitness composing the collective. Thus, according to Williams, the fleetness of the collective should not be seen as a 'group adaptation' but rather as a 'fortuitous group benefit'. It is now widely accepted, both by the proponents of the multilevel selection framework and their opponents, that this argument is a correct one (see for instance Sober & Wilson 2011). In these cases both camps agree that selection only occurs at the particle level, for differences in collective fitness represent merely a by-product of differences in particles fitness. Okasha (2006) calls these by-products 'cross-level by-products'. [98] Although this argument is widely accepted, it is far from obvious what the best way to causally decompose the effects of particle character and those of collective character on particle fitness is.

Another reason why the concept of multilevel selection has been contentious is that different authors have generally meant two distinct things by employing the term 'multilevel selection' (or related terms). The conflation of these two concepts has been a source of confusion in the debate over multilevel selection. These two notions have first been clearly distinguished by Damuth and Heisler (1988) although similar (but less precise) distinctions can be found in earlier

[98] A cross-level by-product is defined by Okasha (2006,76) as "a side effect, or by-product, of selection acting at a different level".

discussions (see Damuth & Heisler 1988, 410; Okasha 2006, 56). Under what Damuth and Heisler call the multilevel selection 1 (MLS1) framework, the focal level from which we assess the selection process(es) is the particle level. Both particle and collective fitnesses are measured with the same metric. Usually particle fitness will be measured as the number of offspring particle produced after some time and the collective fitness as the number of offspring *particles* produced after some time. Under what Damuth and Heisler call the multilevel selection 2 (MLS2) framework, both the particle and collective levels are the focal levels from which we assess the selection process(es). The particle fitness and collective fitness have, in this case, different metrics. Usually, but not necessarily, collective fitness will be measured in terms of offspring *collectives* produced after some time while particle fitness will be measured in terms of offspring particles produced after some time.

Finally, another reason why the concept of multilevel selection is still muddled is that some authors understand multilevel selection as differences in fitness between particles *within a collective* and difference in fitness of collectives *between collectives* (e.g., Heisler & Damuth 1987). Others such as Nunney (1985) understand multilevel selection as differences in fitness between particles and between collectives when *the whole population is taken into account at both levels*.

Recently the notion of multilevel selection has been extended to what has been called 'evolutionary transitions in individuality'. Evolutionary transitions in individuality are events in the course of evolution that lead to the formation of new higher level individuals due to the cooperation of two or more individuals at a lower level of organisation (Michod 2011). The notion of evolutionary transitions in individuality, originates from the literature on major transitions in evolution which rapidly rose after the publication in 1995 of *The Major Transitions in Evolution* (Maynard Smith & Szathmáry 1995). In this book Maynard Smith and Szathmáry

proposed eight major transitions that changed drastically the course of evolution of life. Although to a large extent what Maynard Smith and Szathmáry refer to as 'major transition' is what I will refer to as 'evolutionary transitions in individuality', the notion of major transition is vaguer and does not necessarily concern individuality. For that reason I will stick with 'evolutionary transition in individuality'. One example of evolutionary transition in individuality that will be the example used throughout this chapter, is the transition from uni- to multicellular organisms. A number of other evolutionary transitions in individuality have been proposed, among them the transitions from prokaryote to eukaryote cells and from multicellular organisms to integrated colonies such as colonies of ants or honeybees.

The most accomplished models of evolutionary transition in individuality, called the-export of-fitness view, is certainly the one proposed by Michod and colleagues. In a number of articles and books (Michod 1999, 2005; Michod, et al. 2003; Michod, et al. 2006) they detail the necessary conditions for evolutionary transitions in individuality to occur. Okasha (2006, 2011) has recently set Michod and colleagues' work in the MLS1/MLS2 framework. Although Michod's model is supposed to be applicable to a wide range of transitions, it stems from Michod's preferred organism model, namely organisms from the genus *Volvox*. There are other organism models and conceptual frameworks of evolutionary transitions in individuality to which I cannot give justice in this chapter. The two most notable are *Pseudomonas fluorescens* (e.g., Rainey & Kerr 2010; Rainey & Rainey 2003) and *Saccharomyces cerevisiae* (e.g., Ratcliff, Denison, Borrello, & Travisano 2012). Some examples of other conceptual frameworks can be found in Libby and Rainey (2013), De Monte and Rainey (2014) and the literature on the topic is treated in Calcott and Sterelny (2011). However, some of the argumentation provided in the chapter could be easily

applied to them since elements of the export-of-fitness view can be found in all of these frameworks.

Okasha's and Michod's models of evolutionary transitions in individuality are committed to the classical concept of fitness, that is, measured by the ability of a given entity to survive (viability) and reproduce given that the entity has survived (fertility) in its environment (see Sober 2001 for more on this way of conceptualising fitness). According to this definition, the higher the product of viability and fertility of an entity is, the higher its fitness is. Although this definition has been heavily criticised in the previous chapters as overly restrictive and causally problematic, I will accept it as common ground for the development of the different problems I point out in this chapter. The main aim of this chapter is to argue against the export-of-fitness view of evolutionary transitions in individuality and more generally to provide an argument against the view that MLS1 and MLS2 represent two ontological processes of selection as it appears to be assumed by Okasha and Michod. Some of the arguments in this chapter are related to those made in Section 2.5, where I criticised the concept of individuality for its fuzziness.

The remainder of the chapter will be divided into four sections. In Section 6.2 I briefly review Michod and colleagues' as well as Okasha's models of evolutionary transitions in individuality with respect to the concept of fitness. I present two specific claims defended by both authors: (1) that during the last stage of an evolutionary transition in individuality, once a division of labour is in place, the fitness of the entities constituting the newly emerged individual becomes nil; (2) that there are two fundamentally distinct processes of selection, namely multilevel selection 1 (MSL1) and multilevel selection 2 (MLS2), occurring alternately at the different stages of an evolutionary transition in individuality. Both claim (1) and claim (2) are consequences that follow from the export-of-fitness view.

In the Sections 6.3 and 6.4, I respectively attack claims (1) and (2). Claim (1) has recently been briefly criticised by Godfrey-Smith (2011, 77-78) for its metaphorical nature. In Section 6.3, I propose a more thorough examination of this claim and show first that Michod and colleagues' models rely on one assumption with regards to fitness of the collective which is inconsistent with the claim made about the fitnesses of the particles. I claim that an ambiguity surrounding the concepts of viability (fertility) of the particle and viability (fertility) of the collective exists in Michod's and Okasha's writings and creates the artefact that the fitness of the particle can be nil while the fitness of the collective can be high. Second, I propose that even in cases where the fertility of the particles is literally nil, this is because the concept of reproduction used by Michod and Okasha relies too much on material overlap such as proposed by Griesemer (2000). This concept, I will argue, is unnecessary for Darwinian evolution to occur. Expanding on Godfrey-Smith (2009) and his concept of formal reproduction I show that somatic cells of multicellular organisms formally reproduce and consequently that their fitness is not nil.

In Section 6.4, I turn to claim (2) and demonstrate that if fitness is assessed over the same period of time at the collective level and at the level of its constituents, then there most likely is commensurability between selection at the particle and collective level during an evolutionary transition in individuality. For that reason, they should not be claimed to represent two ontologically distinct processes of selection. I propose instead that they represent two ways of describing one and the same process from the perspective of two spatial and temporal scales. To do so I will use one of Okasha's (2006) example from the literature Yet, I will not deny the epistemological value of describing evolutionary transitions in individuality within the MLS1/MLS2 framework but will conclude that, so far, there is no clear evidence to support the claims given by Okasha about ontology.

Finally, in Section 6.5, I present another source of confusion in the levels of selection debate regarding the fact that it is not recognised that in many cases lower-level and higher-level entities belong to two different hierarchies.

Before going further, I need to make a point clear. In the rest of the chapter, I will at times claim that MLS1 and MLS2 are commensurable or equivalent. This should only be understood in the context of Okasha (2006) where he proposes a model giving an *ontology* to MLS1 and MLS2. Damuth and Heisler's (1988) initial distinction only regards MLS1 and MLS2 as two possible *methods* to study multilevel selection. In fact, they write: "Once one has decided to analyse a given situation in terms of multilevel selection processes both approaches [MLS1 or MSL2] are legitimate within that context and a choice has to be made depending upon what questions are of interest" (Damuth & Heisler 1988), which clearly indicates that their distinction is about method not ontology. I will thus not contend that the two methods are equivalent but rather that using one method instead of another one does not point to a different ontology.

6.2. Michod and Okasha on evolutionary transitions in individuality

Michod (2005) proposes the following model of evolutionary transition in individuality. For new individuals at a collective level to emerge from a particle level, e.g., for multicellular organisms to emerge from unicellular organisms, two things must happen. First, conflicts between members of the collective need to be resolved. Two ways that this can happen are the presence of policing mechanisms in the collectives and the presence of developmental bottlenecks during life cycles. These two mechanisms lead to genetic homogeneity that reduces genetic conflict between the cells of an adult multicellular organism. However, even if genetic

homogeneity is reached between the different members of the same group, this will not necessarily lead to the emergence of a higher individual.

For an evolutionary transition in individuality to take place, Michod (2005) proposes that there must be a division of labour between germ and soma (or its equivalent in evolutionary transitions in individuality other than from uni- to multicellular organisms), since without this division of labour, he argues, the collective fitness will remain strictly proportional to the average particle fitness. As such, the collective will not be an individual with its own fitness (Michod 2005, 569); its fitness will merely be a cross level by-product (see the Introduction of this chapter) of its particles' fitnesses. Thus, for Michod, the concept of individuality involves irreducibility of fitness at that level. We saw in Chapter 2 that Godfrey-Smith on the one hand, and Queller and Strassmann on the other hand, had two different view to define biological individuality. Michod's criteria, although to some extent overlapping with those of Godfrey-Smith and Queller and Strassmann, represent yet another alternative for a definition of individuality.

6.2.1. Claim 1

As I noted in the Introduction of this chapter, Michod defines the fitness of an entity (whether particle or collective) as the product of its viability and fertility, as is often done in life-history models. In the cases of evolutionary transitions in individuality from unicellular to multicellular organisms with full separation of germ and soma, when a cell is not specialised, it invests its resources in both the viability and fertility components of fitness. As a result, following Michod's reasoning, its fitness is positive. However, when a cell invests everything in viability or fertility, its fitness becomes nil because its product of viability and fertility is zero. Michod (2005,

2011) and Okasha (2009) both generalise this argument to other evolutionary transitions in individuality and propose that:

> **(1)** If a particle invests everything in the somatic (or germ) function (or its equivalent) of the future collective individual, it will have a fitness nil since although its viability (or fertility) component of fitness will be positive, its fertility (or viability) component and consequently the product of viability and fertility will be nil.

Conversely, they propose that when the two types of particles combine their investment in both components of fitness (one type investing everything in the soma and the other type investing everything in the germ function) a new collective individual emerges with its own fitness. This reasoning leads Michod to claim that during an evolutionary transition in individuality, there is a transfer or export of fitness from the particle to the collective level, hence the name of his view.

6.2.2. Claim 2

Okasha (2006) and Michod (2005, 2011), mostly relying on Okasha's analysis, both link this work to the two concepts of multilevel selection distinguished by Damuth and Heisler (1988) and presented in the introduction of this chapter, namely MLS1 and MLS2. Let me reiterate that in the MLS1 framework according to Okasha, the focal unit of selection is the particle. For that reason fitness is expressed as the number of particles produced. A collective has a higher *fitness 1* than another if *ceteris paribus* it produces more particles. In the MLS2 framework, the focal units of selection are both the particle and the collective. Fitnesses of the collective and of the particle are measured in different units. The *fitness 2* of a collective is expressed as the number of new collectives it produces independently of the number of particles composing each collective, while the fitness of a particle is simply expressed in the number of particles it produces.

Okasha (2006, 237-238) proposes that an evolutionary transition in individuality can schematically be decomposed in three temporal stages for which MLS1 and MLS2 are alternately more relevant to describe the selection process(es) occurring at that stage. According to Okasha, in the first stage of an evolutionary transition in individuality, the particles of the future collective start to aggregate and cooperate. The fitness of this newly formed collective is merely the average of the particles' fitness, hence MLS1 is the relevant type of selection occurring. During the second stage, the fitness of the collective is no longer defined in terms of the particles, but is proportional to the average fitness of the particles. At that stage, although the MLS2 framework can be applied, so can the MLS1 one. There is a "grey area between MLS1 and MSL2", in Okasha's words (2006, 237). However, the collective lacks individuality, since its fitness is a cross-level by-product of the particles' fitness. During the third stage, when the transition is complete, the fitness of the collective *cannot* be expressed as the average fitness of the particles anymore and it is not proportionate to it. The collective is now an individual on its own and its fitness is not proportional to the fitness of the particles; both fitnesses are now incommensurable. Michod and colleagues use the word 'decoupling' to refer to this phenomenon. By decoupling they mean that the fitness at the collective level becomes expressed in a different currency than fitness at the particle level and that it is not translatable into fitness at that level. Their idea is thus the same as the one proposed by Okasha. This leads Okasha to claim that:

> **(2)** MLS1 and MLS2 are two ontologically distinct *causal processes* of selection as opposed to two conventional ways of expressing selection (2006, 59; 2011, 243). During an evolutionary transition in individuality, they represent a transition in processes of selection. Not only MLS1 and MSL2 are alternately more relevant at the different stages

of an evolutionary transition in individuality, they are represent ontologically distinct processes of selection.

6.3. Particles with nil fitnesses?

In the previous section, I presented (1), the claim that for an evolutionary transition in individuality from uni- to multicellular organisms (or more generally from particle to collective) to be successful, cells must renounce their individuality by specialising either in the soma or germ function of this higher level individual. As a consequence of this specialisation, Michod claims, their fitness taken separately is nil. Godfrey-Smith (2011, 77-78) briefly commented that this way of speaking about fitness is metaphorical. Indeed, he remarks that even in a paradigmatic case of multicellular organisms, germ cells do not literally have a viability of nil and many somatic cell divide, hence their fertility is not nil either.

In this section, I start by making Godfrey-Smith's remark more precise and show that one assumption in Michod and colleagues' model, namely that the viability and fertility of the multicellular organism are fully and only dependent respectively on the viability and fertility of the cells it is composed of, is quite misleading. At that stage, it will become clear that Michod and Okasha confuse viability and fertility *of the cells* with viability and fertility *of the multicellular organism*. Although the assumption that the fertility of somatic cells is nil might in some cases possibly be true if by 'fitness' Michod means 'reproductive output', in the second part of this section, I show that even if the fertility of somatic cells can be conceptualised as nil, it must come from the fact that Michod and Okasha understand 'reproduction' as 'reproduction with material overlap'(for more on reproduction with material overlap see Griesemer 2000, 2014). I argue that material

overlap unnecessarily constrains the idea of ENS. Following Godfrey-Smith, I propose a broader concept of reproduction, namely *formal* reproduction (Godfrey-Smith 2009b, 79-81), to be used when describing foundational Darwinian processes such as evolutionary transitions in individuality. I show that if the concept of formal reproduction is used to measure the fitness of somatic cells in multicellular organism, they do not have a fitness nil because they formally reproduce.

6.3.1. Viability and fertility at different levels

Once an evolutionary transition in individuality from uni- to multicellular organisms with germ soma separation is completed, both types of cells in the organism cannot exist independently from each other and contribute durably to evolutionary change. In fact, without germ cells, the organism would eventually die without being able to reproduce; and without somatic cells, the organism could not survive long enough to reproduce. From a multicellular-organism perspective, it is therefore correct to state that without taking both the germ and somatic lines into account to measure fitness at the multicellular level, the fitness of the organism after one generation would be nil. But that claim is a different one from the one made by Michod and colleagues which is that the fitness *of the cells* becomes nil once the evolutionary transition in individuality is completed (Michod 2005). There is obviously a sense in which the fitness of a cell could be assumed to be nil. This would be when the fitness of a cell refers to the fitness that this cell *would have had* in the absence of a cell-group environment. But this definition of fitness cannot play any causal role in ETIs. This is because a situation in which cells have no group-cell environment does not correspond to an *actual* situation encountered by the cells. In fact, the cells of which Michod claims have a nil fitness are always (or at least most of the time) surrounded by a cell-group environment. It seems thus incorrect to claim that their fitness *is* nil and would be

more accurate to claim that their fitness *would have been* nil. Making this sort of counterfactual claims can be useful in *explanations* but cannot be invoked for claims about *ontology*. Similarly, claiming that you would have had a fitness nil if you did not have received milk from your mother, does not mean that your fitness *is* nil although it could be a useful claim to make for some particular explanation. Thus, although it might be justified to claim that a new individual emerges when one can identify a germ-soma specialisation (although see the problems associated with the concept of individuality I discussed in Chapter 2), attempting to equate viability and fertility of the particle-level entities (the cells) with viability and fertility of the collective-level entity (the multicellular organism) seems misguided.

It is clear that a causal relation between viability and fertility at the two levels exists, but viability and fertility of a cell and of a multicellular organism clearly refer to different biological processes in most examples of multicellular organisms. To clarify this point, let us take a hypothetical multicellular organism such as the one used or modelled by Michod and co-workers (of the genus *Volvox*) in different articles (Michod, et al. 2003; Michod & Roze 1999; Michod, et al. 2006). Let us imagine that this multicellular organism starts to develop through the multiplication of two lines of cells, one which will perform the somatic function of the multicellular organism (for example maintain afloat the multicellular organism), the other one which will perform the germ function of the multicellular organism and which will, once the organism has reached adulthood, give birth to new multicellular organisms.

In this hypothetical case, viability and fertility at the collective level do not respectively equate to viability and fertility at the particle level. In fact, during the phase of development of the multicellular organism, that is, when only the somatic or viability function of the organisms is performed, the somatic line has offspring. For somatic cells to have offspring entails that they

both survive and reproduce. Thus viability at the collective level is a function of viability and fertility at the particle level. Similarly, during the phase of reproduction of the multicellular organism, the germ cells, as a matter of fact, both survive and reproduce. Thus fertility at the collective level is a function of viability and fertility at the particle level. If this reasoning is correct, it results that the most important difference between the developmental and reproductive phases of the multicellular organism from the point of view of cells is not whether cells survive and/or reproduce, but how they are *distributed*, that is, either in the multicellular organism for somatic cells or its future offspring for the germ cells. Hence, not distinguishing the level at which the concepts of viability or survival and reproduction or fertility are applied can lead to the odd claim that the fitnesses of the cells of a multicellular organism are nil. Another way to make the same point is to remark that the viability and fertility of the multicellular organism result from the *effort* of the cells in those variables. Yet, if the invested effort of a particular line of cells in the viability (fertility) of the multicellular organism is nil, it does not follow that the viability (fertility) of the cell is nil. This point is to relate to the examples given in Section 2.5, where I showed that in some cases there is a direct translation possible between reproduction at the particle level and viability and/or developmental growth at the collective level.

The confusion of not distinguishing viability and fertility at both levels is most vivid in Okasha (2009, 567), in which he summarises Michod's model of evolutionary transition in individuality from unicellular to multicellular organisms. In fact Okasha writes:

> [Let us] consider the extreme case where there is complete germ-soma specialisation— a cell either focuses entirely on viability (so [fertility of the i^{th} cell:] $b_i = 0$), or entirely on fecundity (so [viability of i^{th} cell:] $v_i = 0$). Clearly, this means that each cell has an individual fitness ($v_i\, b_i$) of zero, so average cell fitness C is zero too. Group fitness G, however, defined in Michod's way, may be very high—

> essentially, because some cells' null investment in v (or b) may be offset by other cells' high investment in v (or b). Had group fitness been defined as average cell fitness, this would not be possible.[99]

What seems to be a confusion in this quote is that a cell *focuses* entirely on viability or fertility *of the group* and not on fertility and viability *of the cell* as Okasha suggests when he writes $b_i = 0$ and $v_i = 0$. In the case described by Okasha, only the investment in V and B (respectively viability and fertility of the group) are nil.

In their models, Michod *et al* assume that:

> [F]or simplicity, […] the viability and fecundity of the group, V and B, respectively, are simple additive functions of the cell properties given by: $V = \sum_{i=1}^{N} v_i$ and $B = \sum_{i=1}^{N} b_i$ [where vi and bi are respectively the viability and fertility of the i^{th} cell of the group]. (Michod et al 2006, 262)

They couple this assumption with the assumption that when an evolutionary transition in individuality is complete $v_i = 0$ and $b_i = b_{max}$ for the germ cells and $b_i = 0$ and $v_i = v_{max}$ for the somatic cells. This means, at that point, that V and B depends respectively solely on the viability of the somatic cells and the fertility of the germ cells. It may be biologically meaningful to assume that the viability of some multicellular organisms depend solely on the viability of one particular type of cells that never reproduces. However, that should not be regarded as the general case. What is more, assuming that the fertility of the multicellular organism depends solely on the fertility of germs cells with a viability of nil is biologically meaningless, since a necessary condition to have offspring is at least to be alive! In other words following the assumption that the viability of the germ cells forming a multicellular organism is

[99] Okasha and Michod used 'fecundity' instead of 'fertility'. I take both terms to be equivalent.

nil, the fitness of the whole organism should be nil because without existing, they could not reproduce.

6.3.2. Fitness and formal reproduction

The previous argument only undermines Michod's export-of-fitness model with respect to the assumption that the fertility of the organism is only dependent on the fertility of the cells. But so far, as noted above, the assumption that $V = \sum_{i=1}^{N} v_i$ seems plausible, at least for some specific cases, such as when somatic cells never reproduce or when they do not have any offspring in the long term and that b_i measures the fertility rate of cells across multicellular generations. If we follow this reasoning, it can be tempting to assume, as do Michod and colleagues and Okasha, that the fitnesses of somatic cells are nil. In the remainder of this section, I show that one should resist this temptation.

One way to show that the assumption of somatic cells as having a nil fitness is misled involves defining 'reproduction' in a way that is necessary and sufficient for ENS to occur. Godfrey-Smith (2009, 79-85) proposes that a concept of reproduction which necessitates material overlap as proposed by Griesemer (2000) is not suitable for a foundational description of Darwinian processes while Griesemer (2000) argues that for evolution by natural selection to occur, similarity between parents and offspring is insufficient. Material overlap involves the offspring to be made of the parts of their parents (2000, 74). Godfrey-Smith prefers the concept of formal reproduction which only involves an unspecified relation of causality between the form of the 'parent' and of its 'offspring' and thus without the requirement of material overlap between the two. This is the concept I follow here.

The main argument provided by Godfrey-Smith to prefer the notion of formal reproduction to the notion of reproduction with material overlap is that there are cases of entities in which there is no material overlap, yet these entities are classically recognised as evolving by natural selection. The clearest example used by Godfrey-Smith comes from retroviruses in which due to reverse transcription, there is no parental material of the parent present in its offspring (Godfrey-Smith 2009, 80). Many more examples exist.[100] In a recent article, Griesemer (2014) asserts that he "accept[s Godfrey-Smith's] conceptual contrast of formal and material, but not his reasoning about which are the material cases."(p.36). Griesemer shows that depending on the scale or grain of description used and the choice of "central subjects", a term borrowed from Hull (1975), the case of the replication retroviruses can be conceptualised in terms of material overlap. Although I accept that it might be possible to re-describe many cases, which under an intuitive or classical approach would be recognised as cases of formal reproduction, in terms of material overlap, this should not be a concern for me. Suffice to say that *at a given grain of description* and *after having chosen central subjects*, material overlap will not be an important criterion to determine whether a population is evolving by natural selection.

In the case of reproduction of a multicellular organism, there seems to be no material overlap between the somatic cells of the parental organism and the offspring organisms. Thus, somatic cells are not the material parents of the offspring organisms. However things are different if by reproduction we have in mind *formal* reproduction. In such a case, one can argue that the somatic line of a multicellular organism does formally reproduce. To convince ourselves, let us imagine the case of a multicellular organism composed of somatic and germ cells. Let us further

[100] See for instance the cases of prions and LINE transposon provided by Godfrey-Smith (2009, 80). Kin selection, which is standard evolutionary theory, represents another example in which the reproduction of one allele is done "indirectly", that is, without material overlap (e.g., sterile workers helping the queen).

postulate that the fitness of the multicellular organism depends on both its number of somatic cells and number of germ cells. If it is composed of an insufficient number of somatic cells its fitness plummets, for example because this low number does not allow the multicellular organism to remain afloat. Similarly, if it is composed of too many somatic cells its fitness also plummets because with a higher number of germ cells the organism could have produced more offspring organisms. The number of somatic cells in the multicellular organism is thus as important as the number of germ cells for the reproductive output of the organism.

For formal reproduction to happen the only requirement is a causal link between the presence/absence of a somatic cell in the parental organism and the overall number of cells produced (number and quality of offspring). This is precisely what Michod *et al*'s model test:

> [W]e considered how a single new cell could maximize its fitness contribution to the group. In the optimization model we test whether small deviations by two or more cells could increase the fitness of the group. (2006, 262)

Indeed, in Michod *et al*'s models one can establish the causal relation between the parental and offspring cells by first manipulating the presence or absence of only one somatic cell in the parental organism. From there, they can deduce that *ceteris paribus* one more somatic cell in the parental organism will lead to a higher number of offspring cells composing the offspring multicellular organisms (higher number of offspring organisms or of better quality). Causation can thus be understood here under the manipulationist account of causation presented in the first chapter, although Michod and colleagues do not make such connection. Let us recall that the cornerstone of the manipulationist account of causation is according to Woodward (2013) that "if *C* is genuinely a cause of *E*, then if I can manipulate *C* in the right way, this should be a way of manipulating or changing *E*". In our case, manipulating the number of somatic cells in the

right way has consequences on the fitness of the whole organism (see Figure 6.1 for a toy illustration).

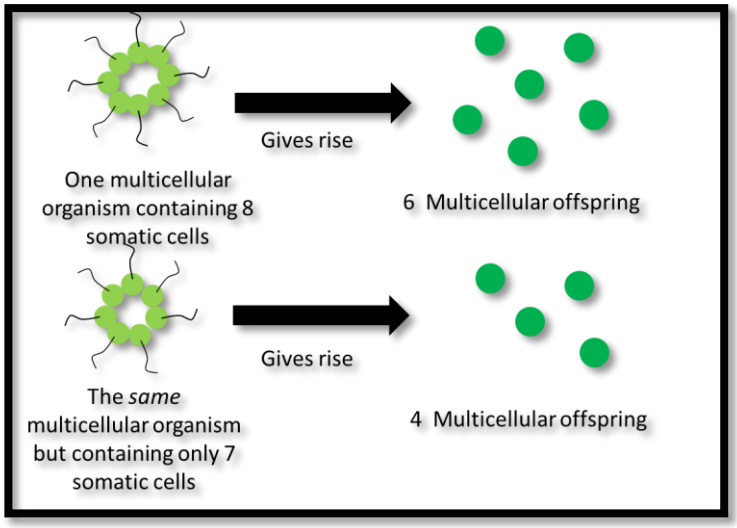


Figure 6.1. Toy illustration of the formal causal relation between the presence/absence of one somatic cell in a multicellular organism of the genus and Volvox the number of multicellular offspring produced.

Everything could be seen here as if somatic cells were sending a message to the germ cells to reproduce more cells at the next generation.[101] But is the fact that somatic cells are causally responsible for the production of somatic *and* germ cells not undermining my argument? In other words, is the fact that the manipulation of a number of somatic cells does not specifically alter the number of somatic cells in offspring organisms not a problem for my view? I claim that it is not for the following reasons. First, it is important to note that for ENS to occur, as shown in Chapter 2, perfect reproduction is unnecessary. It is only necessary for *pure* ENS to occur. The only requirement for *some* ENS to occur, assuming that no *new* type can be produced, is that some

[101] This is to be taken here as a metaphor.

information (the type) from parents and offspring is conserved over time. This case falls perfectly under any other cases proposed in this thesis and is depicted for instance in Figures 5.1 and 5.2, where entities produce offspring of their type with a frequency lower than one.

Another point worth mentioning is that in that respect germ cells and somatic cells are on a par. In fact, germ cells in the parental organism also produce the somatic cells of the offspring organisms. The only important property for a somatic cell of a given type to be evolutionary successful, that is, to evolve by natural selection, is that the relative number of its type increases over time. Whether this occurs through reproduction with material overlap, with formal reproduction or any other mean (such as developmental growth) does not matter. This idea is to relate to the points made first in Chapter 2 and then in Chapter 4 more formally, that what matters for evolutionary success is a causal link between tokens of the same types over time no matter what the causal link is. Another way to make this point is to notice that somatic and germ cells have a relatedness of 1 and thus following standard kin selection theory (W. D. Hamilton 1964), the *indirect* fitness of somatic cells, as opposed to their *direct* fitness, is not nil. Yet, in kin selection theory the total personal fitness of an individual is measured as the sum of the direct and indirect component of fitness (see Gardner & Foster 2008; West, Griffin, & Gardner 2007). Thus having a nil direct fitness does not necessarily mean that the total personal fitness of a cell is nil.

To sum up, I have thus far shown that an evolutionary transition in individuality does not lead the particles forming a collective to have a fitness nil as it has been argued by Michod and Okasha. This becomes clear when one does not apply the concept of viability and fertility of the collective to the particles and when the classical concept of reproduction is substituted with the more general concept of formal reproduction. In this latter case, somatic cells do reproduce at

each new multicellular generation of multicellular organism and consequently their fitness cannot be said to be nil.

6.4. When time makes a difference

I turn now to claim (2). If claim (1) seems unjustified in light of the above arguments, the claim made by Okasha and Michod that during an evolutionary transition in individuality an ontologically new level of selection is created at the collective level, leading to two ontologically distinct levels of selections once the evolutionary transition in individuality is completed, remains untouched by my arguments.

The first thing to note is that, according to Okasha, once an evolutionary transition in individuality is completed, the particles and the collective fitnesses become incommensurable: one could not, even in principle, measure the fitness of the collective in terms of fitness of the particles. Where does such incommensurability come from? To this question there is no clear answer and it is hard to see how there could be one even in principle. It is in fact hard to imagine that collectives could exhibit variations in fitness without their constitutive parts exhibiting a form of variation with consequences on their own fitness if the collective is made of particles and nothing else (I will have more to say about this in Section 6.5). Yet Okasha believes that such scenarios exist (Okasha 2006, 106) and that they materialise when MLS2 is the framework of choice, for the MLS2 framework, he claims, fits two causally distinct processes of natural selection happening in nature (Okasha 2006, 59; 2011, 243). Recall that in the MLS2 framework, the fitness of the collective can be defined as a quantity "that bears no *necessary* relation to average particle fitnesses alone" (2006, 136, my emphasis). Yet, in the same sentence Okasha surprisingly

asserts that "it is *impossible* that the resulting evolutionary change could be expressed in terms of particle fitnesses alone" (Okasha 2006, 136, my emphasis). Beyond the fact that the claim does not follow from the premise (Okasha should have used "sometimes impossible" instead of "impossible"), I propose one important reason why we should doubt this claim in any case.

I will not argue here against the MLS2 framework itself since it is obviously logically true. In fact, one can choose to define fitness of the collective with different metrics than fitness of the particles. There is nothing wrong with that. However, I will show that the arguments provided for the claim that there is incommensurability between the particle and collective fitnesses in cases in which there is ENS is unwarranted. I illustrate the point I make with one of Okasha's own examples of MLS2.

In his chapter 7 Okasha (2006) deals with species selection, the paradigmatic case of MLS2 in the literature on the subject, and embraces Vrba's 'acid test' (1989, 155) to detect true species selection (and more generally MLS2) from mere by-products of selection at lower levels, as in MLS1. Vrba proposes that there is true species selection when the outcome of selection at the species level cannot be explained from the perspective of the organism One stringent way (which represents the test) to know when this happens is to seek different directions of selection at the different levels of organisation. For instance, species selection, if truly independent, could in principle counteract selection at the organism level. Vrba's test will however be inconclusive when both selection processes push in the same direction (see a discussion of this problem in Jablonski 2008), but the most parsimonious attitude, that is the simplest one given the evidence, to adopt in such case will be to consider that selection only really occurs at the lower level (see Figure 6.2), unless one has some clear evidence of the contrary. Okasha (2006, 207) illustrates one example of true species selection satisfying Vrba's test with the evolution/maintenance of

sexual reproduction.[102] He asserts, following a classical interpretation, that asexuality is advantageous at the organism level, because of the two-fold cost of producing males (Maynard Smith 1978), but that sexuality is advantageous at the species level because it allows faster evolutionary responses to rapid changes in environmental conditions. According to this reasoning, sexual lineages would be selected via species selection as a process of natural selection distinct from selection at the organism level which favours asexual organisms.

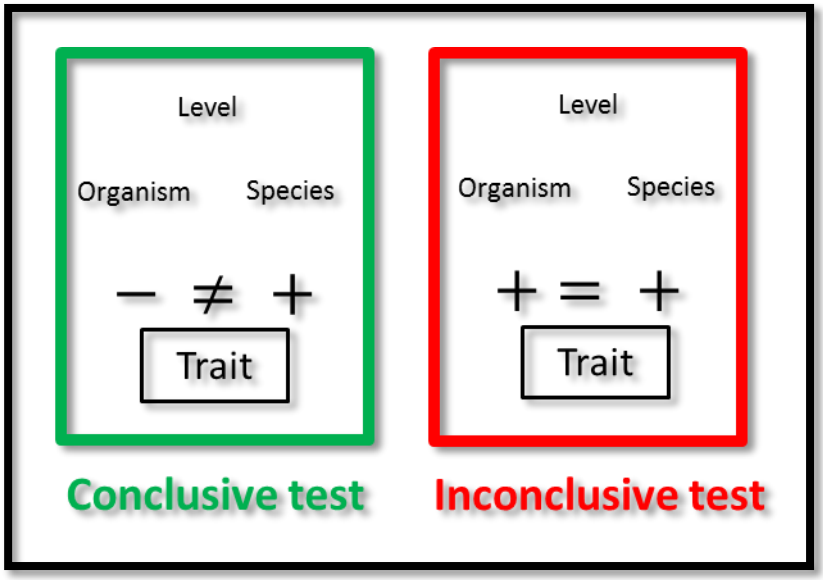


Figure 6.2. Illustration of Vrba's acid test and when it is conclusive.

Some authors (e.g., Damuth & Heisler 1988; Lloyd 1988; Lloyd & Gould 1993) have questioned the validity of Vrba's acid test based on the fact that it should allow for a distinction between 'true species selection' and what she calls 'effect macroevolution' (Vrba 1983). The difference between true species selection and effect macroevolution is the following. The former

[102] For a survey on the issues surrounding the evolution of sex see West, Lively, & Read (1999) and the different commentaries on their article.

refers to a case in which selection at the species level results from an individual character difference. The latter refers to a case in which species level selection results from species level differences that are not reducible to individual level differences (e.g., the frequency of individuals with blue eyes). Okasha (2006, 207), relying on Grantham's (1995) and Stidd and Wade's (1995) analyses, shows however that Vrba's test is valid but that it should not focus on emergent properties but emergent levels of selection. Although I regard the claim that there could be emergent levels of selection without there being emergent properties of fitness bearers as problematic, for the sake of the argument, I accept Okasha's conclusion that Vrba's test is valid. Thus, I will focus here on the notion of 'emergent level of selection' that does not involve an emergent character as this is the case for the evolution and maintenance of sexual reproduction.

One fundamental principle of experimental sciences, concomitant with the manipulationist account of causation I presented in Chapter 1, is that to establish causation, ideally, one would have to change only one variable at a time while the others are kept unchanged or controlled in order to eliminate confounds. Controlled and randomised experiments rely on this idea since they are attempts to reach this goal (Fisher 1970; Shipley 2002). Thus to establish causation as opposed to a mere correlation, if one is interested in measuring the influence of *X* (e.g., a drug) on a population *P*, the experimenter will need to control the effect of *X* on *P* with another population (let us call it *P'* or Control) which was not administered *X* but which is as similar to *P* as possible in all other respects. In such case, if a difference is observed between the two populations, it will *only* be attributable to *X* because no other variable is different. However, if *P* and *P'* are not strictly identical in all respects but *X*, then any observed difference could be attributable to *X or* any of the other different variables between the two populations which could

have the same effect as *X*. Such variables are classically called confounding variables (see Chapter 1). [103]

How is this relevant to our problem of species selection and Vrba's test? Vrba's test is not a scientific experiment per se, but it shares with them the necessity to be controlled. Unless all the variables relevant to selection are strictly identical at both levels in the test, the detection of a different direction of selection at those levels could be attributed either to a different process of selection at each level *or* to any other variable with different values at each level and with some relevance to the direction of evolutionary change. Just like any scientific experiment, Vrba's test requires ideally that only one variable at a time is changed while all the other are kept unchanged.

I noted earlier that Okasha claims that the evolution/maintenance of sexual reproduction is a true case of species selection. He justifies this assertion using Vrba's test. Because, he argues, the test shows that selection pushes in two opposite directions (that is, selection for sexuality at the species level and selection for asexuality at the organism level), a process of selection ontologically different from the process of selection at the organism level must exist at the species level. But does Okasha's comparison eliminate all possible confounding (or at any rate, all known) variables, which would render his conclusion spurious? In other words, is selection at the organism level assessed in the exact same way as it is at the species level? The answer to this question is that it is not; Okasha has overlooked the issue of confounding variables.

To detect this confounding variable, let us use one more time Okasha's example and consider two types of organisms, one asexual and one sexual, under the same selection pressures.

---

[103] Pearl (2000) and Shipley (2002) provide sophisticated statistical tools that rely, among other things, on the simple idea that correlation does not equate causation for situations in which experimentally controlled and randomised experiments are impossible.

The so called two-fold cost of male occurs in sexual organisms, because only half of sexual organisms can reproduce (the females) and they produce offspring of both sexes. On the contrary, asexual organisms are able to reproduce without partner and only produce female organisms. Hence if the two types of organisms are in competition, the asexual ones should quickly out-compete the sexual ones, because of the supplementary costs associated with sexual reproduction. A good example of this phenomenon within the same species is explained by Bouchard (2008, 2011). The Quaking aspen (*Populus tremuloides*) can reproduce both sexually and asexually, but Bouchard (2011) notes that the asexual clonal grove often beats the sexually reproducing individuals. In such cases, it is thus extremely tempting to claim that the fitness of an asexual organism is higher than the fitness of a sexual organism. In fact replacing 'trait' by 'sexual reproduction' or 'asexual reproduction' in Figure 6.2 seems to yield two different levels of selection going in two opposite directions. But, if performed as such, the test is incorrectly applied. To correctly apply it, measures of fitness must be relativised over the same period of time at both levels of organisation. This point has been made by Bouchard (2011) in the context of comparisons between fitness of different of organisms. My point is that it should be made in the context of comparisons of fitnesses between levels (see also Pocheville 2010, Chap. 2).

Why is that? To answer this question, I first need to digress and justify why time is an important parameter to be taken into account in evolutionary explanation. We have seen in Chapter 1 and then in Chapter 5 that the fitness of an organism is a relational property that cannot be directly measured as, for instance, the mass of an object can. As a result, measures of fitness are only proxies for fitness, not fitness in and of itself. Yet, different proxies for fitness can lead to different answers. In fact, the reproductive output *after one generation* of an organism represents

only one proxy for its fitness among an infinite number of proxies (e.g., reproductive output after two generations, survival rate at a specific age, etc.)[104]

In an environment that is perfectly stable all the time, it is very likely that most of these proxies will be equally good. In such cases, having the reproductive output after one or after two generations, for instance, will make no difference in our estimation of whether natural selection occurs between two or more types in a population. However, if every two generations there are changes in the environment that affect differentially the reproductive output of organisms, having only the reproductive output after one generation will be misleading if one wants to know the evolutionary dynamics of a population over periods of time longer than two generations. Similarly having the reproductive output after two generations will be misleading if one wants to predict the evolutionary change for a period of time shorter than two generations.

Thus, there is no 'best' period of time over which one should measure fitness for there is no definite environmental conditions that are the 'true' environmental conditions of an organism. This has led Beatty & Finsen (1989), Sober (2001) and more recently Abrams (2009) to propose a distinction between short-term and long-term fitness. Although they all consider that short-term and long-term reproductive outputs *are* fitness, it should be clear by now that my view is that they represent different *proxies of* fitness.

More generally, proxies of fitness over long periods of time should be preferred if one is interested in evolutionary problems involving long term environmental changes, as it is the case with the evolution and maintenance of sexual reproduction. This is because long term

[104] See also Pocheville (2010, Chap 2) who gives a more definitive argument for the importance for specifying a period of time to measure fitness.

environmental changes and their consequences on selection pressures will be invisible to a proxy for fitness based on the short term reproductive output. Increasing the period of time over which proxies of fitness is evaluated might represent solutions to problems with the concept of fitness identified by Ariew and Lewontin (2004). Yet, many evolutionary problems do not involve such changes and measuring fitness as the reproductive output over one generation is fine because the environment usually is stable enough. This is the case for instance if one wants to know what phenotype is optimal in a constant environment.

The confounding variable or more precisely variables in Okasha's comparison might now become more obvious to the reader. They are the environmental variables that changed with the measures of fitness made at the two levels. At the organism level, fitness is usually measured as the reproductive output after one organism's generation. At the species level, fitness is measured as the rate of extinction or speciation over much longer periods of time, sometimes many millions of years. Thoday (1953) for example considered that fitness should be measured as the probability of leaving a descendent after a period of 10^8 years. Speciation and extinction events are ultimately composed of the deaths, survivals and reproduction events of organisms over many generations, for the former events supervene on the latter ones. Thus, when Okasha applies Vrba's test over the maintenance/evolution of sex, he compares the fitness of organisms over one generation at the organism level with the fitness of organisms over a much higher number of generations.[105] Performed as such, Vrba's test remains inconclusive. Indeed, the difference observed could be either due to two processes of selection pushing in two opposite directions or to two measures of one and the same process of selection over two different periods of time during which the

[105] This is in virtue of the supervenience of speciation and extinction events at the species level on death, survival and reproduction events at the organism level.

environment has changed. In this latter case this process would push in one direction over the short term (in one set of environmental conditions) and in the other over the long term (in another set environmental conditions). In the remainder of the section, I defend the latter possibility.

To start with, let us now perform Vrba's test while controlling the period of time over which fitness is measured (Table 6.2). Controlling for time could be done in two ways: (1) by measuring the two fitnesses at both the species level and the organism level over one organism generation and comparing them over this period of time; (2) by measuring the two fitnesses over the period of time that would normally be used to measure species' fitness, that is, a period long enough to detect events of speciation or extinction. Both alternatives seem to be doomed in practice, since we are neither able to measure the fitness of species over short periods of time, nor able to measure the fitness of organisms over periods of time longer than a few generations. But if we were able to do so, we would certainly find that *ceteris paribus* asexual organisms and asexual species have a higher short-term fitness as measured by (1) than sexual organisms *and* sexual species, but have a lower long-term fitness as measured by (2). The reason for this is not mysterious. Asexual organisms and asexual species on average do better when the environmental conditions remain stable (as it is usually the case over one generation) while sexual organisms and sexual species do better when new environmental conditions arise (which certainly occur over several millions of years). In other words, both selection at the organism and the species level would go in the same direction once the test is controlled for the period of time over which fitness is measured.

Table 6.1 represents Vrba's test performed on the regime of reproduction at the organism and species level when the confounding variable time, a proxy of changes in the environmental

conditions, is not taken into account. Table 6.2 represents the same test performed when time is taken into account. We can see that contrary to what happens when time is not taken into account, we expect selection on the regime of reproduction to go in the same direction at the organism and the species level.

Table 6.1. Vrba's test comparing the direction of selection on the regime of reproduction (sexual or asexual) without controlling for the time over which fitness is measured in organisms and species.

Environment	Reproduction	Organism level	Species level
Not controlled for	Asexual	+	-
	Sexual	-	+

Table 6.2. Vrba's test comparing the direction of selection on the regime of reproduction (sexual or asexual) controlling for the period of time over which fitness is measured in organisms and species.

Environment	Reproduction	Organism level	Species level
Constant – proxy: short period of time	Asexual	+	+
	Sexual	-	-
Changing – proxy: long period of time	Asexual	-	-
	Sexual	+	+

Thus, Okasha's claim that the evolution/maintenance of sexual reproduction occurs as a result of species selection is inexact. If we follow his reasoning keeping the time over which fitness is assessed as a constant (meaning that the environment is the same at both levels), we predict no difference between a measure of selection made at the level of the organism and another one made at the level of the species. The simplest interpretation of difference in the

direction of selection over time is one and the same process of natural selection changes direction over time not that two processes are opposing each other.

There is no logical barrier to extending this argument to all the other cases for which MLS2 has been the framework of choice and in which what is referred as the collective level is made of the particles only (more on this in Section 6.5). In each case, if fitness could be determined over the same period of time or in the same constant environment at each level, what seem to be ontologically different levels of selection could, in principle, be unified under one and the same process. Does it mean that the MLS2 framework should be abandoned and always be replaced by the MLS1 framework? I claim that it should not unless one has the full availability at any point in time of the selection pressures on the particles under consideration. I can only see multilevel *models* as satisfying these criteria. In *practice*, the complete list of selection pressures, keeping the same grain of description, will be unknown and/or constantly changing (e.g., frequency dependent selection) making the particle fitnesses extremely complex to determine over long period of time.

When both particle and collective fitnesses are available, and when the question at stake is about the collective, I propose that the MLS2 framework should be privileged. There are two further reasons for this choice. First, the hard task of measuring fitness of all the particles within a collective (with all the non-linear relations it implies) and over many particles' generations will often materialise at the collective level into a single and easily measurable parameter: the collective's reproductive output. Second, keeping independent the fitness of particles and the fitness of collectives, as it is done in MLS2 framework, can bring different, yet relevant, information about selection pressures which might change over time.

In light of these general considerations on MLS2 what becomes of the MLS1/MLS2 distinction in the context of evolutionary transitions in individuality and especially during their last stage? Would it be in principle possible, at the last stage, to describe the fitness of a collective in terms of the fitnesses of its particles, *contra* Okasha? Following the reasoning I used in the case of the evolution/maintenance of sex, as in any case of MLS2, I see nothing that would prevent it, as long as a multicellular organism is nothing more than the cells it is composed of. [106] In such a case, during an evolutionary transition in individuality, if the fitness of the particles seems incommensurable with the fitness of the collective, it is most probably due to the fact that, during the last stage, both fitnesses are not measured over the same period of time anymore and that the interactions between particles become so complex that tracking back their fitness over longer periods of time than one or two generations appears in practice impossible. Yet, what becomes decoupled in the two levels is not fitness *per se* but generations or life cycles.

Because Michod's proxy for fitness depends on reproductive output after one generation, if 'one generation' does not mean the same thing at the particle and the collective level, it is not surprising that collective and particle fitnesses seem decoupled from each other. This, in fact, is an artefact created by the measure and does not mean that MLS2 represents an ontologically distinct process of selection from MLS1. Rather, it suggests that MLS2 becomes a useful means to carve one single process of natural selection both in time and space, especially once an evolutionary transition in individuality is completed.

[106] Which is certainly a problematic assumption, see Section 6.5.

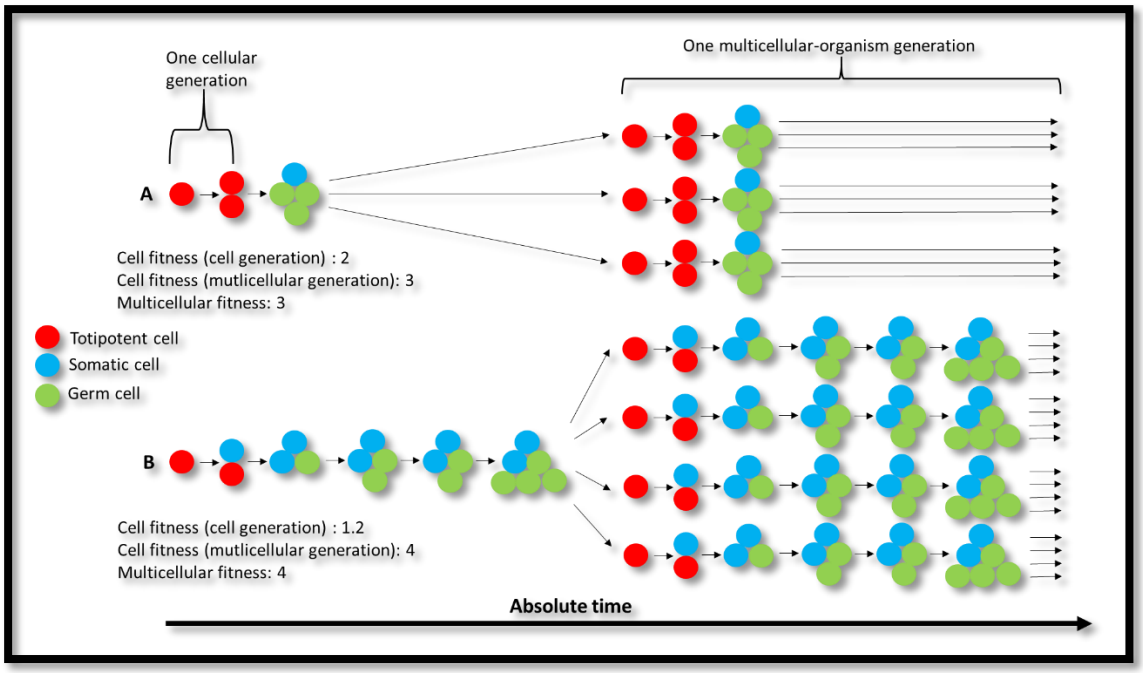


Figure 6.3. Hypothetical case of multicellular organisms with different life cycles.

To see this, let us take again a take a toy example of hypothetical multicellular organisms of two types **A** and **B** that develop through cellular division and then reproduces at the multicellular level (see Figure 6.3). Suppose that type **A** develops through a single (totipotent) cell into two totipotent cells of which one will lead to two germ cells and the other to one germ cell and one somatic cell (which we suppose unable to reproduce). Finally suppose that this hypothetical multicellular organism is able to produce three offspring. Suppose now that type **B**, instead of waiting until the second cellular generation to differentiate the totipotent cells in somatic and germ cells, does so at the first cellular generation. As a result, at the third cell generation, the multicellular organism of type **B** is composed of two somatic cells and one germinal cell, that is, one somatic cell more and two germinal cell less than type **A**. But imagine now that the fact that type **B** has one more somatic cell than type **A** allows it to have three more cellular generations than type **A** before producing the multicellular offspring (see Figure 6.3). This

could be explained, for instance, by supposing that somatic cells gather resources for the multicellular organism and that having more resources allows for more cellular reproduction and a higher number of multicellular offspring, with a priority given to the production of new multicellular offspring until a certain threshold is reached. In the case of type **A**, all the resources gathered go to the production of new multicellular offspring while in the case of type **B**, part of the resources goes to the production of new offspring while the exedent resources allow for more cellular divisions. This results overall in the production of one more multicellular offspring (see Figure 6.3). I am assuming here that it takes the same absolute time to produce a new multicellular generation for both types.

As one can see on Figure 6.3, if we compare the measures of cell fitness of the two types with their measures of multicellular-organisms fitness using respectively one *cellular generation* and one *multicellular-organism generation*, we find that type **A** has a cell fitness of 2 and mutlicellular-organisms fitness of 3 while type **B** has a cell fitness of 1.43[107] and a multicellular-organism fitness of 4. Thus **A** seems to be fitter at the cell level and **B** fitter at the multicellular level. We seem to be in a case where one process of selection pushes in one direction at the cell level and another in the other direction at the mutcellular-organism level, as in the case of species selection developped earlier. Yet, this artefact quickly disappears if we measure the fitness of the cells and multicellular organisms over the same period of time. In this case we can see that both selection at the cell and at the mutlicellular-organism level go in the same direction. In the example on

[107] Since 6 cells are produced over 5 cell generations this means that on average a cell produce 1.43 offspring cell per cell generation.

Figure 6.3, type **A** has a cell fitness (muticellular-organism generation from birth to deah) of 3 and type **B** of 4, which are the same values as those obtained at the multicellular level.

Although the demonstrations provided here suggest that the MLS2 framework does not vindicate an ontologically distinct process of selection from MLS1, it does not render the notion of MLS2 useless. In fact, MLS2 is a very useful mean to carve one single process of natural selection both in time and space and becomes especially useful once an ETI is completed. This echoes a recent criticism made by Waters (2011) about Okasha's fundamentalism over the distinction between MLS1/MLS2 in which he claims that MLS1 and MSL2 frameworks are more conventional rather than fundamental. In a response Okasha (2011, 243) held his ground, restating that they were fundamental. I have provided evidence here that they could be conventional and it becomes apparent once measures of fitness are controlled.

6.5. Changing hierarchies between levels

Before concluding this chapter, I would like to mention another possible source of confusion in the levels-of-selection discussions highly relevant to evolutionary transitions in individuality and which, as far as I am aware, has been largely neglected by the protagonists of the debate. In many cases, when the direction of selection is compared at different levels, it is not uncommon that collectives are constituted of more than the particles with which their fitness is compared to. As a result, the collectives are part of what is referred to as the environment of the particles. In this case, it is simply meaningless to compare the two levels in terms of selection. This is because the two levels is merely two representations of the same world but using different units of measurement. To be meaningful, there should be a perfect mapping between the lower

and the higher level entities in which the higher level entities fully constitute what is referred to as the higher level entities.

To make this clearer, take the case of an allele and an organism. Often, in classical models of population genetics, the genotype at one locus is taken to represent the organism. The problem is that an organism is not made of two alleles and thus this representation is simply false unless many other unrealistic assumptions are made. Other alleles at different loci as well as other developmental factors affect the fate of an organism and those factors, if they have any sort of interaction with the focal allele (and in any organism they will), will be part of the selective environment of this allele. Thus, comparing the direction of selection in a population of alleles reproducing with the direction of selection in a population or organisms and claiming that selection is pushing in two opposite directions is simply problematic. This is because many other factors could be responsible for the difference observed (such as selection on other loci which are not counted as part of the population of alleles considered)

By referring to the same terms at both levels or keeping track of their relations, the claim that selection pushes in two opposite directions can be avoided. Selection in a population of alleles pushes in one direction, but when the same model is applied to a population of alleles and a part of their environment, namely the other alleles of the organism they are part of and other developmental factors, it predicts another direction for the complex allele-organismal environment. Put in those terms, any paradox or deep mystery as to why selection goes in different directions at different levels disappears. Selection could only be assumed to push in two directions if there was a perfect mapping between alleles and organisms so that no part of the organisms would be part of the environment of the alleles. Yet, it is hard see how that would be possible in principle, without invoking a strong notion of emergence that would violate the

supervenience assumption. If this is right, different directions of selection at different levels represent two perspectives on one and the same phenomenon carved in different ways.

A similar point has been made by Sarkar (1994, 1998) in which he explains that two separate hierarchies have often been conflated in discussions over the units of selection. These two hierarchies are on the one hand, allele → locus → gene complex → genotype, etc. and on the other hand, molecule → organelle (including chromosome) → cell → tissue → organism → group, etc. Sarkar believes that the conflation originates from the conflation between organism and genotype. Once the two hierarchies, on which selection can act independently, are distinguished, it does not seem paradoxical that selection can simultaneously 'act' at two levels in the two hierarchies. It also suggests that any conclusion drawn in one hierarchy cannot be straightforwardly translated into the other.

Sarkar's distinction must be linked to the distinction between MLS1 and MLS2 made before. In fact, by noticing that particles (e.g., alleles) and collectives (e.g., organisms) belong to two different hierarchies, we immediately know that the fitness metrics for particles and collectives are different and thus that one should not attempt to compare genotypic fitness (in number of *alleles* present at the next generation) with organismal fitness (in number of successful *gametes* produced). There might be rules of equivalence between the two hierarchies that might render, in some specific conditions, the conclusion(s) drawn within one hierarchy applicable to the other(s), but without first having established those rules, the conclusion drawn within one hierarchy cannot be used in another.

The MLS2 framework might thus be useful in two different scenarios. First, when it is useful to use two different proxies of fitness within one hierarchy as with the case of evolution

of sexual reproduction presented in Section 6.4. Second it might useful to use it when the 'Darwinian substrate' compared at each level is different. Yet, the conclusions made about different processes pushing in different directions are flawed in both cases because the comparisons are either not made within one and the same hierarchy but between two hierarchies or between two different sets of environmental conditions (of which the different periods of times over which fitnesses are assessed are the proxies) as in the case of the evolution and maintenance of sexual reproduction.

6.6. Conclusion of the chapter

In this chapter I have shown several problems associated with the application of a MLS2 thinking in the case of evolutionary transitions in individuality and species selection. First, when doing a multilevel analysis of an evolutionary process, it is important to use comparable concepts at each level. The terms 'viability' and 'fertility' and consequently 'fitness' cannot be used interchangeably when referring to the cell (and more generally the particle) or the multicellular organism (and more generally the collective). Although Michod and colleagues' model formally distinguishes the two, it assumes a direct relation between viability and fertility of the cell and viability and fertility of the organism. I have shown that cell viability and fertility, should not, to be biologically meaningful, straightforwardly be translated respectively into viability and fertility of multicellular organisms even in simple cases such as the ones proposed by Michod and colleagues.

Second I have shown that the concept of *formal* reproduction can be useful to consider during evolutionary transition in individuality and that it has the advantage of avoiding the odd claim that the somatic cells of a multicellular organism have a fitness of nil.

Third, I have demonstrated that time, because it is intimately linked to the notion of environment, is an extremely important parameter to take into account when comparing selection at different level of organisation by the means of fitness proxies. I used the evolution/maintenance of sexual reproduction as an illustrative example to establish that if different proxies of fitness reflecting different time scales are used at the organism and species levels, this will have the consequence of measuring selection pressures over two different sets of environmental conditions. This can lead one to confound the existence of one unique process of selection over two different periods of time with two ontologically distinct processes of selection, one for each level. I applied the same reasoning to evolutionary transitions in individuality and argued that they were not transitions in processes of selection, but rather events for which MLS1 and MLS2 perspectives were alternately more tractable.

Finally, I have shown that the notion of conflicts across levels of selection might be an artefact that results from comparing two hierarchies at different levels of organisation, something that does not lead to a meaningful comparison. At best, as in the case of proxies of fitness being compared over different periods of times, it represents two useful perspectives on one and the same evolutionary process.

The claim that the distinction between collective selection and particle selection is conventional is not new (e.g. : Dugatkin & Reeve 1994; Sterelny 1996) and Kerr & Godfrey-Smith (2002) have formalised this equivalence in the MLS1 context. Yet, as Okasha (2006, 136)

rightly points out, this formalism has been made solely in the context of MLS1. Although a recent criticism calls into question the equivalence between particle and collective selection (conventionalism) in the MLS1 context (see Lloyd, Lewontin, & Feldman 2008) that could potentially undermine the claim about conventionalism I have made in this chapter, whether this criticism is valid does not undermine my claims about time and fitness. Suffice to say that the apparent incommensurability between fitnesses at different levels of organisation, *at least partially* results from an artefact of measuring fitness over different periods of time at different levels of organisation. This is sufficient reason to take time seriously into account in fitness comparisons. By combining the arguments developed in the previous chapter on grains of descriptions and this chapter, it might be possible in the future to make an important step towards a formalism in which the MLS2 framework reduces to a MLS1 framework without denying that the MLS2 framework is a useful, but an idealised one, in many contexts.

7. General conclusion and further directions: A Glimpse on Cultural Evolution

7.1. General conclusion

Throughout this thesis I have aimed at extending the concept of natural selection and the concept of ENS beyond their 'natural' biological boundaries. It is now time to recapitulate what has been achieved. I will then briefly show how these tools can be useful for understanding the domain of cultural evolution from a Darwinian perspective.

I wrote in the Introduction of this thesis that my strategy to delimit the essential features of the concept of natural selection will be to use different conceptual tools. To do so I started, in Chapter 1 by providing a precise definition of fitness and sharply distinguishing fitness from measures or proxies of fitness which have been obstacle to consistently generalise the concept of natural selection across different fields. If fitness is equated with reproductive output, for example, we run into difficulties (some of which were exposed in Chapter 2). I have shown that what is really fundamental to fitness, following one of Godfrey-Smith's distinctions, is the notion of intrinsic-invariable properties. In fact, differences between entities in such properties, when they are causally upstream to growth (relativised to the population size), that is: persistence, reproductive output and developmental growth, are differences in fitness.

With this distinction between, on the one hand, intrinsic-invariable properties and on the other hand, extrinsic and intrinsic-variable properties, I then proposed a possible interpretation of what one can reasonably mean by claiming that natural selection is a *cause* of evolutionary

change. I used causal modelling and directed acyclic graphs to show that the notion of natural selection as a cause of evolution could be justified using the manipulationist account of causation and argued that natural selection and drift (of which I have distinguished two kinds) could be regarded as individual level causes and not population level causes as it has been argued by some in the literature. I showed that the relation between individual properties and evolutionary change is more stable or invariant than between population level properties and evolutionary change. The workability of my arguments in the framework of the manipulationist account of causation gives us a strong incentive to reject the view that natural selection and drift are mere statistical aggregate without real causal efficacy as it has been proposed by the statisticalists (see Chapter 1). Statisticalists argue that the efficacious causal events underlying evolutionary change are death, birth and reproduction events and that consequently natural selection cannot be a cause, but also that the difference between drift and natural selection is a purely statistical one. I showed that if birth, death, reproduction events occur differentially between entities of a population and that those differences are the result of differences in intrinsic-invariable properties, it is justified to regard any resulting evolutionary change as the result of natural selection. Conversely when differences are due to difference in intrinsic-variable or extrinsic properties, it is justified to regard any resulting evolutionary change as the result of a drift process.

Having a precise definition of fitness also allowed me to further investigate the links between natural selection the *process* and its various possible *products* (e.g., ENS, complex adaptation). More particularly, I have defended the view using three different strategies that reproduction and inheritance are conceptually separate from the process of natural selection. Although reproduction and inheritance are one way through which to instantiate the process of natural selection, this by no means represents the only one. In Chapter 2 I argued against the

view that process of natural selection, in and of itself, does not require perfect transmission of traits between generations, a proposal which Godfrey-Smith calls 'evolutionary nominalism.' I argued that in a population in which there is natural selection and no other evolutionary process, perfect transmission between generations was the only possible products. In fact, if transmission is imperfect, because new variation is introduced in the population, another evolutionary force must be at work.

I then showed that the biological and philosophical literature does not have an unambiguous definition of the process of natural selection. Some definitions include the notion of reproduction and/or inheritance whilst other do not. I demonstrated that, strictly speaking, reproduction is conceptually distinct from natural selection, proposing several cases more or less far from classical biological cases in which it would be inconsistent to claim that natural selection is not occurring. Yet, the reason why these cases are not typically thought of as cases in which natural selection is occurring is because they challenge our intuitive concepts of individuality, which I have shown, rely ultimately on arbitrary distinctions. Following a similar strategy as with reproduction, I showed that biased inheritance, which is a quantitative rather than qualitative measure of inheritance at the individual level, is also conceptually distinct from the process of natural selection.

In Chapter 3 I pursued the question of the relations between natural selection, reproduction and inheritance. Using individual-based modelling, I showed that reproduction and biased inheritance are not necessary for the process of natural selection to occur and that ENS can result in spite of them being absent. Following a strategy outlined by Okasha (2006) called the strategy of endogenisation, starting from a population of mere persistors, that is, entities unable to reproduce and unable to transmit reliably their traits, I demonstrated, using a set of

individual-based models, that reproduction and biased inheritance can evolve by cumulative evolution. Since a form of natural selection is the only potential candidate to explain this cumulative evolution, I argued that this demonstrated that natural selection neither requires reproduction nor biased inheritance to produce evolutionary change and more importantly to produce complex adaptations (resulting in cumulative evolution). Inheritance, I showed, can be effectively endogenised.

In Chapter 4 I further reinforced the view that persistence and developmental growth can have in some cases the same role as reproduction. To do so, I demonstrated that there exists a point of tension between the verbal approach to ENS encountered in chapters 2 and 4 and a more formal approach I introduce in this chapter. I presented one version of the Price equation proposed by Okasha that makes apparent a term of heritability which is classically understood as a population level measure of inheritance between parent and offspring and is usually considered to be a necessary ingredient for ENS to occur. I noticed that, in this form, a temporal interpretation of the Price equation in which individuals do not reproduce would still require an interpretation of the heritability term. Yet, without parent and offspring but only individuals persisting differently over time, heritability cannot be interpreted in terms of similarities between parent(s) of offspring as it is classically assumed. From there I argued that the classical notion of heritability was incomplete and that a more general notion of heritability measuring not only parent-offspring resemblance but also the resemblance between persisting entities between two particular times should be included in the verbal formulation of ENS. I followed a similar strategy to show that the concept of fitness in the temporal interpretation of the Price equation cannot be considered as reproductive output since there are no offspring in the cases described under it. I proposed to decompose fitness in the Price equation into the three components that seem to

be general enough to capture all possible case of evolutionary change (including ENS), namely persistence, developmental growth and reproductive output.

The main gain from chapters 1, 2, 3 and 4 has been to show that it is possible to develop conceptual tools both coherent and consistent with the classical tools used in evolutionary theory and that can be used to treat cases that do not fit well with those classical tools. This represents a gain from what can be found in the literature. In many cases, the classical tools have been 'tinkered' to accommodate what are often regarded as marginal cases. But no real conceptual work has made to make those tools workable in many different contexts. By conceptually separating natural selection from its products and fitness from the notion of reproductive output, one becomes more able to apply the concept of natural selection to other domains than classical population biology, while avoiding being trapped by unessential biological specificities.

In Chapter 5, I returned to the concept of fitness and argued that its measures were necessarily bound to the concept of type. I proposed that fitness, once operationalised, is measured by how much of the same type is produced over time and that natural selection leads to a relative change in frequency of types. Yet, type-identity, either between parent and offspring or between two random members of a population depends on what grain of description is used to make the comparison. Thus one population of entities, once described from different perspectives, can be considered either as evolving by natural selection or as evolving by other evolutionary processes. I drew some parallels between this problem (the reference grain problem) and the reference class problem encountered in probability theory, arguing that they have a common origin. Far from having provided a solution to the reference grain problem, I have suggested that one way to accommodate it would be to consider that natural selection is a model

of a process highly sensitive to the grain of description used and thus having a meaningful concept of natural selection requires explicitly stating the grain of description used in each case.

Finally in Chapter 6, I have shown that the claim that fitness is exported from a lower to a higher level during evolutionary transitions in individuality is an artefact of measures of proxies of fitness across different temporal scales at different levels of organisation. Once the same temporal scale is used consistently across levels of organisation, the temptation to call the higher level of organisation a new level of selection ontologically distinct from the lower one can easily be resisted. Missing this implicit change of temporal scale when changing levels of organisation creates the illusion of different processes of selection pushing in different directions *at different levels* when it only reflects that one selection process goes in different directions *over different periods of time* because of a changing environment. Towards the end of Chapter 6, I proposed that MLS2 explanations are the result of referring intuitively to two hierarchies, at two different levels of organisation, in an evolutionary explanation. I argued that comparing selection processes when they refer to two different hierarchies is problematic. The main purpose of this chapter was to oppose the view that MLS2 cases represent ontologically distinct levels of selection. Rather, I argued that they represent one and the same process seen from different perspectives (e.g., temporal).

In sum, the first chapter of the thesis argued that the process of natural selection was an individual-level cause of evolution, chapters 2, 3 and 4 proposed different tools to free the concept of natural selection from its biological specificities to demonstrate its true explanatory generality. In Chapters 5 and 6, I showed the limitations of thinking of natural selection across levels of organisation without paying attention to the different parameters involved by moving

from one level to the other but also how important it is to relate any claim about natural selection (and other evolutionary processes) to the grain of description used.

7.2. Further directions: Cultural evolution

With this picture of the process of natural selection stripped down from its inessential components and the awareness of the limitations of evolution explanation especially with regards to its dependence on levels of organisation, temporal scales and grains of description, one can now assess more confidently whether and to what extent the tools of evolutionary theory can meaningfully be applied to other domains. To demonstrate some of what can be done, I will use cultural evolution as a case study, especially concerning the question as to whether cultural evolution is Darwinian, which has been debated in the literature on cultural evolution. I have already briefly mentioned cultural evolution in Section 3.5. I argued that considering reproduction and inheritance as complex adaptations, once they have been endogenised in a more general conception of ENS, rather than as necessary conditions for ENS, could to free cultural evolution (and more particularly cultural selection) from the relentless pursuit of 'memes' (e.g., Aunger 2000; Blackmore 1999; Dawkins 1976). And yet, problematically, the replicator model has been used as a 'benchmark' to decide whether and to what extant cultural evolution is Darwinian. I think this enterprise is misguided. Many ideas besides the endogenisation of inheritance I developed throughout the thesis demonstrate why using the replicator as a benchmark to decide whether cultural evolution is Darwinian is a non-starter and my work provides some alternative tools.

First of all, let us remark that by asking whether cultural evolution is Darwinian one can mean two different things that can easily be conflated. By *Darwinian* one could mean *in some way the result of natural selection* or *that produces complex adaptations*. I argued in Chapter 2, that natural selection and ENS (including when it leads to complex adaptations) should be conceptually distinguished. We will see towards the end of this section that it is important to make this distinction in the cultural domain.

Second, if one applies the conclusion developed in Chapter 5, they might argue that cultural selection, that is, natural selection in the cultural domain, can and certainly does occur in some cases, provided that a particular grain of description is used. One might recognise at the same time that using a different grain of description would yield a different answer. Thus concretely, if my framework is right, the claim that cultural selection occurs depends on whether biological individuals, ideas, cultural groups, entire cultures or a different kind of entity is taken to be the bearer of intrinsic-invariable properties. Whether cultural selection occur will also depend on which time scale one choses to measure the evolutionary change since intrinsic-invariable properties are only relative to one particular time scale.

As I argued in chapters 5 and 6, there is no non-arbitrary way for deciding what grain of description and what temporal scale should be used to measure evolutionary change. All we know, if my arguments are correct, is that an entity able to participate in a Darwinian process should have some intrinsic-invariable properties leading to some difference in the representation of its type after *some* time in a collection of entities with different intrinsic-invariable properties in a given population. The same conclusion should naturally apply to the case of cultural evolution. *Prima facie*, there seems to be many sorts of cultural entities that could legitimately have some intrinsic-invariable properties depending on the perspective one uses. From a very coarse

grained perspective, an entire culture for instance, certainly has some intrinsic-invariable properties (e.g., agriculture, farming, the wheel). Yet, once described from a much finer-grain, say an individual level perspective, these intrinsic-invariable properties might look highly variable. Thus, if there is disagreement over whether cultural evolution is Darwinian (and we will see that there is), a first step should be to assess whether the different protagonists of the debate use the same grain of description and periods of time. Although those distinctions are crucial, as far as I am aware, analyses of grains of descriptions and temporal scales exist no more in the literature on cultural evolution than in the biological literature.

Dan Sperber and his colleagues have for many years argued that there is a disanalogy between cultural and biological evolution with respect to its Darwinian character (Claidière, et al. 2014; Claidière & Sperber 2007; Sperber 1996, 2000). According to Sperber, the fundamental difference between biological and cultural evolution is that:

> There is much greater slack between descent and similarity in the case of cultural transmission than there is in the biological case. Most cultural descendants are transformations, not replicas. (1996, 108)

Thus for Sperber, the analogy of a cultural replicator (the meme) based on the biological replicator (the gene) is misguided.

Of course Sperber is careful enough to explain that there is a much greater variance in the fidelity of transmission of cultural items than there is in the biological ones, which is overwhelmingly DNA based. To appreciate the variance of cultural transmission compare a computer virus, transmitted with high fidelity at each event of transmission versus the transmission of the folktale 'Little Red Riding Hood' from mouth to ear with parts of the tale

being reconstructed and transformed by the mind of each individual. But the basic point Sperber makes is that *overall* cultural evolution is more mutational than it is selective.

Thus for Sperber, what looks like cultural selection is in fact cultural transformation. And because the human brain is roughly the same across all humanity, this leads to the same sorts of transformations being made in a similar ecological and social environment. Sperber calls this phenomenon cultural attraction. According to this view, if there is only one version of 'Little Red Riding Hood' in our culture, it is not because it has replicated at a higher rate than any other version that might have existed before. Rather, it is because the other versions have been transformed in the mind of the individual receivers in a way that leads to the version we all know. Thus instead of being *replicated*, the tale is *re-produced*. Although Sperber's cultural attraction model is insightful, there is reason to doubt that this distinction is as clear cut as it seems to be.

One might think, as Sperber does, that during each event of transmission in the cultural domain between two humans, there is a lot of transformation and thus that selection is mostly not responsible for the evolution observed. If this is correct, then there are literally no cultural replicators and no entities with intrinsic properties that remain invariable to allow natural (cultural) selection and have important evolutionary consequences. Although this reasoning seems fair, it has unexpected consequences when applied to the biological case. In fact the view that biological evolution is replicative is not exactly correct. If looked from a very fine grained perspective most descendants in the biological domain are also transformations not replicas, even genes. After the replication process has occurred, the end product strains of DNA might contain some modifications from the original. It is to some extent transformed, as in the example of the tale, since, at particular sites the original template is not always used during the replication process. This can happen in a variety of situations such as when the original template has a lesion due to

UV radiation. In such cases errors can be introduced in the replicated DNA molecule by a mechanism called translesion synthesis, in which the lesion is bypassed by incorporating an incorrect nucleotide opposite to a non-instructional or a mis-instructional DNA lesion (Friedberg et al. 2006). Another good example is radioactive labelling in which the radioactivity level can be different between parent and offspring strains of DNA. See for instance the experiment of Hershey and Chase (1952).

Another reason why replication in the biological domain might not be as pervasive as is usually thought, is that DNA is only one of different products transmitted during an event of so called replication. Other products are also transmitted, what many people call epigenetic factors, such as methylation patterns, which can have significant effects on the phenotype and may be transmitted across many generations (Jablonka & Lamb 2005). But methylation patterns and other epigenetic factors are not always transmitted perfectly in a coding-decoding fashion as with DNA. Among these different mechanisms some look more as if they are replicated and others more as if they make a difference in the phenotype of the offspring without the same pattern or 'code' being present in parents and offspring (see Jablonka & Raz 2009 for a survey of those different mechanisms).This point is important if we consider the existing tension between the molecular and evolutionary (or replicator) conceptions of the gene.

The evolutionary gene, the notion of the gene defended by Williams for example, is an informational (and thus abstract) entity, not a physical one (Williams 1985, 1992). The evolutionary gene is at the basis of the replicator view or gene's eye view of evolution. The molecular notion of the gene basically corresponds to a piece of DNA. Yet, these two notions of

genes do not perfectly match with each other. This is a fact that is often unappreciated both in the philosophical and biological literature (see Griffiths & Neumann-Held 1999; Griffiths & Stotz 2013). In fact, although all molecular genes are evolutionary genes, the converse is not true. Some epigenetic factors are also transmitted across generations and sometimes even in a Mendelian fashion (see for instance Cubas, Vincent, & Coen 1999); they are thus effectively evolutionary genes. If some of those factors are not strictly speaking 'replicated' and have substantial evolutionary effects, what in the biological domain is called replication is in fact based on an idealised notion of gene that does not represent accurately the biological processes involved. This is precisely the point made by Sperber and colleagues in the cultural domain Claidière, et al. (2014).

Some might argue at this point that the degree of similarity between the physical reproducing entity and the informational concept of replicator is vastly superior in the biological case than it is in the cultural case. This might be a fair point. There is a concern with this claim however, namely that the discrepancy between the notions of evolutionary gene and molecular gene and what part of evolutionary gene are non DNA based is unknown and hard to establish. It might turn out that the evolutionary gene and the molecular gene have a high degree of mismatch and that many epigenetic factors do not beneficiate of the similar mechanisms of replication than the machinery involved in DNA replication (proofreading, repair etc.). If so, then Sperber's argument for cultural evolution would have to be run in the biological case.

But there might be other reasons why the analogy between biological and cultural evolution could be seen as misguided. Claidière and André (2012), who are Sperberians, offer two different arguments. They argue, first, that contrary to the biological case, cultural transmission influences the way information is transmitted and, second, that in the biological case

there are only a few reliable mechanisms of transmission while there are potentially an infinity in the cultural case. These problems are important and indeed cast some doubt on the analogy between cultural and biological evolution and they might be the main reason why Sperberians find the notion of Darwinian cultural evolution questionable. That said, Claidière and André base their analysis on 'classical' evolutionary biological models which I have myself criticised throughout the thesis as being not general enough even for some biological cases. Furthermore it is not clear to me whether Claidière and André refer to the evolutionary or molecular concept of the gene as they seem to slide between the two. Thus one would need to investigate whether their criticisms still stand after the models and the assumptions made about fitness and transmission have been thoroughly stripped down from their biological specificities.

There is another reason why one might want to question the analogy between cultural evolution and biological evolution as being Darwinian which, as far as I am aware, has been overlooked by the Sperberians. This argument pertains to the second sense I have distinguished under which *Darwinian* can be understood, namely as *that process that produces complex adaptations*. This argument is perfectly compatible with the cultural attractor model and one of its advantage is that it does not rely on the problematic concept of type-identity. But to run this argument, I first need to give some more detail on the cultural attractor model. Following the cultural attractor model, cultural stability, that is, the fact that most individuals of a population exhibit one and the same cultural trait, is not the sole result of a selective process but also involve a 'mutational' one. According to the Sperberians the macro-level stability is not explained solely by micro-level

events of replication during transmission following by selection but also by the partial reconstruction and transformation which I referred to, throughout the thesis, as 'mutation'.[108]

One very common assumption in the biological domain is that evolution or lack of evolution as in cases of stabilising selection, is the result of blind variation and selective retentions to use the famous formula of Donald Campbell (1960). In other words the production of new variation is generally assumed to be random. This is assumed by Dawkins (1986, 41) for example and for him one strength of the concept of natural selection is that it can produce design without directed variation. Some summaries of ENS, such as the one proposed by Fracchia and Lewontin (1999) clearly stipulate that variation is random.

If the Sperberian picture is right, in the cultural domain, mutations occurs more preferentially towards attractors rather than unpredictably. Here is Sperber's 'Little Red Riding Hood' example encountered earlier explaining why only some forms of this tale are encountered using the notion of cultural attraction:

> For instance, suppose an incompetent teller has the hunters extract Little Red Riding Hood from the Big Bad Wolf's belly, but forgets the grandmother. Meme theorists might want to argue - and I would agree with them - that such a version is less likely to be replicated than the standard one. The meme theorists' explanation would be that this version is less likely to have descendants. This is indeed plausible. There is another explanation, however, which is also plausible: that hearers whose knowledge of the story derives from this defective version are likely to consciously or unconsciously correct the story when they retell it, and, in their narrative, to bring the grandmother back to life too. In the logical space of possible versions of a tale, some versions have a better form: that is, a form

[108] In the same way one might want to distinguish evolutionary gene from molecular gene, one might want to distinguish an evolutionary concept of mutation from a molecular concept of mutation. I refer thus here to the evolutionary concept of mutation.

> seen as being without either missing or superfluous parts, easier to remember, and more attractive.(1996, 108)

Instead of a pure explanation in terms of survival and reproduction of successful variants and in which some could mutate accidentally, a substantial part of the explanation of the cultural stability will be provided by factors of attraction that most if not all individuals of a population will be submitted to. Factors of attractions will be considered here as the cause for mutations. Note that factors of attraction are not necessarily cognitive biases.

In Chapter 2 (Section 2.3.3), I have provided a quote from Godfrey-Smith (2007) who states that Darwinian evolution can occur even when variation is directional and that in such cases natural selection might not have as much explanatory power as when it is random. The attractor model in cultural evolution exemplifies why this is the case. Natural selection cannot be considered as the main difference maker leading to the distribution observed. The Sperberians, I take it, consider that the production of variation, in many cases of cultural evolution is so overwhelmingly directional that the Darwinian explanation in which selection is an important factor for the phenomenon observed is thus misleading. Explanation in terms of cultural attractors on the other hand account for a much greater part of the evolutionary change observed.

Thus, putting the pieces together, the debate as to whether cultural evolution is Darwinian does not reduce to answering the question of whether there is selection of cultural items for there is no absolute answer to this question due to the dependence of the presence of natural selection on the grain of description chosen. Furthermore, answering the question in this way might be the result of a misguided picture of biological evolution in which two concepts of gene have been conflated. Instead I propose that what is crucial for a Darwinian explanation is whether the

complex traits observed are the result of blind variation and thus whether natural selection is the only *known* factor that can account for them.

If one wants to answer this question, whether replication occurs becomes a secondary answer for the reasons invoked above and more generally throughout the thesis. However, if cultural variants are modified during events of transmission in such a way that we can predict with higher than chance probability the direction of the modifications (because of some factors of attraction), an explanation in terms of selection will clearly be insufficient, especially if the selection pressure between events of transmission is weak. In such a case, it seems that the cultural evolutionary change resulting will not be Darwinian in the way a biological evolutionary change is.

In light of those considerations, the debate as to whether cultural evolution is Darwinian seems to amount to whether the notion of cultural selection can do some explanatory work in the cultural domain. Although directional mutation seems to be an important factor to take into consideration generally in cultural evolution, I think nevertheless that for at least some cases the notions of cultural selection and cultural adaptation can have an important explanatory role. In fact, in cases where one cultural variant ends up invading the population and having effects on its own success while other variants with the same level of directional mutation appear but are less successful in persisting in time, it is hard to see how neglecting selection and focus solely directed transformation could provide a suitable explanation. Of course, the Sperberians might argue that such case is never or rarely realised.

I agree that from a very fine-grained perspective the case might only rarely be realised. However, I disagree that this can be generalised to any temporal scale or any grain of description.

For example, suppose that one wants to compare different cultures or cultural groups and understand why some are more successful than others. The concept of cultural selection can have an important explanatory role in this case. I do not want to enter here in debate over whether cultural group selection *exist* (or not) as opposed to individual level selection since whether it exists cannot be independent from the grain of description used. The important consideration, in my view, with cultural-group selection is thus not whether selection can occur on cultural groups but whether a cultural group selection *perspective* or *explanation* can be useful. What a 'cultural group' is often not very well defined, but if a systematic criterion compatible with the idea that a cultural group bears some intrinsic-invariable properties at least for some time can be found, there is no reason why cultural groups could not undergo a process of cultural selection resulting in cultural adaptation at that particular grain of description. The only requirement for that to happen is that there is sufficient differential growth or reproduction of the type of each culture over time allowing for multiplication. This is in fact what has been envisioned, although not developed in details, by Bouchard (2007).

Assuming a large number of societies with different cultural items, it should not be excluded on this large scale that among cultural items with the same level of attraction some of those items had an impact on the persistence, growth and reproduction of cultural groups. Of course there will always be particular reasons leading one group to adopt one particular policy at the group level and decisions ultimately reside in individual minds. But facing the 1167 societies of the Ethnographic Atlas (Murdock 1967) or even the more modest 186 societies of the Standard Cross Cultural Sample (Murdock & White 1969) a subset of the Ethnographic Atlas[109], it might

[109] See for instance, in the domain of the evolution of religion, the studies by Bourrat, Atkinson, and Dunbar (2011), Roes and Raymond (2003) and Johnson (2005), which all use the Ethnographic Atlas and/or the Standard Cross Cultural Sample.

be a good strategy to consider that at least in some cases cultural differences from this coarse-grained perspective represent intrinsic and invariable properties of societies or cultural groups. Similarly it might be a good strategy to consider that the most common cultural traits are to some extent optimised due to the result of a cultural selection process or more modestly that these properties *cause* the success of cultures they are part of. The different accounts given by Diamond in *Guns, Germs, and Steel* (1997) or *Collapse* (2005) of why some societies are more successful than other due to large scale factors such as degree of deforestation, latitude at which is found the society etc., deserve at least for some of those factors, in my view, to be called cultural group level explanation.

Diamond often emphasises what he calls 'geographic determinism'. For example, in a 2004 comparative study he and a colleague, Barry Rolett, showed that deforestation in Pacific islands was the result of different geographical factors, such as, but not limited to rainfall, elevation and age of soil. Deforestation is itself a strong predictor for the collapse of a society (Rolett & Diamond 2004). These factors need to be separated from what can be considered as intrinsic-invariable properties of a society. In a 2008 article, Rolett extends his analysis of contributing factors for sustainability, although in a more qualitative way than the 2004 study (Rolett 2008). He distinguishes the geographical factors from what he calls 'cultural adaptations' such as whether the cultural practice irrigated culture or arboriculture. It is in my view justified to consider those traits as invariable-intrinsic factors, for they do not straightforwardly originate from the environmental conditions nearby and they refer to something that is sufficiently stable over time from a particular grain of description. Considering those factors as intrinsic-invariable properties, one can then provide an explanation in terms of cultural-group selection.

Of course this explanation could be thoroughly reduced in terms of individual differences, not differences at the level of the culture. This is precisely what I have shown in Chapter 6 with the case of species selection; but that does not mean a cultural-group selection (or species selection) perspective is not an interesting one when many cultures are compared. By making particular abstractions, they very effectively *explain* why some cultural traits are found more often than other ones in some cultures with particular ecological conditions without having to explain every specificity of each culture. Only their ontological interpretation is problematic.

Other examples of works using a cultural selective framework at the level of the group or culture, although not necessarily explicitly, include the study from Snarey (1996), in which he shows that a very dry environment is a predictor of beliefs in moralising gods.[110] Snarey hypothesises that such beliefs allow societies in dry environments to better persist. Another study from Sherman and Billing (1999) shows that recipes of meats in cultures with higher risks of pathogens spoils more commonly include the use of diverse spices which kill pathogens. In these two cases, it would be hard to explain how directional mutation, without selection, could effectively explain differences in recipes between cultures and how to analyse this phenomenon from a purely individualistic perspective for the commerce of spices involves highly nonlinear relation between individuals. These examples show that both the notion of group selection and the notion of cultural selection is meaningful and can be useful in particular cases. I have provided here a few examples, many more exist (e.g., Fincher, Thornhill, Murray, & Schaller 2008; Wilson 2002).

[110] On this topic see also Roes and Raymond (2003), Johnson (2005) and Atkinson and Bourrat (2011)

I hope I have demonstrated that the conceptual distinctions surrounding the concept of natural selection made throughout the thesis are not superfluous. I have shown that the biological model of evolution and, in particular, its genetic version, has specificities that should not be applied to the general case, I showed this with respect to the cultural case, but there is reason to think that my point generalises to other domains, such as non-zoological branches of biology. Yet, some of these specificities, such as the notions of intrinsicness and invariability seem, although not explicitly stated as such in the classical presentations of the biological case, intimately linked to the notion of selection. That is the reason, I believe, they should be the ones used when moving from a classical biological population to a population of entities that do not reproduce and transmit their traits to their offspring as most macroscopic metazoans do.

References

Abrams, M. (2009). The Unity of Fitness. *Philosophy of Science*, 76(5), 750-761.

Abrams, M. (2012). Measured, Modeled, and Causal Conceptions of Fitness. *Frontiers in Genetics*, 3, 196.

Ariew, A., & Lewontin, R. C. (2004). The confusions of fitness. *The British Journal for the Philosophy of Science*, 55(2), 347-363.

Arnold, S. J., & Wade, M. J. (1984). On the measurement of natural and sexual selection: theory. *Evolution*, 38(4), 709-719.

Atkinson, Q. D., & Bourrat, P. (2011). Beliefs about God, the afterlife and morality support the role of supernatural policing in human cooperation. *Evolution and Human Behavior*, 32(1), 41-49.

Aunger, R. (2000). *Darwinizing culture*. Oxford: Oxford University Press.

Beatty, J. H. (1984). Chance and natural selection. *Philosophy of science*, 51(2), 183-211.

Beatty, J. H., & Finsen, S. K. (1989). Rethinking the propensity interpretation: A peek inside Pandora's box. In M. Ruse (Ed.), *What the Philosophy of Biology Is: Essays Dedicated to David Hull* (pp. 17-30). Dordrecht: Kluwer Publishers.

Beisson, J., & Sonneborn, T. M. (1965). Cytoplasmic inheritance of the organization of the cell cortex in Paramecium aurelia. *Proceedings of the National Academy of Sciences of the United States of America*, 53(2), 275.

Blackmore, S. J. (1999). *The meme machine*. New York: Oxford University Press.

Bonner, J. T. (1959). *The cellular slime molds*. Princeton: Princenton University Press.

Bouchard, F. (2004). *Evolution, fitness and the struggle for persistence*. Unpublished PhD diss., Duke University.

Bouchard, F. (2007). *Ideas that stand the [evolutionary] test of time*, Interdisciplines: adaptation and representation. Paris: C.N.R.S.

Bouchard, F. (2008). Causal processes, fitness, and the differential persistence of lineages. *Philosophy of Science*, 75(5), 560-570.

Bouchard, F. (2011). Darwinism without populations: a more inclusive understanding of the "Survival of the Fittest". *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 42(1), 106-114.

Bouchard, F., & Huneman, P. (Eds.). (2013). *From Groups to Individuals: Evolution and Emerging Individuality*. Cambridge, MA: MIT Press.

Bouchard, F., & Rosenberg, A. (2004). Fitness, probability and the principles of natural selection. *The British Journal for the Philosophy of Science*, 55(4), 693-712.

Bourrat, P. (2014). From survivors to replicators: evolution by natural selection revisited. *Biology & Philosophy*, 29(4), 517-538.

Bourrat, P. (in press-a). How to read "heritability" in the recipe approach to natural selection. *British Journal for The Philosophy of Science*.

Bourrat, P. (in press-b). Levels of Selection Are Artefacts of Different Fitness Temporal Measures. *Ratio*.

Bourrat, P., Atkinson, Q. D., & Dunbar, R. I. M. (2011). Supernatural punishment and individual social compliance across cultures. *Religion, Brain & Behavior*, 1(2), 119-134.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Boyd, R., & Richerson, P. J. (2005). *The Origin and Evolution of Cultures*. Oxford: Oxford University Press.

Brandon, R. N. (1978). Adaptation and evolutionary theory. *Studies In History and Philosophy of Science Part A*, 9(3), 181-206.

Brandon, R. N. (1990). *Adaptation and environment*. Princeton: Princeton University Press.

Brandon, R. N. (2005). The difference between selection and drift: A reply to Millstein. *Biology and Philosophy*, 20(1), 153-170.

Brandon, R. N. (2014). Natural Selection. *Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/archives/spr2014/entries/natural-selection/

Brandon, R. N., & Carson, S. (1996). The Indeterministic Character of Evolutionary Theory: No" No Hidden Variables Proof" but No Room for Determinism Either. *Philosophy of Science*, 63(3), 315-337.

Burnet, M. (1958). *The Clonal Selection Theory of Acquired Immunity*. Cambridge: Cambridge University Press.

Burt, A., & Trivers, R. (2006). *Genes in conflict: the biology of selfish genetic elements*: Harvard University Press.

Buss, L. W. (1983). Evolution, development, and the units of selection. *Proceedings of the National Academy of Sciences*, 80(5), 1387-1391.

Buss, L. W. (1987). *The Evolution of Individuality*. Princeton, NJ: Princeton University Press

Byerly, H. C., & Michod, R. E. (1991). Fitness and evolutionary explanation. *Biology and Philosophy*, 6(1), 45-53.

Cairns-Smith, A. G. (1990). *Seven clues to the origin of life: a scientific detective story*. Cambridge: Cambridge University Press.

Calcott, B., & Sterelny, K. (2011). *The major transitions in evolution revisited*. Cambridge, MA: MIT Press.

Campbell, D. T. (1960). Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological review*, 67(6), 380.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution : a quantitative approach*. Princeton, N.J: Princeton University Press.

Charlesworth, B. (1994). *Evolution in age-structured populations (2nd edition)* (Vol. 2). Cambridge, UK: Cambridge University Press.

Claidière, N., & André, J. B. (2012). The transmission of genes and culture: A questionable analogy. *Evolutionary Biology*, 39(1), 12-24.

Claidière, N., Scott-Phillips, T. C., & Sperber, D. (2014). How Darwinian is cultural evolution? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1642), 20130368.

Claidière, N., & Sperber, D. (2007). The role of attraction in cultural evolution. *Journal of Cognition and Culture*, 7, 89-111.

Cleland, C. (2013). Conceptual Challenges for Contemporary Theories of the Origin of Life. *Current Organic Chemistry*, 17(Special Issue on Prebiotic Chemistry), 1704-1709.

Crow, J. F., & Kimura, M. (1970). *An introduction to population genetics theory*. New York: Harper and Row.

Cubas, P., Vincent, C., & Coen, E. (1999). An epigenetic mutation responsible for natural variation in floral symmetry. *Nature*, 401(6749), 157-161.

Damuth, J., & Heisler, I. L. (1988). Alternative formulations of multilevel selection. *Biology and Philosophy*, 3(4), 407-430.

Darwin, C. (1859). *On the origin of species by means of natural selection*. London: J. Murray.

Dawkins, R. (1976). *The Selfish Gene*. Oxford, UK: Oxford University Press.

Dawkins, R. (1982). *The Extended Phenotype: The long reach of the gene*. New York: Oxford University Press, USA.

Dawkins, R. (1986). *The Blind Watchmaker*. Harlow: Longman.

Dawkins, R. (1996). *Climbing mount improbable*: WW Norton & Company.

De Monte, S., & Rainey, P. B. (2014). Nascent multicellular life and the emergence of individuality. *Journal of Biosciences*, 39(2), 237-248.

Dennett, D. C. (1995). *Darwin's dangerous idea*. New York: Touchstone.

Diamond, J. M. (1997). *Guns, germs and steel*. New York: W.W. Norton & Co.

Diamond, J. M. (2005). *Collapse: how societies choose to fail or succeed: revised edition*. London: Lane.

Dobzhansky, T., & Pavlovsky, O. (1957). An experimental study of interaction between genetic drift and natural selection. *Evolution*, 11, 311-319.

Doolittle, W. F. (2014). Natural selection through survival alone, and the possibility of Gaia. *Biology & Philosophy*, 29(3), 415-423.

Downes, S. M. (2009a). Heritability. *Stanford Encyclopeadia of Philosophy*. Retrieved from http://plato.stanford.edu/archives/spr2014/entries/heredity/

Downes, S. M. (2009b). Moving past the levels of selection debates. *Biology and Philosophy*, 24(5), 703-709.

Dugatkin, L. A., & Reeve, H. K. (1994). Behavioral ecology and levels of selection: dissolving the group selection controversy. *Advances in the Study of Behavior*, 23, 101-133.

Dunbar, R. I. M. (1982). Adaptation, fitness and evolutionary tautology. In K. s. C. S. Group (Ed.), *Current Problems in Sociobiology*. Cambridge: Cambridge University Press.

Dyson, F. J. (1999). *Origins of life*. Cambridge: Cambridge University Press.

Eagle, A. (2004). Twenty-one arguments against propensity analyses of probability. *Erkenntnis*, 60(3), 371-416.

Earnshaw-Whyte, E. (2012). Increasingly Radical Claims about Heredity and Fitness. *Philosophy of Science*, 79(3), 396-412.

Edelman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection*. New York: Basic Books.

Edwards, A. W. F. (1994). The fundamental theorem of natural selection. *Biological Reviews*, 69(4), 443-474.

Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10), 465-523.

El Mouden, C., André, J. B., Morin, O., & Nettle, D. (2014). Cultural transmission and the evolution of human behaviour: a general approach based on the Price equation. *Journal of Evolutionary Biology*, 27(2), 231-241.

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

Ewens, W. J. (1989). An interpretation and proof of the fundamental theorem of natural selection. *Theoretical population biology*, 36(2), 167-180.

Falconer, D. S. (1981). *Introduction to quantitative genetics*. New York: Longman.

Fincher, C. L., Thornhill, R., Murray, D. R., & Schaller, M. (2008). Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism. *Proceedings of the Royal Society B: Biological Sciences*, 275, 1279-1285.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.

Fisher, R. A. (1970). *The design of experiments* (8th ed.). New York: Hafner.

Folse, H. J., & Roughgarden, J. (2010). What is an individual organism? A multilevel selection perspective. *The Quarterly Review of Biology*, 85(4), 447-472.

Fracchia, J., & Lewontin, R. C. (1999). Does culture evolve? *History and theory*, 38(4), 52-78.

Frank, S. A. (1997). The Price equation, Fisher's fundamental theorem, kin selection, and causal analysis. *Evolution*, 51(6), 1712-1729.

Frank, S. A. (1998). *Foundations of Social Evolution*. Princeton: Princeton University Press.

Frank, S. A. (2012a). Natural selection. III. Selection versus transmission and the levels of selection. *Journal of Evolutionary Biology*, 25(2), 227-243.

Frank, S. A. (2012b). Natural selection. IV. The Price equation. *Journal of Evolutionary Biology*, 25(6), 1002-1019.

Friedberg, E. C., Walker, G. C., Siede, W., Wood, R. D., Schultz, R. A., & Ellenberger, T. (2006). *DNA repair and mutagenesis* (2nd ed.). Washington, DC: American Society for Microbiology (ASM).

Fry, I. (2011). The role of natural selection in the origin of life. *Origins of Life and Evolution of Biospheres*, 41(1), 3-16.

Futuyma, D. J. (2005). *Evolution*. Sunerland, MA: Sinauer Associates, Inc.

Futuyma, D. J. (2006). *Evolutionary Biology*. Sunderland, MA: Sinauer Associates, Inc.

Gardner, A., & Conlon, J. P. (2013). Cosmological natural selection and the purpose of the universe. *Complexity*, 18(5), 48-56.

Gardner, A., & Foster, K. R. (2008). The evolution and ecology of cooperation–History and concepts. In J. Korb & J. Heinze (Eds.), *Ecology of social evolution* (pp. 1-36). Heidelberg: Springer.

Gardner, A., & Grafen, A. (2009). Capturing the superorganism: a formal theory of group adaptation. *Journal of Evolutionary Biology*, 22(4), 659-671.

Gillespie, J. H. (2004). *Population genetics: a concise guide*. Baltimore, MD: Johns Hopkins University Press.

Glymour, B. (2001). Selection, indeterminism, and evolutionary theory. *Philosophy of Science*, 68(4), 518-535.

Godfrey-Smith, P. (2001). Three kinds of adaptationism. In S. H. Orzack & E. Sober (Eds.), *Adaptationism and optimality* (pp. 335-357). Cambridge: Cambridge University Press.

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy*, 21(5), 725-740.

Godfrey-Smith, P. (2007). Conditions for evolution by natural selection. *Journal of Philosophy*, 104(10), 489.

Godfrey-Smith, P. (2009a). Abstractions, idealizations, and evolutionary biology. In A. Barberousse, M. Morange & T. Pradeu (Eds.), *Mapping the Future of Biology* (pp. 47-56). Dordrech: Springer.

Godfrey-Smith, P. (2009b). *Darwinian populations and natural selection*. New York: Oxford University Press, USA.

Godfrey-Smith, P. (2009c). Models and fictions in science. *Philosophical Studies*, 143(1), 101-116.

Godfrey-Smith, P. (2011). Darwinian populations and transitions in individuality. In B. Calcott & K. Sterelny (Eds.), *The Major Transitions in Evolution Revisited*. Cambridge, MA: MIT Press.

Godfrey-Smith, P. (2012). Darwinism and cultural change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1599), 2160-2170.

Godfrey-Smith, P., & Lewontin, R. C. (1993). The dimensions of selection. *Philosophy of Science*, 60(3), 373-395.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society B: Biological Sciences*, 205(1161), 581-598.

Grafen, A. (2000). Developments of the Price equation and natural selection under uncertainty. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1449), 1223-1227.

Grantham, T. A. (1995). Hierarchical Approaches to Macroevolution: Recent Work on Species Selection and the "Effect Hypothesis". *Annual Review of Ecology and Systematics*, 26, 301-321.

Graves, L., Horan, B. L., & Rosenberg, A. (1999). Is indeterminism the source of the statistical character of evolutionary theory? *Philosophy of science*, 66(1), 140-157.

Griesemer, J. R. (2000). The units of evolutionary transition. *Selection*, 1(1), 67-80.

Griesemer, J. R. (2005). The informational gene and the substantial body: On the generalization of evolutionary theory by abstraction. In M. R. Jones & N. Cartwright (Eds.), *Idealization XII: Correcting the Model—Idealization and Abstraction in the Sciences* (pp. 59-115). Amsterdam: Rodopi.

Griesemer, J. R. (2014). Reproduction and the Scaffolded Development of Hybrids. In L. R. Caporael, J. R. Griesemer & W. C. Wimsatt (Eds.), *Developing Scaffolds in Evolution, Culture, and Cognition* (pp. 23). Cambridge, MA: MIT Press.

Griffiths, P. E., & Neumann-Held, E. M. (1999). The many faces of the gene. *BioScience*, 49(8), 656-662.

Griffiths, P. E., & Stotz, K. (2013). *Genetics and Philosophy: An Introduction*. Cambridge: Cambridge University Press.

Grimm, V., & Railsback, S. F. (2005). *Individual-based modeling and ecology*: Princeton university press.

Haig, D. (2012). The strategic gene. *Biology and Philosophy*, 27(4), 461-479.

Hájek, A. (2007). The reference class problem is your problem too. *Synthese*, 156(3), 563-585.

Hájek, A. (2012). Interpretations of probability. *Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/archives/win2012/entries/probability-interpret/

Hamilton, M. (2009). *Population genetics*. Oxford: Wiley-Blackwell.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. I. *Journal of theoretical Biology*, 7(1), 1-16.

Hamilton, W. D. (1975). Innate social aptitudes of man: an approach from evolutionary genetics. In R. Fox (Ed.), *Biosocial anthropology* (pp. 133-153). London: Malaby Press.

Harper, J. L. (1977). *Population biology of plants*. New York: Academic Press.

Hartl, D. L., & Clark, A. G. (1997). *Principles of population genetics* (Vol. 116). Sunderland: Sinauer associates.

Heisler, L., & Damuth, J. (1987). A method for analyzing selection in hierarchically structured populations. *American Naturalist*, 130(4), 582-602.

Helanterä, H., & Uller, T. (2010). The Price equation and extended inheritance. *Philosophy & Theory in Biology*, 2.

Hershey, A. D., & Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *The Journal of general physiology*, 36(1), 39-56.

Hodgson, G. M., & Knudsen, T. (2010). *Darwin's conjecture: The search for general principles of social and economic evolution*. Chicago: University of Chicago Press.

Horan, B. L. (1994). The statistical character of evolutionary theory. *Philosophy of science*, 61(1), 76-95.

Hull, D. L. (1975). Central subjects and historical narratives. *History and Theory*, 14(3), 253-274.

Hull, D. L. (1980). Individuality and selection. *Annual Review of Ecology and Systematics*, 11(1), 311-332.

Huneman, P. (2012). Natural selection: A case for the counterfactual approach. *Erkenntnis*, 76(2), 171-194.

Jablonka, E., & Lamb, M. J. (2005). *Evolution in four dimensions: Genetic, epigenetic, behavioral, and symbolic variation in the history of life*. Cambridge, MA: MIT Press.

Jablonka, E., & Raz, G. (2009). Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *The Quarterly Review of Biology*, 84(2), 131-176.

Jablonski, D. (2008). Species selection: theory and data. *Annual Review of Ecology, Evolution, and Systematics*, 39, 501-524.

Jacquard, A. (1983). Heritability: one word, three concepts. *Biometrics*, 39, 465-477.

Jerne, N. K. (1955). The natural-selection theory of antibody formation. *Proceedings of the National Academy of Sciences of the United States of America*, 41(11), 849-857.

Johnson, D. D. P. (2005). God's Punishment and Public Goods: A Test of the Supernatural Punishment Hypothesis in 186 World Cultures. *Human Nature*, 16(4), 410-446.

Kerr, B., & Godfrey-Smith, P. (2002). Individualist and multi-level perspectives on selection in structured populations. *Biology and Philosophy*, 17(4), 477-517.

Kerr, B., & Godfrey-Smith, P. (2009). Generalization of the Price equation for evolutionary change. *Evolution*, 63(2), 531-536.

Kettlewell, H. B. D. (1955). Selection experiments on industrial melanism in the Lepidoptera. *Heredity*, 9(3), 323-342.

Krimbas, C. B. (2004). On fitness. *Biology and Philosophy*, 19(2), 185-203.

Laland, K. N., & Brown, G. R. (2002). *Sense and nonsense : evolutionary perspectives on human behaviour*. Oxford; New York: Oxford University Press.

Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics*, 1(1), 1-18.

Lewontin, R. C. (1985). Adaptation In R. Levins & R. C. Lewontin (Eds.), *Dialectics and reductionism in ecology* (pp. 65–84). Cambridge, MA: Harvard University Press.

Libby, E., & Rainey, P. B. (2013). A conceptual framework for the evolutionary origins of multicellularity. *Physical biology*, 10(3), 035001.

Lloyd, E. A. (1988). *The structure and confirmation of evolutionary theory*. New York: Greenwood Press.

Lloyd, E. A., & Gould, S. J. (1993). Species selection on variability. *Proceedings of the National Academy of Sciences*, 90(2), 595-599.

Lloyd, E. A., Lewontin, R. C., & Feldman, M. W. (2008). The Generational Cycle of State Spaces and Adequate Genetical Representation. *Philosophy of science*, 75(2), 140-156.

Lumsden, C. J., & Wilson, E. O. (1981). *Genes, Mind, and Culture*. Cambridge, MA: Harvard University Press.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of science*, 67(1), 1-25.

Mameli, M. (2004). Nongenetic selection and nongenetic inheritance. *The British journal for the philosophy of science*, 55(1), 35-71.

Matthen, M., & Ariew, A. (2002). Two ways of thinking about fitness and natural selection. *The Journal of Philosophy*, 99(2), 55-83.

Matthen, M., & Ariew, A. (2009). Selection and Causation. *Philosophy of Science*, 76(2), 201-224.

Maynard Smith, J. (1978). *The evolution of sex*. Cambridge: Cambridge University Press.

Maynard Smith, J. (1988). *Evolutionary progress and levels of selection*. In M. H. Nitecki (Ed.), Evolutionary progress (Vol. 219, pp. 219-230). Chicago: University of Chicago Press.

Maynard Smith, J., & Szathmáry, E. (1995). *The Major Transitions in Evolution*. Oxford: Freeman & Co.

McCallum, H., & Jones, M. (2006). To lose both would look like carelessness: Tasmanian devil facial tumour disease. *PLoS biology*, 4(10), e342.

McShea, D. W. (1996). Metazoan complexity and evolution: is there a trend? *Evolution*, 50(2), 477-492.

McShea, D. W., & Brandon, R. N. (2010). *Biology's first law: the tendency for diversity and complexity to increase in evolutionary systems*. Chicago: University of Chicago Press.

Mesoudi, A., Whiten, A., & Laland, K. N. (2004). Perspective: Is human cultural evolution Darwinian? Evidence reviewed from the perspective of The Origin of Species. *Evolution*, 58(1), 1-11.

Michod, R. E. (1999). *Darwinian dynamics*. Princeton: Princeton University Press.

Michod, R. E. (2005). On the transfer of fitness from the cell to the multicellular organism. *Biology and Philosophy*, 20(5), 967-987.

Michod, R. E. (2011). Sex and multicellularity as evolutionary transitions in individuality. In B. Calcott & K. Sterelny (Eds.), *The Major Transitions in Evolution Revisited*. Cambridge, MA: MIT press.

Michod, R. E., Nedelcu, A. M., & Roze, D. (2003). Cooperation and conflict in the evolution of individuality: IV. Conflict mediation and evolvability in Volvox carteri. *BioSystems*, 69(2-3), 95-114.

Michod, R. E., & Roze, D. (1999). Cooperation and conflict in the evolution of individuality. III. Transitions in the unit of fitness. In C. L. Nehaniv (Ed.), *Mathematical and Computational Biology: Computational Morphogenesis, Hierarchical Complexity, and Digital Evolution* (pp. 47-92). Providence, Rhode Island: American Mathematical Society.

Michod, R. E., Viossat, Y., Solari, C. A., Hurand, M., & Nedelcu, A. M. (2006). Life-history evolution and the origin of multicellularity. *Journal of theoretical Biology*, 239(2), 257-272.

Mills, S. K., & Beatty, J. H. (1979). The propensity interpretation of fitness. *Philosophy of Science*, 46(2), 263-286.

Millstein, R. L. (2000). *Is the evolutionary process deterministic or indeterministic? An argument for agnosticism*.

Millstein, R. L. (2002). Are Random Drift and Natural Selection Conceptually Distinct? *Biology and Philosophy*, 17(1), 33-53.

Millstein, R. L. (2006). Natural selection as a population-level causal process. *The British Journal for the Philosophy of Science*, 57(4), 627-653.

Millstein, R. L. (2013). Natural Selection and Causal Productivity. In H.-K. Chao, S.-T. Chen & R. L. Millstein (Eds.), *Mechanism and causality in biology and economics* (pp. 147-163). Dordrecht: Springer.

Murdock, G. P. (1967). *Ethnographic atlas*. Pittsburgh: University of Pittsburgh Press.

Murdock, G. P., & White, D. R. (1969). Standard cross-cultural sample. *Ethnology*, 8, 329-369.

Nanay, B. (2011). Replication without replicators. *Synthese*, 179(3), 455-477.

Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press.

Nilsson, D.-E., & Pelger, S. (1994). A pessimistic estimate of the time required for an eye to evolve. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 256(1345), 53-58.

Northcott, R. (2010). Walsh on Causes and Evolution. *Philosophy of Science*, 77(3), 457-467.

Nunney, L. (1985). Group selection, altruism, and structured-deme models. *American Naturalist*, 126(2), 212-230.

O'Malley, M. A., & Dupré, J. (2007). Size doesn't matter: towards a more inclusive philosophy of biology. *Biology & Philosophy*, 22(2), 155-191.

Okasha, S. (2006). *Evolution and the Levels of Selection*. New York: Oxford University Press, USA.

Okasha, S. (2008). Fisher's fundamental theorem of natural selection—a philosophical analysis. *The British Journal for the Philosophy of Science*, 59(3), 319-351.

Okasha, S. (2009). Individuals, groups, fitness and utility: multi-level selection meets social choice theory. *Biology and Philosophy*, 24(5), 561-584.

Okasha, S. (2010). Replies to my critics. *Biology and Philosophy*, 25(3), 425-431.

Okasha, S. (2011). Reply to Sober and Waters. *Philosophy and Phenomenological Research*, 82(1), 241-248.

Orzack, S. H., & Forber, P. (2010). Adaptationism. *Stanford encyclopedia of philosophy*. Retrieved from http://plato.stanford.edu/archives/win2012/entries/adaptationism/

Paul, L. A., & Hall, N. (2013). *Causation: A User's Guide*. Oxford: Oxford University Press.

Pearl, J. (2000). *Causality: models, reasoning and inference*. Cambridge: Cambridge University Press.

Pocheville, A. (2010). *La Niche Ecologique: Concepts, Modèles, Applications*. Ecole Normale Supérieure de Paris-ENS Paris, Paris.

Price, G. R. (1970). Selection and covariance. *Nature*, 227, 520-521.

Price, G. R. (1972a). Extension of covariance selection mathematics. *Annals of human genetics*, 35(4), 485-490.

Price, G. R. (1972b). Fisher's 'fundamental theorem' made clear. *Annals of human genetics*, 36(2), 129-140.

Queller, D. C. (1992). A general model for kin selection. *Evolution*, 46(2), 376-380.

Queller, D. C., & Strassmann, J. E. (2009). Beyond society: the evolution of organismality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1533), 3143-3155.

Railsback, S. F., & Grimm, V. (2011). *Agent-based and individual-based modeling: a practical introduction*. Princeton: Princeton University Press.

Rainey, P. B., & Kerr, B. (2010). Cheats as first propagules: A new hypothesis for the evolution of individuality during the transition from single cells to multicellularity. *Bioessays*, 32(10), 872-880.

Rainey, P. B., & Rainey, K. (2003). Evolution of cooperation and conflict in experimental bacterial populations. *Nature*, 425, 72-74.

Ratcliff, W. C., Denison, R. F., Borrello, M., & Travisano, M. (2012). Experimental evolution of multicellularity. *Proceedings of the National Academy of Sciences*, 109(5), 1595-1600.

Reisman, K., & Forber, P. (2005). Manipulation and the Causes of Evolution. *Philosophy of Science*, 72(5), 1113-1123.

Reutlinger, A., Schurz, G., & Hüttemann, A. (2014). Ceteris paribus laws. *Stanford encyclopedia of philosophy*. Retrieved from http://plato.stanford.edu/archives/spr2014/entries/ceteris-paribus/

Rice, S. H. (2004). *Evolutionary theory: Mathematical and conceptual foundations*. Sunderland, MA: Sinauer.

Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago: University of Chicago Press.

Ridley, M. (1996). *Evolution (2nd edition)*. Oxford: Blackwell

Robertson, A. (1966). A mathematical model of the culling process in dairy cattle. *Anim. Prod*, 8(95), 108.

Roes, F. L., & Raymond, M. (2003). Belief in moralizing gods. *Evolution and Human Behavior*, 24(2), 126-135.

Rolett, B. V. (2008). Avoiding collapse: pre-European sustainability on Pacific Islands. *Quaternary International*, 184(1), 4-10.

Rolett, B. V., & Diamond, J. M. (2004). Environmental predictors of pre-European deforestation on Pacific islands. *Nature*, 431(7007), 443-446.

Rosenberg, A. (1985). *The structure of biological science*. Cambridge: Cambridge University Press.

Rosenberg, A. (1988). Is the Theory of Natural Selection a Statistical Theory? *Canadian Journal of Philosophy*, 18(sup1), 187-207.

Rosenberg, A. (1994). *Instrumental biology, or the disunity of science*. Chicago: University of Chicago Press.

Rosenberg, A. (2001). Discussion note: Indeterminism, probability, and randomness in evolutionary theory. *Philosophy of science*, 68(4), 536-544.

Rosenberg, A., & Bouchard, F. (2005). Matthen and Ariew's obituary for fitness: reports of its death have been greatly exaggerated. *Biology and Philosophy*, 20(2), 343-353.

Roughgarden, J. (1979). *Theory of population genetics and evolutionary ecology*. New York: Macmillan.

Salmon, W. C. (1994). Causality without counterfactuals. *Philosophy of Science*, 61(2), 297-312.

Salmon, W. C. (1998). *Causality and explanation*. Oxford: Oxford University Press.

Sarkar, S. (1994). The Selection of Alleles and the Additivity of Variance. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1994, 3-12.

Sarkar, S. (1998). *Genetics and reductionism*. Cambridge: Cambridge Univ Press.

Schaffer, J. (2005). Contrastive causation. *The Philosophical Review*, 114(3), 327-358.

Scriven, M. (1959). Explanation and prediction in evolutionary theory. *Science*, 130, 477-482.

Sesardic, N. (2005). *Making sense of heritability*. Cambridge: Cambridge University Press.

Shapiro, L., & Sober, E. (2007). Epiphenomenalism–The Do's and the Don'ts. In G. Wolters & P. Machamer (Eds.), *Thinking about Causes: From Greek Philosophy to Modern Physics* (pp. 235–264). Pittsburgh: University of Pittsburgh Press.

Sherman, P. W., & Billing, J. (1999). Darwinian Gastronomy: Why We Use Spices Spices taste good because they are good for us. *BioScience*, 49(6), 453-463.

Shipley, B. (2002). *Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference*. Cambridge: Cambridge University Press.

Smolin, L. (1992). Did the universe evolve? *Classical and Quantum Gravity*, 9(1), 173-191.

Snarey, J. (1996). The natural environment's impact upon religious ethics: a cross-cultural study. *Journal for the Scientific Study of Religion*, 35(2), 85-96.

Sober, E. (1984). *The nature of selection*. Cambridge, MA: MIT Press.

Sober, E. (2001). The two faces of fitness. In R. S. Singh, C. B. Krimbas, D. B. Paul & J. Beatty (Eds.), *Thinking about evolution: historical, philosophical, and political perspectives*. Cambridge: Cambridge University Press.

Sober, E. (2002). The two faces of fitness *Thinking about evolution: historical, philosophical, and political perspectives*. Cambridge: Cambridge University Press.

Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

Sober, E., & Wilson, D. S. (2011). Adaptation and natural selection revisited. *Journal of evolutionary biology*, 24(2), 462-468.

Sperber, D. (1996). *Explaining culture: A naturalistic approach*. Cambridge: Cambridge University Press.

Sperber, D. (2000). An objection to the memetic approach to culture. In R. Aunger (Ed.), *Darwinizing culture: the status of memetics as a science* (pp. 163-173). Oxford: Oxford University Press.

Stamos, D. N. (2001). Quantum indeterminism and evolutionary biology. *Philosophy of science*, 68(2), 164-184.

Stearns, S. C. (1976). Life-history tactics: a review of the ideas. *Quarterly review of biology*, 51(1), 3-47.

Stephens, C. (2004). Selection, Drift, and the "Forces" of Evolution. *Philosophy of Science*, 71(4), 550-570.

Sterelny, K. (1996). The return of the group. *Philosophy of Science*, 63(4), 562-584.

Stidd, B. M., & Wade, D. L. (1995). Is species selection dependent upon emergent characters? *Biology and Philosophy*, 10(1), 55-76.

Strevens, M. (2003). *Bigger than chaos: Understanding complexity through probability*. Cambridge, MA: Harvard University Press.

Strevens, M. (2011). Probability out of determinism. In C. Beisbart & S. Hartmann (Eds.), *Probabilities in Physics* (pp. 339-364). New York: Oxford Univeristy Press.

Sydes, M. A., & Peakall, R. (1998). Extensive clonality in the endangered shrub Haloragodendron lucasii (Haloragaceae) revealed by allozymes and RAPDs. *Molecular Ecology*, 7(1), 87-93.

Thoday, J. (1953). Components of fitness. *Symposium of the Society for Experimental Biology*, 7, 96-113.

Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., et al. (2006). Convergent adaptation of human lactase persistence in Africa and Europe. *Nature genetics*, 39(1), 31-40.

Van Valen, L. M. (1976). Energy and evolution. *Evolutionary theory*, 1(1), 179-229.

Van Valen, L. M. (1989). Three paradigms of evolution. *Evolutionary theory*, 9(1), 1-17.

Venn, J. (1876). *The logic of chance: An essay on the foundations and province of the theory of probability, with especial reference to its application to moral and social science (2nd edition)*. London: Macmillan.

Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era—concepts and misconceptions. *Nature Reviews Genetics*, 9(4), 255-266.

Vrba, E. S. (1983). Macroevolutionary trends: new perspectives on the roles of adaptation and incidental effect. *Science*, 221(4608), 387-389.

Vrba, E. S. (1989). Levels of selection and sorting with special reference to the species level. *Oxford Surveys of Evolutionary Biology*, 6, 111-168.

Walsh, D. M. (2000). Chasing shadows: natural selection and adaptation. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 31(1), 135-153.

Walsh, D. M. (2007). The Pomp of Superfluous Causes: The Interpretation of Evolutionary Theory. *Philosophy of Science*, 74(3), 281-303.

Walsh, D. M. (2010). Not a Sure Thing: Fitness, Probability, and Causation. *Philosophy of Science*, 77(2), 147-171.

Walsh, D. M., Lewens, T., & Ariew, A. (2002). The Trials of Life: Natural Selection and Random Drift. *Philosophy of Science*, 69(3), 429-446.

Waters, K. C. (2011). Okasha's Unintended Argument for Toolbox Theorizing. *Philosophy and Phenomenological Research*, 82(1), 232-240.

Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids. *Nature*, 171(4356), 737-738.

Weber, M. (2001). Determinism, Realism, and Probability in Evolutionary Theory. *Philosophy of science*, 68(3), S213-S224.

Weisberg, M. (2012). Getting Serious about Similarity. *Philosophy of Science*, 79(5), 785-794.

Weisberg, M. (2013). *Simulation and similarity: using models to understand the world*. New York: Oxford University Press USA.

West, S. A., Griffin, A. S., & Gardner, A. (2007). Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology*, 20(2), 415-432.

West, S. A., Griffin, A. S., & Gardner, A. (2008). Social semantics: how useful has group selection been? *Journal of Evolutionary Biology*, 21(1), 374-385.

West, S. A., Lively, C. M., & Read, A. F. (1999). A pluralist approach to sex and recombination. *Journal of Evolutionary Biology*, 12(6), 1003-1012.

Wilensky, U. (1999). NetLogo. http://ccl.northwestern.edu/netlogo/ (Version 5.0). Evanston, IL: Center for Connected Learning and Computer-Based Modeling, Northwestern University.

Williams, G. C. (1966). *Adaptation and natural selection: a critique of some current evolutionary thought*. Princeton: Princeton University Press.

Williams, G. C. (1985). A defense of reductionism in evolutionary biology. *Oxford surveys in evolutionary biology*, 2, 1-27.

Williams, G. C. (1992). *Natural selection: domains, levels, and challenges*. New York: Oxford University Press.

Wilson, D. S. (1975). A theory of group selection. *Proceedings of the national academy of sciences*, 72(1), 143-146.

Wilson, D. S. (2002). *Darwin's cathedral: Evolution, religion, and the nature of society*. Chicago: University of Chicago Press.

Wilson, D. S., & Wilson, E. O. (2007). Rethinking the theoretical foundation of sociobiology. *The Quarterly Review of Biology*, 82(4), 327-348.

Woodward, J. (2000). Explanation and invariance in the special sciences. *The British Journal for the Philosophy of Science*, 51(2), 197-254.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press, USA.

Woodward, J. (2010). Causation in biology: stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, 25(3), 287-318.

Woodward, J. (2013). Causation and manipulability. *Stanford encyclopedia of philosophy*. Retrieved from http://plato.stanford.edu/archives/win2013/entries/causation-mani/